Studies in Emotion and Social Interaction

Demystifying Emotions

A Typology of Theories in Psychology and Philosophy

AGNES MOORS

Demystifying Emotions

Demystifying Emotions provides a comprehensive typology of emotion theories in psychology (evolutionary, network, appraisal, goal-directed, psychological constructionist, and social) and philosophy (feeling, judgmental, quasi-judgmental, perceptual, embodied, and motivational) in a systematic manner with the help of tools from philosophy of science, allowing scholars in both fields to understand the commonalities and differences between these theories. Agnes Moors also proposes her own novel, skeptical theory of emotions, called the goal-directed theory, based on the central idea that all kinds of behaviors and feelings are grounded in goal-striving. Whereas most scholars of emotion do not call the notion of emotion itself into question, this review engages in a critical examination of its scientific legitimacy. This book will appeal to readers in psychology, philosophy, and related disciplines who want to gain a deeper understanding of the controversies at play in the emotion domain.

AGNES MOORS is an Associate Professor in the Faculty of Psychology and Educational Sciences, KU Leuven, Belgium, where she is a member of the Research Group of Quantitative Psychology and Individual Differences and the Center for Social and Cultural Psychology. Her research combines theoretical work, informed by philosophy, with empirical work.

STUDIES IN EMOTION AND SOCIAL INTERACTION

Second Series

Titles published in the Second Series:

The Psychology of Facial Expression, edited by James A. Russell and José Miguel Fernández-Dols

Emotions, the Social Bond, and Human Reality: Part/Whole Analysis, by Thomas J. Scheff

Intersubjective Communication and Emotion in Early Ontogeny, edited by Stein Bråten

The Social Context of Nonverbal Behavior, edited by Pierre Philippot, Robert S. Feldman, and Erik J. Coats

Communicating Emotion: Social, Moral, and Cultural Processes, by Sally Planalp

Emotions across Languages and Cultures: Diversity and Universals, by Anna Wierzbicka

Feeling and Thinking: The Role of Affect in Social Cognition, edited by Joseph P. Forgas

Metaphor and Emotion: Language, Culture, and Body in Human Feeling, by Zoltán Kövecses

Gender and Emotion: Social Psychological Perspectives, edited by Agneta H. Fischer

Causes and Consequences of Feelings, by Leonard Berkowitz

Emotions and Beliefs: How Feelings Influence Thoughts, edited by Nico H. Frijda, Antony S. R. Manstead, and Sacha Bem

Identity and Emotion: Development through Self-Organization, edited by Harke A. Bosma and E. Saskia Kunnen

(Continued after the index)

Demystifying Emotions

A Typology of Theories in Psychology and Philosophy

Agnes Moors
KU Leuven

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05-06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107066342
DOI: 10.1017/9781107588882

First published 2022

A catalogue record for this publication is available from the British Library.

Library of Congress Cataloging-in-Publication Data
Names: Moors, Agnes, author.
Title: Demystifying emotions : a typology of theories in psychology and philosophy / Agnes Moors, KU Leuven.
Description: Cambridge, United Kingdom ; New York, NY : Cambridge University Press, 2022. | Series: Studies in emotion and social interaction. Second series | Includes bibliographical references and index.
Identifiers: LCCN 2021062724 (print) | LCCN 2021062725 (ebook) | ISBN 9781107066342 (hardback) | ISBN 9781107683204 (paperback) | ISBN 9781107588882 (epub)
Subjects: LCSH: Emotions.
Classification: LCC BF531 .M658 2022 (print) | LCC BF531 (ebook) | DDC 152.4–dc23/eng/20220118
LC record available at https://lccn.loc.gov/2021062724
LC ebook record available at https://lccn.loc.gov/2021062725

ISBN 978-1-107-06634-2 Hardback

Contents

Figures

Tables

Boxes

Preface

Almost every book on emotions starts with a description of a variety of emotions – ranging from joy over being kissed back by a love interest, the nagging regret over a short-sighted decision, pride in the accomplishment of a difficult task, sadness over a definitive goodbye, fear of being attacked in a dark alley, and anger about being mistreated or underestimated. This is followed by emphasizing how emotions are the spice of our lives without which we would be condemned to leading a dull existence. After that, the reader is alerted to the fact that the history of our science shows eras of emotional drought when scientists did not know how to fit emotions in their impoverished frameworks, followed by eras of emotional flood in which a wealth of emotion theories have mushroomed but that show a striking lack of consensus. Authors then promise to provide solace by presenting a novel way of organizing the differences and similarities between theories, and if the reader is lucky, a novel theory that deals with all the challenges of its predecessors and beyond.

I share these authors' ambition to present a novel way of organizing the literature, and even in proposing a novel theory of the phenomena called emotions. Yet, whereas most scholars of emotion do not call the notion of emotion itself into question nor its causal powers, I will critically examine its scientific legitimacy and fruitfulness. Without denying the scientific interest in explaining the phenomena themselves, that is, the fact that our everyday routines are traversed by ups and downs, I will join the small brigade of skeptical theorists in their attempt to demystify the phenomena called emotions.

The first aim of the book is to compare major emotion theories in psychology (i.e., evolutionary, network, appraisal, goal-directed, psychological constructionist, and social theories) and philosophy (i.e., feeling, judgmental, quasi-judgmental, perceptual, embodied, and motivational theories) in a systematic manner with the help of tools from philosophy of science. Although previous comparisons of theories have provided many useful insights, they tend to be partial. All too often, scholars are guilty of portraying rival theories in a caricatured way so that their own theory emerges as the miracle solution. I believe the time is ripe for a more in-depth and systematic approach. This will be to list phenomena that merit explanation, to describe and compare possible explanations for these phenomena, using a fine-tooth comb, and to examine whether and how these explanations can be tested empirically and whether they are internally consistent. It is not my aim to settle any of the debates in a definitive manner nor to enforce premature consensus. My first hope is that this exercise will lead to the clearing up of misunderstandings among emotion theorists. This may increase consensus in some respects but it may also lead to a sharpening of the real differences. Once real differences are identified, ways of examining them can be developed and targeted efforts can be made to further investigate these differences. This may eventually move the field forward.

The second aim of the book is to propose my own skeptical theory of the phenomena called emotions. Much of my previous theoretical and empirical work on emotions was inspired by appraisal theories. However, several elements contributed to a shift in my thinking and culminated in the recently developed goal-directed theory. One element is my own lab's failures to empirically confirm hypotheses of appraisal theories about the influence of appraisals on action tendencies, especially when appraisals were manipulated experimentally and action tendencies were measured with indirect objective methods. Another element is the range of challenges articulated by existing skeptical theories. A final element is my increasing acquaintance with other literatures such as those on motivation, action, and operant learning. Applying insights from these literatures to the study of emotion combined with a sizable skepticism about popular dual-system models in psychology changed my thinking for good.

The book is structured as follows. The first part ("Introduction") comprises two chapters. Chapter 1 ("Theory Development and Concepts") lays out a meta-theoretical framework composed of insights borrowed from philosophy of science. This framework forms the backbone of the book and is indispensable for understanding the following chapters. Chapter 2 ("Demarcation-Explanation Cycle Applied to Emotion Theories") applies this framework to the emotion domain, providing a blueprint for the analysis of emotion theories discussed in the second part ("Emotion theories one by one"). This part comprises seven chapters. Chapter 3 ("General Precursors") kicks off with Darwin (1872), followed by James (1890b), whose theory is called a feeling theory in philosophy. Chapter 4 ("Evolutionary Theories") covers evolutionary theories in psychology, also known as motivational theories in philosophy. Chapter 5 ("Network Theories") discusses network theories in psychology and kindred theories in philosophy. Chapter 6 ("Stimulus Evaluation Theories") handles stimulus evaluation theories, which include appraisal theories in psychology, and judgmental theories, quasi-judgmental theories, perceptual theories, and embodied theories in philosophy. Chapter 7 ("Response Evaluation Theories") is dedicated to response evaluation theories, in particular my own version of them, the goal-directed theory. Chapter 8 ("Psychological Constructionist Theories") discusses psychological constructionist theories. After having discussed these "personal" emotion theories, I discuss the "social" versions of these theories in Chapter 9 ("Social Theories"). The book closes with a third part and Chapter 10 ("Conclusion"), in which I examine whether some form of integration of the discussed theories is possible.

Each of the seven theory chapters includes a description of important lines of empirical research that have been carried out under the umbrella of the respective theory. The empirical parts are not meant as exhaustive overviews but try to give insight into the ways in which research has tackled the questions identified in the theoretical parts.

The book is interspersed with three boxes that delve deeper into issues that transcend individual theories. Box 2.1 pitches the distinction between stimulus-driven and goal-directed processes, which I consider to be one of the central axes on which theories differ. Box 2.2 organizes empirical research methods. Box 7.1 clarifies the nuts and bolts of emotion regulation.

Acknowledgment

They say writing a book is like getting a tattoo. It stings for a while and once the ink is dry, there is no turning back. Consider this an apology to all authors that I may have unwittingly misrepresented. I am indebted to a number of wise people for giving me opportunities, inspiring me, and trying to keep me sharp: Jan De Houwer, Nico Frijda, Jim Russell, and Klaus Scherer, for all these things at once; Yannick Boddez, Maja Fischer, Eike Buabang, Massimo Köster, Sander Van de Cruys, and Yael Ecker, for making the goal-directed theory come to life also in other areas than emotion; Andrea Scarantino, Stéphane Lemaire, Alan Fridlund, Fabrice Teroni, Kris Goffin, Constant Bonard, and Phoebe Ellsworth, for valuable discussions and answering several last-minute questions over the past year; Keith Oatley for the invitation to write the book and guiding me through the first steps; Brian Parkinson for being exactly the editor that I needed; Cees van Leeuwen, Batja Mesquita, Deb Vansteenwegen, Jessica Moors, and Frank Baeyens for their moral support; my parents, other siblings, other colleagues, and other friends – but above all – my love interest David, and my sons Jascha, Lewi, Ilja, and Sinn, for their generous care and for getting on with their lives when mine stood still. The book is dedicated to my father Simon Moors.

Abbreviations

CNS	central nervous system
PNS	peripheral nervous system
ANS	autonomous nervous system
SNS	somatic nervous system
ES	endocrine system
S	raw stimulus
eS	raw external stimulus
iS	raw internal stimulus
[S]	afferent/stimulus representation
[ecS]	external concrete stimulus representation
[eaS]	external abstract stimulus representation
[icS]	internal concrete stimulus representation
[iaS]	internal abstract stimulus representation
[O/R]	efferent representation
[O^v]	non-behavioral goal
[R]	abstract behavioral goal or behavior representation
[r]	concrete behavioral goal or behavior representation
R	overt response
sR	somatic response
mR	motor response
smR	subtle motor response
cmR	coarse motor response
[R_{as}]	strategy of assimilation
[R_{ac}]	strategy of accommodation
[R_{im}]	strategy of immunization
S–R	stimulus-response
O^v	valued outcome
S:R–O^v	relation between a response and a valued outcome given a certain stimulus

https://doi.org/10.1017/9781107588882.002 Published online by Cambridge University Press

US	unconditioned stimulus
CS	conditioned stimulus
UR	unconditioned response
CR	conditioned response
DS	dynamic systems
SET	stimulus evaluation theory
RET	response evaluation theory
PCT	psychological constructionist theory
SCT	social constructionist theory
FACS	facial action coding system
EMG	electromyography
RT	reaction time
MEP	motor evoked potential
TMS	transcranial magnetic stimulation
IAT	implicit association task
EPT	evaluative priming task

PART I

Introduction

The emotion domain suffers from a proliferation of theories. The traditional way of organizing these theories is to group them into discrete categories. I propose a novel typology, in which theories are placed on several axes. Axes correspond to questions that arise during theory development. The places that theories occupy on these axes correspond to the answers they provide to these questions. Chapter 1 describes theory development as a cycle composed of various stages. It also lays out a number of basic concepts and background assumptions required for understanding the chapters that follow. Chapters 2–9 make use of this cycle to compare a range of emotion theories from psychology and philosophy.

https://doi.org/10.1017/9781107588882.003 Published online by Cambridge University Press

https://doi.org/10.1017/9781107588882.003 Published online by Cambridge University Press

CHAPTER 1

Theory Development and Concepts

Theory development can happen via different paths. Section 1.1 describes one such path: the "demarcation-explanation cycle."[1] This path will turn out to be particularly suitable to describe theory development in the emotion domain. Section 1.2 introduces different types of definitions and ways to evaluate their adequacy. Section 1.3 introduces different types of explanations, and related to this, the notion of levels of analysis. This section also digs deeper into the ingredients of mechanistic explanations such as representations, operations, and operating conditions (related to automaticity). It also briefly pauses to discuss dual-process and dual-system models, different types of rationality, and different usages of the term cognition.

1.1 Demarcation-Explanation Cycle

Scientists develop theories with the aim of explaining, predicting, and/or controlling phenomena (Barnes-Holmes & Hughes, 2013). Although prediction and control are in principle possible without explanation, many agree that explanation is an aim worth pursuing in itself, and that it does have invaluable benefits for prediction and control. "Explanation" is an activity in which an explanandum (i.e., a to-be-explained phenomenon) is linked to an explanans (i.e., an explaining entity or set of entities). To illustrate with a toy example, one type of explanation of the phenomenon of water links it to H_2O. Researchers need to demarcate the explanandum before they can search for an explanans. Rather than being a linear process, however, demarcation and explanation are better understood as alternating activities that can be embedded in a series of cycles.

A first cycle comprises the following four stages (see Figure 1.1(a)). In the first stage, researchers present a provisional demarcation or working definition of the explanandum. If the explanandum is a single entity, the working definition can be a collection of superficial properties.

[1] This path combines elements from Bechtel's (2008) path towards "reconstitution of the phenomenon" with elements from Carnap's (1950) path towards "explication."

https://doi.org/10.1017/9781107588882.004 Published online by Cambridge University Press

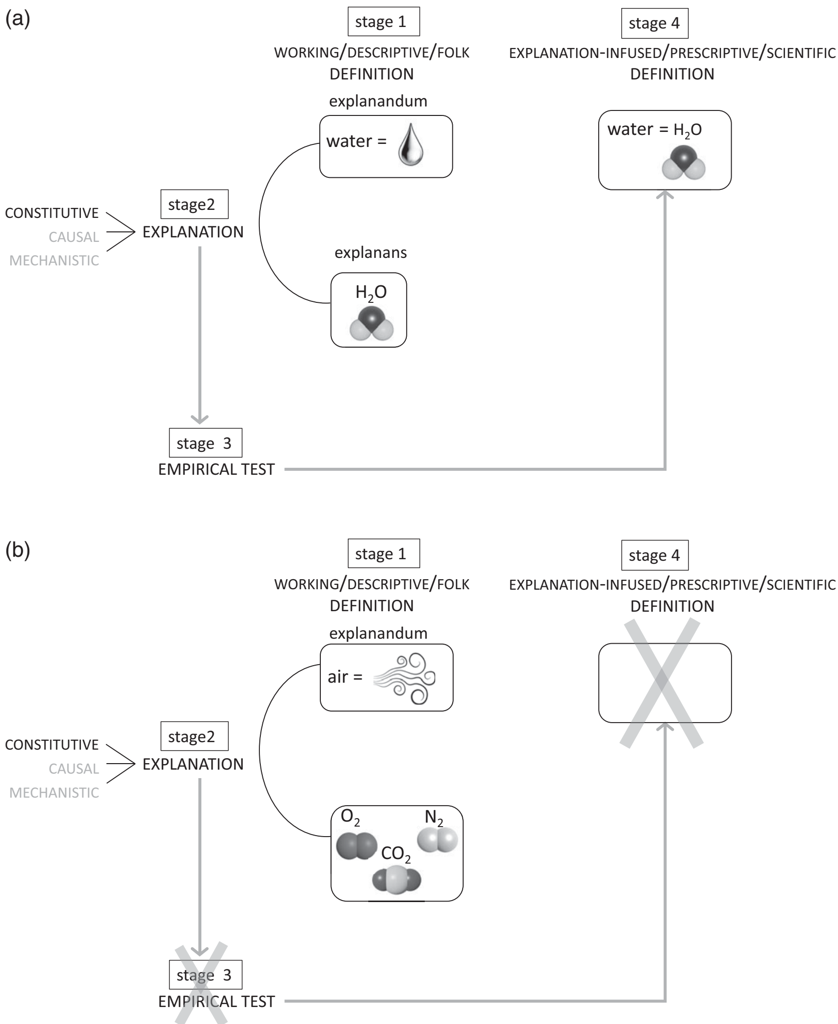


Figure 1.1 Demarcation-explanation cycle: (a) water; (b) air

For instance, water is a transparent, odorless fluid that runs in rivers and falls out of the sky. In the second stage, researchers develop an explanation of some type, in which they link the explanandum to an explanans. In the water example, they discover that the molecular structure of water is H_2O. In the third stage, the explanation is validated by testing it in empirical research. In the water example, researchers take samples of water according to their working definition and they check whether the

https://doi.org/10.1017/9781107588882.004 Published online by Cambridge University Press

molecular structure of these samples is indeed H_2O. If this is sufficiently confirmed, in a fourth stage, the explanans may eventually become part of the definition of the phenomenon, where it replaces the superficial features of the working definition. This definition has now become an explanation-infused definition.[2] Instead of demarcating water as a clear, odorless fluid, it is now equated with H_2O. From now on, water defined as H_2O may figure in new explananda such as the phenomenon that certain substances (e.g., sugar) dissolve in water whereas others (e.g., oil) do not. Note that this new explanandum is no longer a single entity (water), but a regularity between entities (i.e., the mixing of water with other substances and the resulting substance). When new explanations are developed and tested, a scientific theory of water gradually develops.

The entities in science can be understood as sets that have members. This allows us to portray the cycle as follows. Theorists take the working definition of a set as the starting point and develop an explanation in the hope that this will yield a common denominator for the members in this set. If the quest for a common denominator is successful, it forms the basis for the explanation-infused definition of the set.

The demarcation-explanation cycle not only describes (one path towards) theory development in the natural sciences but also in the behavioral and mind sciences, in which all kinds of behaviors and experiences can be targets of explanation. It is especially suitable to describe theory development in the emotion domain, as this domain is still in the stage of figuring out what emotions are. Before we can get our teeth into the emotions, we need to elaborate on the present framework. The following sections discuss types of definitions, types of explanations, and related concepts.

1.2 Types of Definitions and Adequacy

Parallel to what I said about "explanation," "definition" can be thought of as an activity that links a definiendum (i.e., to-be-defined entity) to a definiens (i.e., defining expression) in an identity relation. The demarcation-explanation cycle contains two types of definitions: a *working definition* in Stage 1 and an *explanation-infused definition* in

[2] This corresponds to Bechtel's (2008) "reconstitution of the phenomenon." Several other authors have accepted explanantia at the heart of definitions (e.g., Eilan, 1992; Gordon, 1974; Green, 1992; Reisenzein, 2012; Reisenzein & Junge, 2012; Reisenzein & Schönpflug, 1992; Siemer, 2008). A well-known example is that of "sunburn defined as inflammation of the skin caused by overexposure to the sun" (Gordon, 1978). Note that the credo to avoid conflating explanandum with explanans, although violated in the fourth stage, remains important for the first three stages.

Stage 4. The working definition is often a *descriptive* or *folk* definition, that is, a description of the way in which laypeople understand an entity. The explanation-infused definition is a *prescriptive* or *scientific* definition, that is, a definition in which scientists prescribe how the entity should be understood in scientific discourse (Widen & Russell, 2010).

Another type of distinction pertains to different formats of definitions (J. Lyons, 1977, p. 158). *Intensional* definitions specify the conditions or criteria for a member to belong to a set (i.e., the intension): a single condition that is both necessary and sufficient or a conjunction of necessary conditions that are together sufficient. The conditions are often expressed as properties (Orilia & Paolini Paoletti, 2020). For instance, the set of bachelors has the properties "men" and "unmarried." Note that intensional definitions often do not list all the necessary conditions of a set, but only those that help demarcate the set from specific other sets. The non-mentioned necessary conditions either are implicated in some of the mentioned ones, or they are implicitly assumed. In the bachelor example, the condition "men" implies a bunch of conditions that make the existence of men possible (e.g., that there is a world, and a galaxy) and a bunch of implicit conditions (e.g., that the men are human and that they are adults not babies).

Extensional definitions list the members within a set (i.e., the extension). Intensional and extensional definitions are reciprocal: A set with the intension "all integers between 2 and 7" fixes the extension to {3, 4, 5, 6}. Conversely, a set with the extension {3, 4, 5, 6} leaves room for several intensions, of which a simple one is "integers between 2 and 7" and a more complex one could be "integers that subtract 7, 6, 5, 4, and 3 from 10." A complete extensional definition is only possible for finite sets. For infinite sets, the most one can do is give a sampling definition in which a few prototypical members are listed.

A special type of extensional definitions, which I call divisio definitions, specify the subsets within a set.[3] Divisio definitions not only help to demarcate a set, similar to intensional and extensional definitions, but also to organize the variety within a set. Sets can often be partitioned in more than one way. The set {3, 4, 5, 6} can be split on a low level into subsets that correspond to each of the members ({3},{4},{5},{6}). On a higher level, it can be split into the broad subsets of small ({3, 4}) and large numbers ({5, 6}), but also into the broad subsets of even ({2, 4}) and odd ({3, 5}) numbers. The way in which theorists partition a set thus involves an element of choice.

[3] The term was originally used by Cicero (*Topics*, V. 28; cited in Ierodiakonou, 1993).

The sets, subsets, and members that science is interested in qualify as *types* (i.e., abstract entities) that can be exemplified or instantiated by *tokens* (i.e., concrete entities in space-time; Wetzel, 2018). It could be argued that when members are understood as types, they are in fact subsets of tokens. For this reason, I will continue to talk about "divisio definitions" instead of "extensional definitions."

In principle, both working definitions and scientific definitions can take on an intensional format (i.e., a list of properties) and a divisio format (i.e., a list of subsets). While scientific definitions strive for completeness and precision, working definitions are first approximations. This is why working definitions will often be partial or incomplete.

The scientific definitions in Stage 4 can be evaluated in terms of their adequacy using meta-criteria such as similarity, fruitfulness, and simplicity, to name the most important ones (Carnap, 1950). I first discuss what these criteria entail in the case of intensional definitions before turning to divisio definitions.

In the case of intensional definitions, the similarity meta-criterion entails that the extension of the scientific definition bears sufficient overlap with the extension of the working definition. This means that the scientific definition should tie in with common sense (Green, 1992; Scarantino, 2012b). For instance, the members of the scientific set "water" should show substantial overlap with members of the folk set "water."

The fruitfulness meta-criterion requires that a set allows for scientific extrapolation, that is, the generalization of discoveries about one exemplar to other exemplars in the set (Griffiths, 2004a; Scarantino, 2012b). Scientific extrapolation is only possible when the set is homogeneous in a non-superficial way. Exemplars must share a deep similarity such as a common constitution, a common causal mechanism, or even a common function. If the set is too heterogeneous, not enough generalizations can be made from one exemplar to another. According to this criterion, "diamond" is an adequate set because all its members are constituted by one mineral whereas "jade" is inadequate because its members can be constituted by two different minerals: jadeite and nephrite. Discoveries for jadeite may not generalize to nephrite.

The meta-criterion of simplicity or parsimony, finally, requires that the conditions in a scientific definition be few. Demarcating the set of water using H_2O as the only condition is simple. In fact, the simplicity meta-criterion is hard to separate from the fruitfulness meta-criterion. The ideal is to find a simple common ground among the members of a set, not a complex disjunction of several partially common grounds as this would again hamper extrapolation. This can be captured in the term "fruitfulness-annex-simplicity meta-criterion" but for ease of communication

https://doi.org/10.1017/9781107588882.004 Published online by Cambridge University Press

I will continue to use the term fruitfulness and treat the simplicity meta-criterion as part of it.

Theorists must strike a balance between similarity and fruitfulness even though there are no guidelines for how to establish their relative weights (Swartz, 1997). If the folk set is heterogeneous at the outset, a trade-off between these meta-criteria is inevitable. Maximizing similarity comes at the cost of fruitfulness and maximizing fruitfulness comes at the cost of similarity. Take again the folk set "jade," which is composed of the minerals of jadeite and nephrite. If the scientific definition keeps both minerals on board, this would ensure maximal similarity at the expense of fruitfulness. If the scientific definition keeps only one mineral on board and throws out the other, this would ensure maximal fruitfulness at the expense of similarity. In between these extreme forms of prioritizing similarity or fruitfulness, more subtle forms can be identified.

One moderate form of prioritizing similarity over fruitfulness consists in giving up the quest for a classic intensional definition (with one condition that is both necessary and sufficient or a conjunction of necessary conditions that are jointly sufficient) and turning instead to a cluster-type definition. Simply put, a cluster-type definition is a weak form of intensional definition in which the status of the conditions is relaxed from necessary to typical (Boyd, 1999, 2010; Searle, 1958; Wittgenstein, 1953). For instance, the conditions used to demarcate the set of lemons are typical instead of necessary: oval (some lemons are round), yellow (some lemons are green), and acid (some lemons are bitter). Members belong to the set when they show more or less resemblance with a prototype (Rosch, 1999), understood as an average of all members of the set (Posner & Keele, 1968) or a salient member (Kahneman & Miller, 1986; see Russell, 1991). More formally, cluster-type definitions can be expressed as a disjunction of sets of jointly sufficient properties (Longworth & Scarantino, 2010). The set of lemons has the properties "oval, yellow, and sour" or "oval, yellow, and bitter," or "round, yellow, and sour," and so on. Thus, cluster-type definitions still count as intensional definitions but they are more complex than their classic counterparts and they may hamper smooth extrapolation. Cluster sets are common in science. In addition to lemons, other popular examples are biological species, games, art, and mental disorders. Proponents of this approach argue that the cost for fruitfulness, although in principle increased, remains low in practice. The fact that a strict intensional definition has not been found for lemons does not bother people who need to buy lemons to make lemonade. If it tastes and smells like lemon, it will do.

Moderate forms of prioritizing fruitfulness over similarity, on the other hand, consist in trimming the folk set to a smaller or larger degree.

For instance, when the folk set "fish" turned out to contain not just cold-blooded vertebrates that have gills throughout life (like guppies and sharks) but also a small number of warm-blooded species that breathe through lungs (like dolphins and whales), the latter were trimmed off from the scientific set of fish. The case discussed above in which nephrite is thrown out of the set of jade is more radical in that much more from the initial set is lost. Another solution to handle heterogeneity in this case would be to split the folk set into two equally valid subsets. In this way, more can be rescued from the folk set than just a single subset.

The most radical form of prioritizing fruitfulness over similarity consists in the elimination of the set altogether. If the quest for a common ground turns out to be unsuccessful, scientists may conclude that the set cannot reach a scientific status. Take the example of air (see Figure 1.1 (b)).[4] Just like water, air was once thought to be a fundamental building block of nature. The working definition of air contained superficial features such as that it is a transparent, odorless gas that fills our lungs and the sky. Scientists discovered that all members of the set of air are composed of varying molecules such as oxygen, nitrogen, and carbon dioxide. The lack of a stable common denominator led them to conclude that air is not an adequate scientific set (at least not in chemistry). The question of whether the folk set "emotion" is more like "water," "fish," "jade," or "air" is one that I will be considering later in this book.

Turning to the case of divisio definitions then, the similarity meta-criterion entails that the scientific definition carves up the set in a similar way to the working definition. The fruitfulness meta-criterion stipulates that subsets should be created on the basis of simple criteria that allow for extrapolation between the members of each subset. For instance, a scientific divisio definition with subsets solid, fluid, and gasiform H_2O is similar to the working divisio definition with subsets ice, running water, and steam. The partitioning is fruitful because it is based simply on temperature differences and allows extrapolation within each of the resulting subsets.

Once a set has reached the status of a scientific set, it can be called a scientific or investigative kind (Brigandt, 2003; Griffiths, 2004a). Some scientific kinds are called natural kinds. A natural kind not only requires a common denominator that allows for extrapolation, but also that the common denominator be natural, as is captured in the aphorism that natural kinds carve nature at its joints. Natural kinds are typically contrasted with arbitrary or conventional kinds, in which the members are held together by a common feature that is not natural but resides, at least

[4] I owe this example to Jim Russell.

in part, in the minds of the people making the classification. Examples are the set of weeds and the set of pet animals. The differences between weeds and cultivated plants or between pets and other animals cannot be easily captured in natural terms. Weeds and pets can nevertheless be considered as investigative kinds in certain scientific disciplines such as domestication science (Griffiths, 2004a). The question of whether emotion is a natural kind or a conventional kind has gathered some interest among emotion theorists. It is good to realize, however, that the debate about emotions as natural kinds is complicated by the fact that some scholars have stretched the meaning of natural kinds and use it as synonymous with scientific kinds. Such an extension of meaning is based on the ideas that (a) "natural" is not synonymous with "material" but can also be "mental" and (b) "natural" does not need to equate with a "natural essence" (as per a classic intensional definition) but can also include a "cluster of natural features" (as per a cluster-type intensional definition) (for discussions see Barrett, 2006a; Boyd, 1999; Griffiths, 2004a; Scarantino, 2012b; Scarantino & Griffiths, 2011).

1.3 Types of Explanations and Levels of Analysis

Explanations come in various types. Three types will turn out to be relevant for present purposes: constitutive explanations, causal explanations, and mechanistic explanations (see Figure 1.2). I illustrate these types with the hangover example. A *constitutive explanation* specifies the constituents or components of a phenomenon. For instance, a hangover is comprised of a headache, nausea, and a dry mouth. This constitutive explanation is not yet a definition because the presence of these components is not sufficient to demarcate hangovers from other phenomena. Indeed, a headache, nausea, and a dry mouth may also occur when someone has the flu. To demarcate hangovers from viral infections we probably need a *causal explanation*, in which a hangover is linked to excessive drinking the night before. In such an explanation, a phenomenon is explained by pointing at an antecedent cause. A *mechanistic explanation* specifies the detailed steps of the mechanism that mediates between the cause and the explanandum. Drinking allows alcohol to flow into the bloodstream, part of which is transformed by the liver into acetaldehyde (via a mechanism called alcohol dehydrogenase) and further into acetate (via a mechanism called acetyl dehydrogenase). This causes the contraction of blood vessels in the brain, ending up in a headache, and so on.

The nature of these three types of explanations is best understood if we place them within a levels-of-analysis framework. Levels can be distinguished on the basis of several criteria (e.g., scientific disciplines, strata

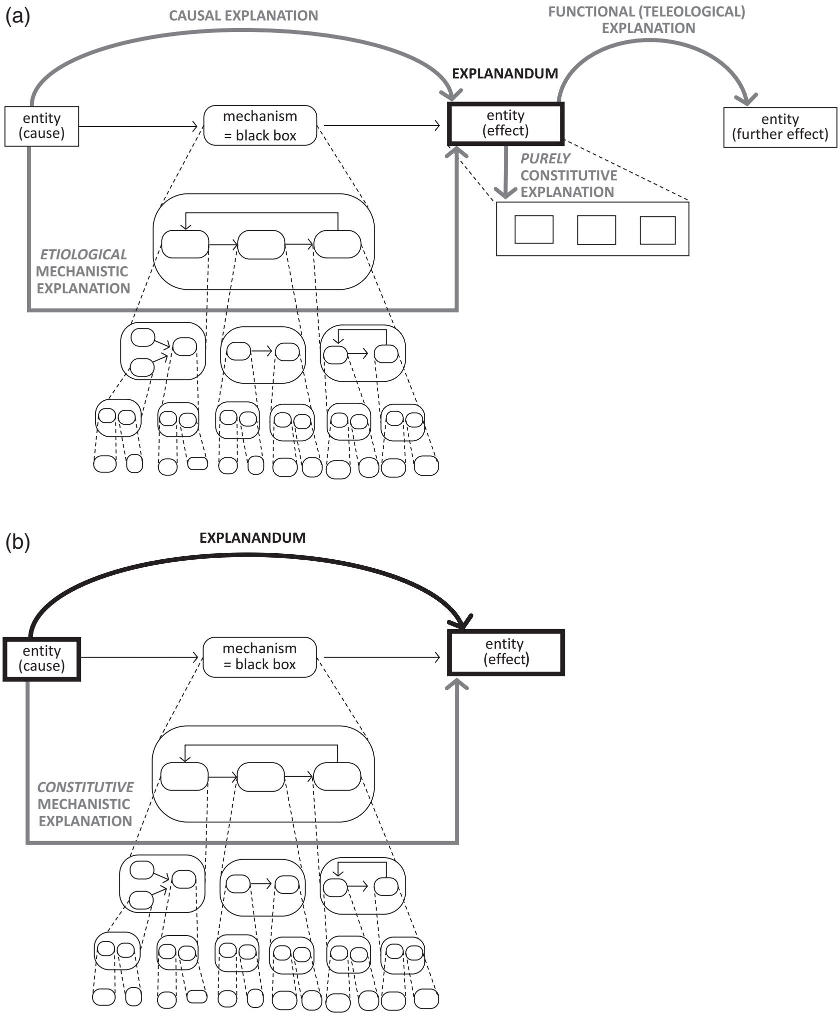


Figure 1.2 Types of explanations: (a) explanandum is an entity; (b) explanandum is a causal relation between entities

across nature, mere aggregates, size, and complexity; see Bechtel, 2008; Craver, 2015). I follow the proposal of mechanistic philosophers of science (e.g., Craver, 2015) to distinguish levels on the basis of mereological (i.e., part–whole) relationships: Level A is lower than level B if the entities at level A are parts of the entities at level B.

In a causal explanation, the explanantia are causal factors situated at the same level of analysis as the explanandum (Craver & Bechtel, 2007, 2013). In constitutive and mechanistic explanations, the explanantia are parts. Constitutive explanations specify the parts of the explanandum, whereas mechanistic explanations specify the parts of the mechanism that mediates between the cause and the explanandum. Thus, mechanistic explanations start from and build on causal explanations in that they specify the mechanisms at a lower level of analysis that mediate between the causal entities (specified in the causal explanation) and the explanandum (Craver, 2013).

In the case in which the explanandum is itself a causal relation between entities (and not a simple entity), explanations that specify the parts of the mechanism mediating between the two entities count as constitutive explanations, strictly speaking. Craver and Tabery (2019; Salmon, 1984) treat the latter type of explanation as a subform of mechanistic explanations, calling them *constitutive mechanistic explanations* (Figure 1.2(b)), next to the subform of *etiological mechanistic explanations* (i.e., which correspond to what I called mechanistic explanations simpliciter so far; Figure 1.2(a)). This leads to an extension of the taxonomy of explanations into four types: purely constitutive ones, causal ones, etiological mechanistic ones, and constitutive mechanistic ones. The first three are suitable when the explanandum is an entity; the fourth is suitable when the explanandum is a causal relation between entities.

Mechanistic explanations not only specify *parts* but also *activities* that spell out the causal relations between parts. The parts in mechanistic explanations are not like marbles in a bag, but hang together in a causal fashion.[5,6] Minimal descriptions of activities only mention that they are causal; more elaborate descriptions specify that the causal relations are also excitatory or inhibitory, for instance, or that they involve certain types of computations.

In addition to specifying parts and activities, mechanistic explanations also specify the way in which different parts and activities are *organized*. An organization can be linear, describing the linear transition from input to output, but it can also be cyclical, in which case the output of a previous cycle forms the input to a new cycle. In sum, mechanistic

[5] Activities figure in etiological as well as constitutive mechanistic explanations. In purely constitutive explanations, on the other hand, information about activities relating to parts is optional. The parts of an atom (neutron, electron, proton), for instance, are working parts, whereas the parts of a marble statue (head, rump, limbs) are not. Purely constitutive explanations that do report activities are nearly indistinguishable from constitutive mechanistic explanations.

[6] Activities have also been characterized as the manifestations of dispositions (also called powers or capacities; Piccinini & Craver, 2011). Some authors have argued that the task of science is not to uncover the activities themselves but rather these dispositions (Manicas & Secord, 1983).

https://doi.org/10.1017/9781107588882.004 Published online by Cambridge University Press

explanations not only look downwards (specifying parts; i.e., decomposition) and sideways (specifying causal activities among parts), but also upwards (specifying the organization of parts into wholes; i.e., recomposition) (e.g., Bechtel, 2008; S. Bem & Looren de Jong, 2013).

Parts are presented as the structural aspect of mechanistic explanations whereas activities are presented as the functional aspect of these explanations (Bechtel, 2005). Parts are structural in that they have a location, shape, and orientation, even if they resist a neat description in these terms (Piccinini & Craver, 2011). Activities are functional in that they are specified in terms of what they do or accomplish, that is, the output parts they produce given a certain input part. In the hangover example, the mechanism of alcohol dehydrogenase takes ethanol as its input and produces acetaldehyde as its output, after which the latter substance forms again the input of the mechanism of acetaldehyde dehydrogenase producing acetate as its output. Explanations that specify activities but leave out structural details are dubbed *functional analyses*. Instead of contrasting the latter with mechanistic explanations, Piccinini and Craver (2011; Craver & Kaplan, 2020) have portrayed them as elliptical or incomplete sketches of mechanistic explanations that may form the first steps towards a complete mechanistic explanation.

In mechanistic explanations and functional analyses, the output of the activities is the explanandum. If the consequences of an explanandum are envisaged, however, we speak of a *functional – in the sense of teleological – explanation* (Mundale & Bechtel, 1996). Functional explanations in psychology and biology, for instance, specify the role that the explanandum plays for an organism's long-term goals or survival or for the species or society as a whole. In the hangover example, it might be speculated that hangovers help to avoid alcohol abuse in the future. Hangovers could alternatively be considered as purely epiphenomenal, defying a functional explanation. Functional explanations can be added to the taxonomy as a fifth type of explanation (see Figure 1.2(a)).

The mereological (i.e., part–whole) view of levels of analysis presented thus far is still compatible with a rough division of levels into three broad super-levels inspired by the levels pioneered by Marr (1982) and others (e.g., Bechtel & Shagrir, 2015): an observable super-level, a mental super-level, and a brain super-level. These levels correspond to strata that are relevant for behavioral and mind sciences. At the *observable super-level*, a system produces an observable output (effect) in response to an observable input (cause). The transition from input to output can be called a process, and is mediated by the mechanism as a whole. At this level, a process is described in terms of its observable input, its observable output, and the relation between the two. Typically, the observable input is called the stimulus, and the observable output is a behavioral or physiological response. The mechanism between input and output is treated here as a black box. At the *mental super-level*, this mechanism is decomposed into submechanisms, which can themselves be described in terms of their inputs, outputs, and interrelations. The intermediate inputs and outputs, which are not observable, are called mental *representations* and the relations or activities among them are called mental *operations* (see more below). Each of the submechanisms at the mental super-level may be decomposed further into even finer-grained submechanisms until, at the final stages of decomposition, they correspond to brain processes situated at the *brain super-level*. In other words, the big black box is recursively decomposed into little black boxes all the way down (i.e., heuristic identity relation between levels; Bechtel, 2008). The three super-levels mentioned here are all situated in the individual. In the social sciences, a fourth, *social super-level* can be proposed, where regular patterns of interactions between individuals are specified (Bunge, 2004).

There is debate about how to understand (a) the relations between (all kinds of) mereological levels, which is an ontological question, and (b) the relations between the scientific theories that occupy the four super-levels, which is an epistemological question. Regarding the ontological question, mechanistic philosophers see inter-level relations as constitutive and therefore identity relations. If, in addition, a view of causation is endorsed in which causes should be separate from and precede their effects, it follows that causal relations are strictly intra-level (Bechtel, 2008, p. 153; Craver & Bechtel, 2007, 2013; Crisp & Warfield, 2001; Romero, 2015; but see Baumgartner & Gebharter, 2016; Krickel, 2017; Leuridan, 2012; Ylikoski, 2013). In line with this view, apparent cases of top-down and bottom-up causation can be recast in terms of mechanistic mediation, that is, hybrids of constitutive and causal relations (Craver & Bechtel, 2007, 2013).

Regarding the epistemological question, approaches to inter-theory relations range between (a) classic (i.e., smooth) reductionism in which higher-level theories (e.g., mental theories) are explained away by lower-level theories (e.g., neuroscientific theories; Oppenheim & Putnam, 1958); (b) new-wave (i.e., "bumpy" and "patchy") reductionism in which higher-level and lower-level theories constrain and inspire each other (Churchland & Churchland, 1992; see Mundale & Bechtel, 1996); (c) the "mosaic unity of sciences" view, in which each science contributes in a non-reductive but still interdependent way (Craver, 2007); and (d) explanatory pluralism that lets many flowers bloom and grants each level full explanatory autonomy (see S. Bem & Looren de Jong, 2013; McCauley, 1996; McCauley & Bechtel, 2001). Mechanistic philosophers profess non-reductionist relations among levels based on the argument that mechanistic explanations span at least two levels (e.g., Bechtel, 2008; McCauley & Bechtel, 2001). Critics are unassuaged by this argument, maintaining that the identity relation between levels inevitably invites some form of reductionism (e.g., Fazekas & Kertész, 2011; Glauer, 2012). Without digging further into the details of this debate, I believe it is useful to separate ontological issues from epistemological ones. According to mechanistic philosophers, who understand the relations between levels in a mereological sense, different levels of analyses do not house different realities, but different ways to parse and look at the same reality. Thus, despite the fact that they assume identity relations between the entities of different levels (i.e., ontological issue), the theories situated at each level can still be granted explanatory individuality, whether in a mosaic with, or independent of, theories on other levels (i.e., epistemological issue). All this is to say that the mechanistic approach adopted in this book does not imply epistemological reductionism.

To take stock, causal explanations are intra-level, with the explanantia situated on the same level as the explanandum. Constitutive and mechanistic explanations cross different levels of analysis. If the individual is taken as the unit of analysis, mechanistic explanations can reside at the mental super-level (i.e., mental mechanistic explanations) or the brain super-level (i.e., neural mechanistic explanations). Let us now consider the ingredients of these two types of explanations in more detail.

1.3.1 Mental Mechanistic Explanations

So far, we learned that mechanisms are made up of parts and activities, and that in the case of mental mechanisms, the parts are representations and the activities are operations (Bechtel, 2008). Mechanisms, moreover, vary in the conditions they require to operate. In the following sections, I will clarify my usage of each of these notions – representations (Section 1.3.1.1), operations (Section 1.3.1.2), and operating conditions (Section 1.3.1.3) – and propose ways in which to organize the variety in each (see Figure 1.3). As it turns out, researchers tend to dichotomize this variety. This has tricked them into binary thinking and the formation of dual-process and dual-system models (Section 1.3.3).

1.3.1.1 Representations

Representations have been invoked to cater for the feature of most mental processes that they are directed at something beyond themselves, a feature that philosophers call Intentionality[7] or aboutness (Brentano,

[7] I capitalize the term to indicate the difference with intentional in the ordinary sense, following Searle (1983). The minimal meaning of the term intentional is "directed" (Jacob, 2019). In philosophical usage, *Intentional* refers to the property of a mental state by which it is directed at something beyond itself (Brentano, 1874). In ordinary usage, *intentional* refers to the fact that an agent is directed to (i.e., willing to engage in) an (overt or covert) act (Moors & De Houwer, 2006a).

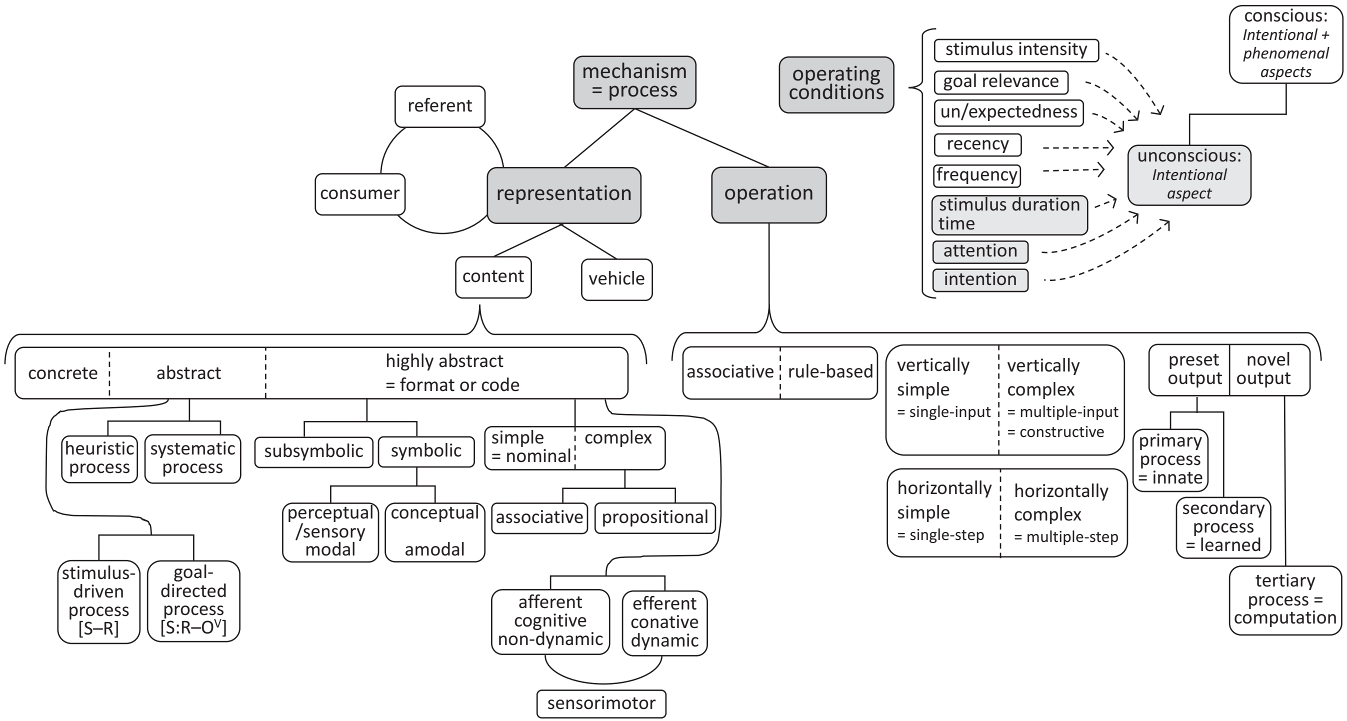


Figure 1.3 Ingredients of mental mechanistic explanations
Note: Full/dotted partitions indicate categorical/gradual distinctions.

1874). One way in which a system can be directed towards something is by forming a representation of it. The notion of representation can be unpacked as being part of a three-place relation, in which it is linked to a referent and a consumer. The referent is an object in the organism's external or internal world. The representation makes the referent available to the consumer or interpretant (Millikan, 1993; Peirce, 1940).

The representation itself is composed of a vehicle and a content. The vehicle is typically identified with the physical brain activity underlying activation of the representation. The content is the referent as represented. This content can be organized along various dimensions. One important dimension is the degree of abstraction. Concrete content can be virtually anything. At the high end of the spectrum, content is so abstract that it slides into what some authors have called the format of representations or the representational code. I will list a number of popular contrasts that have been proposed regarding format: symbolic vs. subsymbolic, conceptual vs. perceptual, simple vs. complex, associative vs. propositional, and afferent vs. efferent. After that, I discuss two dichotomies that concern specific types of representational content: heuristic vs. systematic and stimulus-driven vs. goal-directed.

Let us start with the contrast between symbolic and subsymbolic representations. A symbolic representation is one in which the content is a meaningful entity such as an object. In subsymbolic representations, the meaningful entity is distributed across representations that each refer to a separate feature of the entity. Symbolic representations split further into conceptual and perceptual representations. Conceptual representations are verbal-like or word-like. Perceptual or sensory representations are image-like or picture-like, but they can in principle also be sound-like, smell-like, taste-like, or touch-like.[8] The analogies with words and pictures are not literal – as if there are words and pictures in the head – but rather structural. Perceptual representations have a higher level of detail than conceptual ones and are more vivid, but at the same time, they are schematic in the sense that they allow a form of indeterminacy that actual pictures do not (Barsalou, 1999). For example, the perceptual representation of a tiger has stripes but the number of stripes may remain

[8] The distinction between sensation and perception is not a sharp one. Sensation is used more to refer to the detection of low-level stimulus features by sensory receptors whereas perception is used more to refer to the identification of stimuli based on a combination of sensory stimulus features. Note that the number of senses need not be limited to the "Aristotelian five" listed here (see Macpherson, 2011). Sensation can be organized into (a) external sensation (or exteroception), including vision, hearing, smell, taste, touch, and temperature, and (b) internal sensation, including interoception (internal body state) and proprioception (position of body parts in space, and kinesthesia or movement of body parts).

indeterminate. Another contrast linked to the conceptual-perceptual contrast is that between amodal and modal representations. Conceptual representations are seen as amodal whereas perceptual representations come in a specific modality, linked to a specific sensory channel (e.g., visual, auditory, olfactory, gustatory, and tactile) (e.g., Barsalou, 2008; Barsalou et al., 2003; Garcia-Marques & Ferreira, 2011).

Another contrast has to do with the complexity of the content of representations: Simple or nominal representations hold a single entity (e.g., a cat) whereas complex representations house multiple entities (e.g., a cat and a mat) and relations among these entities. Relations can be of two kinds. Unqualified relations are called "associations" (e.g., the cat and the mat are related but it is not specified how). Qualified relations are called "propositions" (e.g., spatial relations: the cat is on the mat; temporal relations: night follows day; causal relations: allergens cause allergic reactions) (De Houwer, 2014; Moors, 2014c). These are states of affairs that can be expressed in a that-clause (e.g., that the cat is on the mat). This is why propositional representations are often characterized as sentence-like or sentential. Here again, the analogy with sentences is more structural than literal. Propositions are composed of elements that can be recombined (i.e., compositionality and productivity; Fodor, 1981) but they need not be sentences in the head made up of words.

A further contrast is that between afferent (sensory-perceptual/cognitive) and efferent (motor/conative) representations. This contrast can be expressed in terms of a different direction of fit (Searle, 1983): An afferent representation (e.g., perception or belief) has a world-to-mind direction of fit. This means that it is fitting if its content fits with the world, that is, if it is accurate. An efferent representation (e.g., goal) has a mind-to-world direction of fit. This means that it is fitting if the world fits with its content, that is, if it is satisfied. The contrast between afferent and efferent representations also squares neatly with the contrast made in motivation psychology between pure knowledge representations, which are seen as non-dynamic representations, and goals, which are seen as dynamic representations. A dynamic representation typically leads to behavior and its activation accumulates over time (until the goal is fulfilled or overridden by stronger goals), even in the face of obstacles (Bargh & Barndollar, 1996; Bargh et al., 2010). A non-dynamic representation does not lead to behavior and its activation diminishes over time. To illustrate, activation of the goal to have an apple (i.e., the representation of the apple with a value attached to it) leads to behavior to get the apple and does not diminish but rather increases until the apple is obtained or until a more important goal intervenes. The mere thought of an apple (activated by the instruction to think of an apple or a priming procedure in which a picture of an apple is shown), on the other hand, creates a spike in the activation of the representation of an apple, that is, an initial increase that gradually diminishes over time. Some authors have also voted for the existence of sensorimotor representations: combined afferent and efferent representations, also called "embodied" (Barsalou, 2008) or "pushmi-pullyu" representations (Millikan, 1995). These are representations of stimuli that not only contain perceptual features but also information about what can be done with these stimuli, so-called "affordances" (i.e., a term borrowed from Gibson, 1979, but put in a representational jacket here, see Scarantino, 2003).[9]

The distinction between afferent and efferent representations seems to dissolve when the content of the representation is a behavior. This idea stems from James's (1890b, p. 522) ideomotor hypothesis, which states that when an action is carried out, the action command gets bound up with its sensory effects, resulting in bidirectional action–effect links. Once these links are in place, execution of the action conjures up its sensory effects ([R→E]), and perceiving or thinking about the sensory effects is sufficient to put the action in motion ([E→R], i.e., ideo-motor). Thus, for a behavior to occur, the mere thought of the behavior in terms of its immediate outcome is sufficient for it to become executed. No extra fiat is required. Instead, an extra goal to suppress the behavior is required if the person does *not* want to execute the behavior. This may explain why some people avoid standing on the edge of a cliff. If James (1890b) is right, the mere thought of jumping should cause one to jump unless it is suppressed by the goal not to jump. Thus, if the person thinks of jumping but the suppression is temporarily lowered or lifted (e.g., because of inattentiveness), jumping may become a real risk. A similar idea is voiced in W. Prinz's (1990, 1997) common-coding hypothesis, which states that a common representation or code is used to perceive a behavior and to prepare for it. This hypothesis is supported by research showing that the same brain activity (in mirror neurons) occurs when people are instructed to watch someone else carry out a movement as when they are asked to prepare making this movement themselves (e.g., Fadiga et al., 2000; see Hommel et al., 2001).

Let me close with two popular dichotomies that are based on particular types of representational content. A first content dichotomy – central in theories on reasoning and persuasion – is that between systematic and heuristic information processing. An example of systematic information

[9] Gibson (1979) held a non-representationalist view, which places affordances in external objects and not in the minds of agents. Scarantino (2003) and others have taken a more liberal interpretation in which they allow affordances to figure in the content of representations.

are persuasive arguments of a speaker; an example of heuristic information is the attractiveness of the speaker (Chaiken et al., 1989).

A second content dichotomy – central in behavior theories – is that between stimulus-driven and goal-directed processes (Heyes & Dickinson, 1990), two candidate mechanisms of behavior causation. In a stimulus-driven process, behavior is caused by the activation of a representation whose content is the association between the stimulus and a response or behavior ([S–R]). In a goal-directed process, behavior is caused by a representation that contains information about the outcome of one or more response options given a certain stimulus ([S:R–O^v]), more precisely, information about the value of these outcomes and about their expectancy (i.e., the probability that these outcomes will occur). As this dichotomy will be a central principle for organizing emotion theories in this book, it will be discussed in more detail in Box 2.1.

Let us now turn to the consumer of the representation. The consumer is another part of the system, typically another mechanism, that takes this representation as its input. In philosophy, however, a homuncular interpretation of the consumer proves hard to shake off. There, the consumer is said to have an Attitude[10] towards the content of a (usually propositional) representation. Attitudes vary in mode, with the most common ones being an Attitude of belief and an Attitude of desire. Believing that the train is late is different from desiring that the train is late. To believe it is to judge it as true, to desire it is to want it to come true. A "belief" is the combination of a (propositional) representation and an Attitude of belief. A "desire" is the combination of a (propositional) representation and an Attitude of desire. A belief has a mind-to-world direction of fit (it fits if its content fits with the world); a desire has a world-to-mind direction of fit (it fits if the world fits with its content). It may be noted that beliefs and desires in philosophy show overlap with afferent and efferent representations in psychology. The difference is that in beliefs and desires, the direction of fit is located in the Attitude towards a representation whereas in afferent and efferent representations, it resides in the format of the representation.

So far, a realist picture of mental representations has been drawn, with their vehicle corresponding to actual brain activity, although the precise mapping between representations and brain activity has been left unspecified. However, representations can also be understood in purely

[10] I capitalize Attitude, used here in the philosophical sense, to mark the distinction with attitude in psychology, where it refers to the liking or preference of a person for a certain object, often understood on the mental super-level as the association between an object and a valence label (e.g., apple – positive; Greenwald et al., 2002; but see De Houwer, Gawronski, et al., 2013).

functional terms as entities that help explain variable input–output relations, without making any commitment regarding the ontological status of these representations. If a response to the same stimulus varies across occasions, it makes sense to posit an intervening entity such as a representation (Bermúdez, 1995; Fodor, 1981; Moors, 2014c). Representations may be nothing but metaphors, as Skinner (1945, 1977) argued, but in this capacity, they do still play an important heuristic role (De Houwer, Barnes-Holmes, & Moors, 2013).

1.3.1.2 Operations

Operations are the activities carried out on representations. Examples of types of operations cited in the literature are associative and rule-based ones (S. A. Sloman, 1996). An associative operation is the activation of an association between at least two representations. A rule-based operation is the application of an abstract rule to representations. To make the distinction intuitive, imagine a person ordering two lemonades at the counter and trying to figure out how much to pay. To solve the problem, she can engage in an associative operation in which she remembers the price that she paid last time. She can also engage in a rule-based operation in which she applies the rule "multiply the price of one lemonade by the number of lemonades ordered." Although intuitive at first sight, the distinction between associative and rule-based operations has turned out to be fairly elusive (Hahn & Chater, 1998; Moors, 2014c). Perhaps it can best be characterized in terms of degrees of abstraction of the representations, rather than as different types of operations, with representations in associative "operations" situated at the more concrete end of the spectrum and those in rule-based "operations" at the more abstract end (for an extensive justification, see Moors, 2014c).

Operations can also be classified with regard to their complexity. It is worth distinguishing between a vertical and horizontal type of complexity. Vertical complexity refers to the number of inputs that an operation integrates simultaneously: Single-input operations take a single input to produce their output; multiple-input operations, also called constructive operations, take two or more inputs to produce their output (Moors, 2010b). The number of inputs can be regarded as independent of the types of operations involved. Indeed, both associative and rule-based operations can be single-input or multiple-input (Moors, 2014c).

Horizontal complexity refers to the number of sequential steps that must be carried out to arrive at an output. Some operations are single-step, others are multiple-step (Logan, 1988). Again, the types of operations involved in the steps is open. They can be rule-based or associative. In the psychology of language, a reduction of steps is called chunking or entrenchment (Hartsuiker & Moors, 2017). In computer science and

https://doi.org/10.1017/9781107588882.004 Published online by Cambridge University Press

artificial intelligence, a reduction of steps or of inputs is known as compilation and the reverse movement as decompilation (A. Sloman & Croucher, 1981).

Another distinction has to do with whether the output of the operation was preset or is novel (e.g., Panksepp, 2012). Primary processes are innate. They rely on outputs that were preset during phylogenesis (i.e., the evolution of the species). Secondary processes are learned. They rely on outputs that were preset during ontogenesis (i.e., evolution of the individual). Tertiary processes are computations. They can use raw stimulus input or preset representations but their output is freshly produced during microgenesis (i.e., evolution of some process in real time). Although the preset or novel output of operations is in principle independent of types of operations and complexity, a compelling intuition is that primary and secondary processes suffice with single-input associative operations, whereas computation involves a multiple-input operation, whether it is rule-based or associative. To bake a cake (i.e., a novel entity), you typically have to combine several ingredients (i.e., multiple inputs).

1.3.1.3 Operating Conditions and Automaticity Features

In addition to the representations and operations that fix the nature of a mental process, it is worth pointing at factors that count as conditions under which a process can operate or that influence the strength of a process. Examples are the duration, intensity, goal relevance, and un/expectedness of the input, the amount of attention directed at the input, the recency and frequency of the input, and the goal to engage in the processing of the input. Stimuli that are longer-lasting, more intense, more goal-relevant, more *or* less expected, more attended to, recently and frequently processed, and intended to be processed by the person are more likely to be processed or are processed better. Different taxonomies have been proposed to organize these factors. The social psychological literature, for example, groups factors into the categories of opportunity (e.g., stimulus duration and intensity), capacity (e.g., attentional resources), and motivation (e.g., the goal to engage in the process). I recently proposed a more detailed taxonomy (Moors, 2016), based on the distinctions between current vs. prior factors and between observable vs. mental factors, combining them in a four-field table with (a) current observable factors (e.g., stimulus duration and intensity), (b) prior observable factors (e.g., recency and frequency), (c) current mental factors (i.e., quality of the current representation), and (d) prior mental factors (e.g., quality of a prior representation in working memory).

Only a small subset of the above-listed factors, namely duration/time, attention, and intention, have been linked to the dichotomy between automatic and non-automatic processes. Automatic processes have been characterized as fast, efficient, and unintentional, but also as difficult to counteract and unconscious; non-automatic processes have been bestowed with the opposites of these features (Bargh, 1994; Moors, 2016; Moors & De Houwer, 2006a). While the first three features can be recast in terms of operating conditions (as above), the last two features escape this framing. Saying that a process is automatic in the sense of fast, efficient, or unintentional comes down to saying that this process requires little time, little attentional capacity, or no goal to engage in the process, respectively. However, a process that is difficult to counteract does not operate due to, but despite, the presence of the goal to counteract the process. And whether a process is conscious or unconscious is more aptly considered as a consequence of the presence of other conditions (e.g., time) rather than as a condition for the operation of the process itself (although it can be a condition for the operation of subsequent processes; Moors, 2016).

Starting from the premises that all processes require an input of sufficient quality to get launched, and that many mental processes take a representation as their input, I have proposed that the quality or activation level of this input representation is the proximal factor that determines the occurrence and strength of these mental processes (Moors, 2016). A first threshold of activation must be exceeded for the representation to serve as the input to an unconscious process; a second threshold must be reached for the representation to become conscious and serve as the input to a conscious process. In addition, I proposed that the various factors listed above (e.g., stimulus duration, stimulus intensity, attention, frequency, recency) feed into this proximal factor, and that they do so in an additive way. If stimulus duration is reduced, for instance, an increase in stimulus intensity or attention may compensate so that the total activation level is sufficient to launch the process and/or to make it conscious.

A few more words about consciousness are in order. Being conscious of a process requires being conscious of the input and output of a process as well as of their interrelation. These inputs and outputs must be representations and they must be situated on a high level of analysis. It is unlikely that people can be conscious of raw stimulus input and of low-level mental processes. A person can become conscious of the fact that watching advertisements influences her urge to go shopping, for instance, but not of the many detailed representations that go into this process on a lower level of analysis. On a final note, philosophers sometimes use the terms personal-level vs. subpersonal-level processes to refer to conscious vs. unconscious processes.

As mentioned above (see Section 1.3.1.1), the representations involved in mental processes allow for Intentionality, the characteristic of being directed at or about something (Brentano, 1874; Searle, 1983). Both conscious and unconscious mental representations have Intentionality. In conscious representations, moreover, this Intentional aspect is combined with a phenomenal aspect. The phenomenal aspect consists of the qualia or the non-representational content of experience, the aspect of experience that remains after the Intentional aspect is stripped away (Block, 1995; Searle, 1983). For instance, it is what seeing red or having an itching toe feels like, when all there is to know about redness and itches is removed. Sensations like redness and itches have a felt aspect (i.e., there is "something it is like" to have them; Nagel, 1994) that defies any verbal description. Qualia may not be confined to sensations, but also apply to conscious verbal and even abstract thoughts. Some theorists believe such thoughts can only give rise to qualia in an indirect way, however, via the mental images they conjure up (J. J. Prinz, 2010). While unconscious representations are supposed to have Intentionality without phenomenality, some authors have ventured the existence of conscious mental entities that have phenomenality but lack Intentionality, such as objectless sensations and positive or negative feelings (e.g., Reisenzein, 2012; but see Brentano, 1874).[11] Finally, the first-order consciousness discussed so far must be distinguished from second-order consciousness or the state of being conscious of one's first-order conscious entities (Block, 1995; Wegner & Bargh, 1998). Non-human animals are typically assumed to be capable of the former but not the latter type (see Heyes, 2008).

1.3.2 Neural Mechanistic Explanations

The parts and activities in neural explanations correspond to neural representations and neural operations. Neural representations have sometimes been identified with populations or patterns of firing neurons (Bechtel, 2001). In the domain of perception, for instance, the neurons in cortical area V4 code for color whereas those in cortical areas MT and V5 code for motion. In other domains, additional criteria are used to demarcate working parts such as patterns of connectivity (e.g., Bechtel, 2005,

[11] There is debate about how to cash out the distinction between Intentional and phenomenal aspects of consciousness and about how to relate both aspects to one another. So far, I have equated the Intentional aspect with the content of representations, and the phenomenal aspect with non-representational content. The phenomenal aspect can be absent but if it is present, it is supervenient or dependent on the Intentional aspect, like the icing on a cake (Byrne, 2001). Another proposal is that the Intentional aspect depends on the phenomenal aspect in the sense that the phenomenal aspect is what gives the Intentional aspect its meaning (Natsoulas, 1981), as can be illustrated by the argument that the abstract thought of a circle remains meaningless until it is injected with phenomenal experience. Still another proposal is that both aspects are mutually dependent and interwoven (Eilan, 1998).

p. 316; Mundale, 1998). Identifying operations in the brain turns out to be more challenging, however (Bechtel, 2005). In connectionist models (e.g., Rumelhart et al., 1986), the only operation allowed is the firing or activation of the neurons. Yet some scholars have argued that this level of characterizing operations may be too low (Bechtel, 2005).

1.3.3 Dual-Process/System and Multiple-Process/System Models

The various dichotomies listed so far have all been used as grounds for splitting the realm of processes into two exhaustive subsets. Dual-process models have been based, for instance, on formats of representations (e.g., modal vs. amodal; associative vs. propositional), contents of representations (e.g., heuristic vs. systematic; Chaiken et al., 1989; stimulus-driven vs. goal-directed; Balleine & Dickinson, 1998), types of operations (e.g., associative vs. rule-based; S. A. Sloman, 1996), operating conditions (e.g., automatic vs. non-automatic; see Moors & De Houwer, 2006a), and brain locations (e.g., subcortical vs. prefrontal cortical).

Many dual-*process* models have mapped two or more dichotomies onto each other, which has turned them into dual-*system* models (e.g., J. S. B. T. Evans, 2003; Hofmann et al., 2009; Kahneman & Frederick, 2005; E. R. Smith & DeCoster, 2000; Strack & Deutsch, 2004). System 1 houses processes that represent heuristic information (in reasoning models) or information on stimuli and responses (i.e., stimulus-driven process in behavioral models), represented in the form of associations (e.g., handsome – reliable; snake – flee), activated via associative operations, in an automatic way, and implemented by subcortical brain areas. System 2 houses processes that represent systematic information (in reasoning models) or information about outcomes of responses (i.e., goal-directed process in behavioral models), in propositional format, handled by rule-based operations, in a non-automatic way, and implemented in prefrontal cortical brain areas.

Dual-system models have met with serious criticism (Keren & Schul, 2009; Melnikoff & Bargh, 2018a, 2018b; Moors, 2014c; Moors & De Houwer, 2006b). In brief, it has been argued that assumptions of alignment should not be made a priori but should be investigated empirically. This empirical research is complicated by the fact that some dichotomies (e.g., associative vs. rule-based operations) resist a clear definition, thereby making it nearly impossible to diagnose them (Hahn & Chater, 1998; see Moors, 2014c, for a review). Other dichotomies, such as that between automatic and non-automatic modes of processing (and perhaps also that between associative vs. rule-based operations), are gradual in nature instead of binary, thereby allowing only for relative conclusions. Keeping these caveats in mind, empirical research does provide evidence for *non*-alignment between several dichotomies. For example, several studies have shown that goal-directed processes can be relatively automatic (Aarts & Dijksterhuis, 2000). Other work has shown that people can learn to categorize stimuli based on two sources of information (e.g., angle and density) that follow a complex rule even though participants are unable to articulate the rule (Hélie, Roeder, & Ash, 2010; Hélie, Waldschmidt, & Ash, 2010; Kovacs et al., 2021). Still other research has shown that rule-based reasoning can be fast (Newman et al., 2017). These findings have led some scholars to propose single-system models in which dissociations are understood in terms of complexity rather than in terms of qualitatively different systems (e.g., Kruglanski & Gigerenzer, 2011; Osman, 2004).

Some scholars have proposed triple-system models (e.g., Leventhal & Scherer, 1987; Panksepp, 2012; Panksepp & Watt, 2011).[12] Panksepp (2012), for instance, distinguished between three layers of organization in the brain. The first layer located in subcortical areas houses innate (i.e., primary) processes, mostly stimulus-driven ones, which are supposed to be triggered automatically. The second layer houses basic learning (i.e., secondary) processes such as those involved in classical and operant conditioning. Learning does not start from a blank slate, but builds further on innate processes. Learning processes may range in complexity and there is debate about the extent to which they involve computation. Once installed, however, deployment of innate and learned knowledge is assumed to happen via single-input associative processes that are automatic. The third layer is located in neocortical regions and houses computations (i.e., tertiary processes), which are assumed to rely on complex rule-based operations that are non-automatic. Here again, the alignments are assumed a priori and may not survive empirical testing.

1.3.4 Rationality

The two systems in dual-system models have also been aligned with another dichotomy: that between rationality and irrationality. This dichotomy does not concern properties of the processes or their operating conditions, but points at an outsider evaluation relative to certain standards (Davidson, 1985b). The presence of such standards reveals a normativist approach (as distinct from a descriptivist approach, e.g., Elqayam & Evans, 2011). Before turning to the alignment, let me explain a few basic distinctions.

Rationality comes in different shapes. Theoretical or epistemic rationality refers to the accuracy of an entity to represent the external

[12] Some scholars have also proposed multi-process or multi-system models with four types of processes (e.g., Conrey et al., 2005; Sherman, 2006).

https://doi.org/10.1017/9781107588882.004 Published online by Cambridge University Press

world. Practical rationality or adaptiveness refers to the degree to which an entity satisfies goals and leads to well-being. Afferent representations, such as perceptions and beliefs, have a mind-to-world direction of fit. They are evaluated in terms of how well they fit with the external world, that is, their accuracy or theoretical rationality. Efferent representations, such as desires and goals, have a world-to-mind direction of fit. They are evaluated in terms of how well the world fits with them, that is, how well they are satisfied. The satisfaction of goals must be done by subgoals and ultimately by behavior. Thus, it is subgoals and behaviors that are typically evaluated in terms of how well they satisfy goals or well-being, that is, how practically rational they are.

In addition to the afferent and efferent representations and behavior that are produced as the outputs of processes, rationality can also be judged for the processes themselves (Elster, 2010). While the rationality of the outputs of processes is judged based on their *actual* fit (accuracy in the case of theoretical rationality; satisfaction in the case of practical rationality), the rationality of processes is judged based on their *potential* to produce a fitting output, irrespective of the output itself. Rationality in the output-sense and the process-sense may dissociate. Indeed, a reasoning process that uses the right logic may still produce a false belief whereas one that uses the wrong logic may still produce a correct belief. Likewise, a process of behavior causation designed to satisfy goals (i.e., a goal-directed process) may fail to do so whereas one that is not so designed (e.g., a stimulus-driven process) may still accidentally satisfy goals.

One of the reasons for these dissociations is that the rationality of human thought and behavior is never Olympian, but always bounded by the information that is available to the individual, along with the opportunity, capacity, and motivation of the individual to process this information (Bechtel & Richardson, 2010; Simon, 1983). Individuals facing a decision lack information about all possible outcomes, and especially all possible long-term outcomes, of their behavior. Rationality can also be judged from a global or enlightened point of view, which depends on the best evidence that is actually or conceivably available (and hence still not Olympian) (Salmela, 2008). A decision process is rational according to an enlightened standard if the individual makes use of the best available information, but irrational if the individual does not use this information, for instance, due to a lack of opportunity, capacity, and/or motivation.[13]

A few additional distinctions are worth making regarding practical rationality. For one thing, the goals and well-being can be those of one

[13] It could be argued that opportunity does not belong in this list because a lack of opportunity implies a lack of access to information.

individual or those of a community of people. The former type of practical rationality can be called prudential rationality; the latter type can be called moral rationality.[14] Orthogonal to the beneficiary of the rational behavior (individual, community), it can also be specified who does the evaluating: the individual or the community. Taken together, the behavior of an individual can be evaluated by an individual or a community to be good or bad for the individual or the community. Finally, well-being can be understood in terms of the satisfaction of short-term goals (local rationality), long-term goals (ultimate rationality), or an optimal mix of the two (Lemaire, 2021). These are just a few distinctions that highlight the versatility and complexity of the notion of rationality.

Turning back to the alignment of dichotomies, System 1 is typically mapped onto irrationality and System 2 onto rationality. I would even argue that the apparent deviations from rationality observed in daily life are what motivated the creation of dual-system models in the first place. It is when people talk or act dumb that explanations arise in terms of a dumb system taking over. People tend to think that dissociations on the observable super-level of analysis match with dissociations on the mental and neural super-levels and they disregard the possibility that the same mechanism may produce different outputs given different inputs and different other conditions. Yet as several scholars have argued, even double dissociations (e.g., Lieberman et al., 2004) are ultimately unsuitable to settle debates between dual-system models and alternative, single-system models (Chater, 2003; Keren & Schul, 2009).

But how can we make the alignment between ir/rationality and System 1/2 more intelligible? A first source of this alignment is the obvious connection between heuristic content and irrationality and systematic content and rationality. Buying a car because it has the best cost-benefit ratio is more rational than buying it because the salesperson is handsome. A second source is the widely (but often implicitly) assumed trade-off between automaticity and rationality (reminiscent of the better-known trade-off between speed and accuracy), which is tied to the complexity and hence flexibility of processes (e.g., Strack & Deutsch, 2004; see Moors et al., 2017). In short, complex processes are supposed to be non-automatic, or less automatic, than their simple counterparts because integrating more inputs (i.e., vertical complexity) or going through more steps (i.e., horizontal complexity) takes more time and effort. At the same time, complex processes are more flexible than their simple counterparts

[14] Moral rationality is thus one way to cash out morality. The alignment between morality and the well-being of a group of people is grounded in an extrinsic, relational view of moral values (see Rodogno, 2016).

https://doi.org/10.1017/9781107588882.004 Published online by Cambridge University Press

because the more sources of information are taken into account, the better the output can be attuned to these sources. For instance, a goal-directed process that takes into account the outcomes of response options allows for more flexibility than a stimulus-driven process in which this information is absent. More flexible processes, in turn, are more likely to produce accurate (i.e., theoretically rational) and hence adaptive (i.e., practically rational) outcomes (see Box 2.1).

The alignment of the two systems with the rational-irrational dichotomy has recently come under fire. Not all forms of complexity require more time and effort, or the differences may be negligible. Eventually, it is an empirical question to know just how much complexity can be handled under poor operating conditions (see Moors, 2014c; Moors et al., 2017).

1.3.5 Cognition

Now is perhaps a good time to turn to the multifarious meaning of the term cognition. Cognition is a contrastive notion: It takes on different meanings depending on the entities with which it is contrasted (Moors, 2007, 2009). When contrasted with the body, cognitive means mental. For scholars who believe that the realm of the mental is exhausted by representational entities (e.g., Brentano, 1874), cognitive is also synonymous with representational. Thus, cognitive processes are representation-mediated processes. For scholars who leave room within the mental for representational (i.e., Intentional) as well as non-representational (i.e., purely phenomenal) entities, cognitive also refers to representational but they use it to mark the boundary with the phenomenal part of the mental.

The term cognitive has also been used to point at a specific content or format of representation, to a specific type of operations, and to non-automatic processes. Thus, when contrasted with emotional representations, cognitive representations refer to representations with cold, non-valenced content. When contrasted with motivational or conative (i.e., dynamic or efferent) representations, cognitive representations refer to pure knowledge (i.e., non-dynamic or afferent) representations. When contrasted with perceptual or sensory representations, cognitive representations refer to conceptual or propositional representations. When contrasted with associative operations, cognitive operations are understood as rule-based. And when contrasted with automatic processes, cognitive processes are understood as non-automatic. A final meaning of cognitive is when it is used to refer to the mental super-level and contrasted with the observable and brain super-levels. In this book, I will specify the meaning of cognition I have in mind if the contrasting category is not obvious.

On a final note, scholars who take representations as the bearers of information are called representationalists. Throughout the history of cognitive science, several scholars have explored non-representationalist alternatives (e.g., Gibson, 1979; Hutto & Myin, 2017; Stich, 1983; Wakefield & Dreyfus, 1991; see discussions by Bechtel, 1998a, 1998b). They maintain that cognition and Intentionality are possible without representations (see more in Chapter 6).

In the coming chapters, I apply the demarcation-explanation cycle to the emotion domain. As mentioned, theories in this domain are still trying to figure out what emotions are. That is, they try to find an adequate scientific definition for the set of emotions, or – if this turns out to be impossible – to replace it with sets that promise to be more fruitful. The sober fact that the first cycle has so far not led to a consensual scientific definition of emotion has not stopped researchers from moving on to further cycles in which emotions figure in other explananda, such as the influence of emotions on attention, perception, memory, judgment, decision-making or behavior, and psychopathology. The focus of this book will nevertheless be on emotion theories concerned with the first cycle.

In psychology, theories are known as evolutionary theories (e.g., Ekman, 1992a), network theories (e.g., Leventhal, 1984), appraisal theories (e.g., Lazarus, 1991; Roseman, 2013; Scherer, 2009b), the goal-directed theory (e.g., Moors, 2017a), psychological constructionist theories (e.g., Barrett, 2006b; Russell, 2003; Schachter, 1964), and social theories (e.g., Mesquita & Parkinson, 2022). In philosophy, theories go by the names of feeling theories (e.g., James, 1890b), judgmental theories (e.g., Green, 1992; Solomon, 1993), quasi-judgmental theories (e.g., Greenspan, 1988), perceptual theories (e.g., Tappolet, 2016), embodied theories (e.g., Colombetti, 2014; Deonna & Teroni, 2012; Griffiths, 2004b; Hutto, 2012; J. J. Prinz, 2004a), and motivational theories (e.g., Scarantino, 2014).

To facilitate the comparison of emotion theories, I organize them in a new typology built around the various stages in the demarcation-explanation cycle. Each of the stages presents questions for which different theories have provided different answers. The typology outlines a multi-axis space in which the axes correspond to the questions that are encountered during the consecutive stages of the cycle. Emotion theories can be placed and compared within this space depending on the answers they have provided to these questions. It may be noted upfront that this exercise is complicated by the fact that theories are not static entities, but evolve continuously. In addition to an analytic approach, in which I try to do justice to the idiosyncrasies of individual theories, I will also adopt a more synthetic approach, in which I try to identify axes that allow drawing fault lines (FL) between larger groups of theories. Chapter 2 identifies axes and fault lines and provides a broad overview

https://doi.org/10.1017/9781107588882.004 Published online by Cambridge University Press

of the possible choices that have been taken by emotion theories (see Table 2.4). Chapters 3–9 discuss emotion theories one by one. To structure my discussion of these theories, I had to create discrete categories again. As a basis for this, I selected one axis, that of the causal-mechanistic explanations of theories, which resulted in the following overarching categories or families: (a) evolutionary theories (including motivational theories), (b) network theories, (c) stimulus evaluation theories (including appraisal theories, judgmental theories, quasi-judgmental theories, perceptual theories, and embodied theories), (d) response evaluation theories (including the goal-directed theory), (e) psychological constructionist theories, and (f) social theories. Prior to my discussion of these theories, I start with two general precursors: Darwin (1872) and James (1890b; a feeling theory).

CHAPTER 2

Demarcation-Explanation Cycle Applied to Emotion Theories

Emotion theorists take emotions as the explanandum, at least at the start of the first cycle, and they seek an answer to the questions of (a) how to demarcate the set of emotions from other sets, and (b) how best to partition or organize the set of emotions internally. To this end, they first choose a working definition of emotions (Stage 1). After this, they offer constitutive, causal, and/or mechanistic explanations of emotions (Stage 2). Extensive empirical validation of these explanations (Stage 3) ideally precedes the proposal of a scientific definition (Stage 4) but most theories already present a scientific definition before the empirical validation is complete, however premature that might seem.

Prior to the start of the demarcation-explanation cycle, theories already differ in their expectations about how successful the cycle will turn out to be for emotion. A first difference has to do with demarcation. Theories differ in their outlook on whether the folk set of emotions (or a portion of this set) will eventually achieve a scientific status (**Axis 1, FL**). On one side are vindicators who expect this to be the case. For them, theory development simply consists in discovering an explanans that all (or most) members of the folk set of emotions have in common and hence forms the basis for demarcating the set. Most emotion theories are vindicators. On the other side are skeptics, who express no hope that the folk set of emotions will be turned into an adequate scientific set (Bindra, 1969; Duffy, 1934, 1941a, 1941b; Dunlap, 1932; Fantino, 1973; Hunt, 1941; Meyer, 1933; Moors, 2017a; Russell, 2003; Skinner, 1953; Verplanck, 1954; see Kleinginna & Kleinginna, 1981). Instead of vindicating common sense, skeptics aim to critically examine it. Vindicators expect emotion to be like water, a set with an identifiable essence, lying in wait to be discovered. But vindicators can also live with an intensional definition of emotion that is more like that of "lemons," "fish," or "jade." Skeptics expect emotion to be more like "air," a set that has no clear essence, and that will have to be replaced by other sets that do. If their expectations come true, skeptics turn into eliminativists.

A further difference between theories has to do with partitioning. Theories vary in whether they expect that the emotion types from natural language (e.g., fear, anger, sadness, joy) will provide a good basis for carving up the set of emotions (**Axis 2, FL**). On one side are again vindicators, who try to vindicate the "vernacular" emotion types. On the other side are skeptics or eliminativists, who doubt or reject them.[15]

2.1 Stage 1: Working Definition

Working definitions of emotions articulate and elaborate on our ordinary understanding of emotions (Griffiths, 1997). Their role is to specify the agenda of emotion theories, the desiderata that an adequate theory of emotions should meet (Deonna & Teroni, 2012; Mulligan & Scherer, 2012; but see Charland, 2005b). Intensional working definitions ideally specify a number of properties that serve to demarcate the set of emotions from other sets. Given that it is far too soon at this stage to establish the necessity of properties, intensional working definitions fit best with a cluster-type format in which properties have the status of being "typical" at best. Moreover, given that properties at this stage are properties of phenomena, they also have the status of being "apparent." Even theories that ultimately do not accept certain properties as necessary or real in their scientific definitions (in Stage 4) should still be able to explain why they are typical or apparent (Reisenzein, 2012). Divisio working definitions specify samples of subsets of emotions – the emotion types – that theories will have to account for. This not only helps to draw the contours of the set but also suggests a way for how the set should be partitioned.

The content of working definitions is sampled via the armchair or by consulting prior theories (revealing the intuition of scholars), via conceptual analysis (revealing the intuition of a community crystallized in language), and/or via descriptive empirical research (e.g., by studying the lexicon and/or laypeople's intuitions; Averill, 1975; Clore et al., 1987; Shaver et al., 1987; Storm & Storm, 1987). In short, intuition and observation are used as windows for discovering the desiderata for emotion theories (Hatzimoysis, 2010). Although several desiderata reflect pretheoretical intuitions about emotions, they are not presented in pretheoretical form but linked to the concepts laid out in Chapter 1.

Theories vary in the content of the working definitions that they adopt during Stage 1 (**Axis 3**) and this is true regardless of whether they take an intensional (**Axis 3a**) or a divisio format (**Axis 3b**). This variety will complicate the eventual comparison of their scientific definitions during Stage 4. Fortunately, the consensus in Stage 1 is still reasonable enough to

[15] It should be noted that some theories deny a scientific status to the set of emotions as a whole, but still grant scientific status to individual emotions (e.g., Griffiths, 1997; Panksepp, 1998; Rorty, 1980).

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

carry on, at least when it comes to typical and apparent properties of emotions and prototypical emotion types.

2.1.1 Intensional Working Definition

This section reviews typical and apparent properties for which there is fair (although not perfect) consensus that adequate theories of emotion should explain them (see Table 2.1). The properties roughly fall into five classes related to the questions of the "where" (Section 2.1.1.1), "when" (Section 2.1.1.2), "what" (Section 2.1.1.3), "how" (Section 2.1.1.4), and "why" (Section 2.1.1.5) of emotions and their close correlates.[16] For each of the properties, I shall indicate how they help to provisionally demarcate the set of emotions from other sets. The section closes by pointing at alignments that theorists have drawn among several properties (Section 2.1.1.6).

2.1.1.1 Where

Most scholars agree that emotions are located in the individual. A notable exception are social theorists who have toyed with the idea that emotions reside in interactions between individuals (Boiger & Mesquita, 2012; Colombetti & Krueger, 2015; De Rivera & Grinkis, 1986). There is also fair consensus that emotions are not limited to human adults, based on observed analogies with behaviors of infants (e.g., crying, smiling, cringing, fussing) and non-human animals (e.g., flight, fight, freezing, fright). Despite the presumed ontogenetic and phylogenetic continuity, the impression is nevertheless that the range of emotions and/or the level of their sophistication is higher in human adults than in these other groups. For instance, infants and non-human animals can have frustration and distress whereas human adults can have indignation, shame, guilt, and regret.

2.1.1.2 When

Emotions typically arise upon the occurrence of certain (a) stimuli, which are physical events in the external and internal world such as an aggressive opponent or a sudden pain in the foot[17] and (b) certain representations, which are mental events such as the memory, expectation, or

[16] The reason for also considering properties of close correlates of emotions (i.e., entities that are non-trivially connected to emotions) at this stage (Stage 1) is that it is too soon to tell which entities are part of the emotion proper and which entities count as causes and consequences. In fact, this is a question that should be addressed by the constitutive explanation in Stage 2 (see Section 2.2.1).

[17] Following the behaviorist tradition, I reserve the technical term "stimulus" for events in the person's external or internal physical environment (see also Heider, 1926/1959, cited in Hommel, 2009). Thus, I will not stretch the term to also include mental events.

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

Table 2.1. *Typical and apparent properties of emotions listed in the working definition*

	typical and apparent properties of *emotions*	atypical properties of *emotions*	*other entities* lacking the typical and apparent properties of emotions
where	IN INDIVIDUAL adult humans, infants, non-human animals		
when	AFTER EVENTS OF MAJOR IMPORTANCE: physical events = stimuli, mental events	AFTER AESTHETIC AND FICTIONAL STIMULI	
what: aspects	MENTAL ASPECT:		*purely physical entities*
	Intentional aspect: formal + particular object		*- object-less mental entities: e.g., purely phenomenal sensations* *- mental entities with formal object but indeterminate or overgeneral particular object: e.g., moods*
	phenomenal aspect		*unconscious mental representations*
	BODILY ASPECT: peripheral physiological activity, musculoskeletal behavior: subtle, coarse		*purely intellectual mental entities: e.g., thoughts and beliefs*
	ACTION TENDENCY = again mental aspect but with bodily content		

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

Table 2.1. (*cont.*)

	typical and apparent properties of *emotions*	atypical properties of *emotions*	*other entities* lacking the typical and apparent properties of emotions
what: features	HOT	COLD	
	qualitative: - positive/negative valence - emotion-specific quality	qualitative: neutral	
	quantitative: - intense - short-lived	quantitative: - weak - long-lived	*- weak mental entities: e.g., moods, sentiments, personality traits* *- long-lived mental entities: e.g., moods, sentiments, personality traits*
how	AUTOMATIC	NON-AUTOMATIC	*non-automatic mental entities: e.g., pensive states*
	CONTROL PRECEDENCE: - priority claiming - persistence in face of obstacles and of attempts to stop them at will		
	DIS/ ORGANIZATION		
why	IR/RATIONAL: - theoretically ir/ rational - practically irrational	practically rational	*instrumental behavior*

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

imagination of an aggressive opponent or of having pain. Not all events elicit emotions, thus creating a task for emotion theories to discover the regularities. Typical events are those that have major importance for individuals such as the death of a friend, a love declaration, a major achievement, a vicious dog, and betrayal. Another class of events that seem to have strong emotional powers are fictional and aesthetic stimuli (stories, films, music, and other works of art). In addition to identifying the types of events that elicit emotions, theories need to explain the striking variety that events have in their capacity to elicit emotions in different people and on different occasions. For instance, Sam cries at the movie while Sunny remains indifferent, but the next time Sam watches the same movie, his tears have already dried out.

2.1.1.3 What

This section specifies types of ingredients or aspects of emotional episodes as well as their qualitative and quantitative features.

INGREDIENTS OR ASPECTS

Theorists who locate emotions in the individual agree that emotions are constituted or accompanied by mental as well as bodily aspects.[18] The mental aspect sets emotions apart from purely physical, non-mental phenomena such as digestion and thermoregulation (e.g., Clore & Ortony, 2000; Frijda, 1986; Leventhal & Scherer, 1987). Their bodily aspect sets them apart from purely intellectual phenomena such as thoughts and beliefs (James, 1890b).

Focusing on the mind, emotions are said to have both an Intentional aspect and a phenomenal aspect (Deonna & Scherer, 2010; Deonna & Teroni, 2012; Mulligan & Scherer, 2012). That they have an Intentional aspect means that they carry a directedness. This is exemplified by expressions like being afraid *of* the dark, being angry *with* someone *about* betrayal, being sad *about* someone's death, and being happy *about* rediscovering a lost love (Aristotle, *Rhetoric*; Averill, 1980; Brentano, 1874; Clore & Ortony, 2000; Meinong, 1894; Sartre, 1939; Solomon, 1973; but see Hume, 1739). In all these examples, emotions are directed at objects in the world. This makes them similar to cognitive states (perceptions, beliefs), which have a mind-to-world direction of fit.

It has been suggested that the object of an emotion can, but does not have to, coincide with the cause of the emotion. Anger can be directed at the perpetrator who caused one's anger, but it can also be directed at the

[18] In many emotion theories (except enactivist ones), some form of dualism between mind and body proves difficult to throw off (see Jaworski, 2018).

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

next person who crosses one's path. Likewise, fear directed at a vicious dog can be caused by the presence of the vicious dog or by rumors about this dog or even a coffee overdose (Deonna & Scherer, 2010; de Sousa, 2003; Solomon, 1973; but see J. J. Prinz, 2004a).

A common distinction in philosophy is that between the particular and the formal object of emotions (de Sousa, 1987, p. 126, 2003; Kenny, 1963, chapter 9; Roberts, 2003; Teroni, 2007). The *particular object* or target is the specific object that an emotion is directed at, such as the person one is afraid of, angry with, or sad about. The *formal object* is the property that is common to all particular objects in all tokens of an emotion type, such as that fear is about danger, anger is about offense, and sadness is about loss. Whether one is afraid of the dark or of snakes, they both entail danger. The difference between particular and formal objects goes beyond their level of abstraction, however. Talk of formal objects points at the intuition that emotions can play an epistemic role, that they convey information. Sam's fear tells us that there is danger (i.e., the formal object), for instance, but not what he is afraid of (i.e., the particular object).[19]

Theorists quarrel about the type of information that is conveyed by emotions. *Externalists* emphasize that emotions inform us about events in the world. This is captured in a realist interpretation of the formal object, the idea that dangers, offenses, and losses can be objectively established (D'Arms & Jacobson, 2000). For instance, rattlesnakes are objectively dangerous while stuffed animals are not. *Interactionists*, on the other hand, emphasize that emotions inform us about the way in which the world interacts with the mind. This is captured in a relative interpretation of the formal object, the idea that dangers, offenses, and losses depend on person factors such as goals, values, and learning histories (Helm, 2001, 2009; Lépine, 2016; Roberts, 2003; Salmela, 2008).[20] For instance, whether Sam is offended by a blue joke depends on his sensitivities, and whether the frost killing Sunny's cucumber plants constitutes a true loss depends on whether she values her cucumber plants. The upshot is still this: Emotions refer to and inform us about events in the world, regardless of whether they do so in an objective (i.e.,

[19] In addition to the *particular object*, the *formal object*, and the *cause* of emotions, de Sousa (1987, 2003) also mentioned the *focus* of an emotion, which is the property of the particular object that the emotion is directed at (e.g., indignation at the tone of a remark not the content), and the *aim* of the emotion, which is the action tendency (e.g., the tendency to fight the offender; see below).

[20] The terms "externalist" and "interactionist" are my own. As far as I know, they have not been used by philosophers in this context.

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

externalist view) or subjective way (i.e., interactionist view). This is captured in the notion of world-directed Intentionality.[21]

The fact that we can dig up properties of emotions from introspection indicates that emotions are not just mental but also conscious; they are experienced or felt (Clore, 1994; Hatzimoysis, 2007; Ortony, 2021). This means that in addition to an Intentional aspect, they also have a phenomenal aspect or "qualia." The phenomenal aspect of emotions sets them apart from unconscious mental representations, which have Intentionality but lack phenomenality. Their Intentional aspect sets them apart from objectless mental entities such as purely phenomenal sensations (Block, 1995), and according to some authors, from moods (e.g., an anxious mood; Averill, 1980; Clore & Ortony, 2000; Roberts, 2003; Schwarz & Clore, 1996). Rather than seeing moods as entirely objectless, however, several authors argue that moods have an intact formal object, but an indeterminate or overgeneral particular object (Goldie, 2000; Price, 2006; Scarantino, 2010; Solomon, 1993). For instance, in an anxious mood, the formal object is danger, but the particular object is nothing in particular or is waiting behind every corner. Tappolet (2018) alternatively proposed that emotions have real objects whereas moods have possible objects.

Few authors disagree that Intentionality and phenomenality are typical and apparent properties of emotions, even if some of them deny the real status of Intentionality (I. Goldstein, 2002; Hume, 1739; Reisenzein, 2012; Shargel, 2015; Whiting, 2011) and others reject the necessity status of phenomenality with the aim to make room for unconscious (i.e., unfelt) emotions (Damasio, 2004; J. J. Prinz, 2004a; Roberts, 1988; A. Sloman & Croucher, 1981).[22]

Turning to the body, emotions are said to be characterized or accompanied by (a) peripheral physiological activity such as rises and falls in blood pressure, heart rate, blushing, and crying, and (b) motor responses or musculo-skeletal behavior, which can be split into subtle, nonverbal expressions in the face, the voice, and gestures, on the one hand, and coarse or full-body behaviors such as approach, avoidance, flight, fight, and reparation, on the other hand. Introspection teaches us, however, that even if behavior (both subtle and coarse) is absent, emotions can still be

[21] Note that the world in world-directed Intentionality can be outside (external) or inside (internal) the skin. My fear can be about an impending hurricane or about the pain in my foot. To simplify, however, I will focus on external stimuli from now on.

[22] These scholars either treat feelings as an optional property of emotions or as an entity separate from emotions. Postulating unconscious emotions allows them to accommodate cases of repressed or denied emotions (e.g., when a person shouts "I am not angry!").

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

characterized by the inclination or urge to act, so-called action tendencies or modes of action readiness (Frijda, 1986). It should be noted, however, that action tendencies belong again to the mental realm; they qualify as efferent representations of behavior. This has led some authors to argue that the Intentionality of emotions is not only world-directed but also self-directed (e.g., Helm, 2009; Roberts, 2013). According to them, emotions have a double direction of fit: the mind-to-world direction of fit of cognitive states and the world-to-mind direction of fit of conative states.

FEATURES

Taken together, however, the mere presence of mental and bodily aspects does not suffice to demarcate emotions from neutral daily activities like doing the dishes or walking home. A crucial property that is still missing is their so-called "heat," which has a qualitative and a quantitative side (Frijda, 2007b, p. 26). On the qualitative side, emotions are said to have a marked positive or negative valence (Barrett & Russell, 1998; Ben-Ze'ev, 2000; Cabanac, 2002; Charland, 2005b; Clore & Ortony, 2013; Colombetti, 2005; Damasio, 1994; Deonna et al., 2015; Deonna & Teroni, 2012; Fossum & Barrett, 2000; Frijda, 2005; Helm, 2009; Lambie & Marcel, 2002; Lazarus, 1991; Ortony, 2021; Ortony et al., 1987, 1988; Panksepp, 1998; J. J. Prinz, 2010; Russell, 2003; but see Solomon, 2001; Solomon & Stone, 2002). Some authors have relaxed this property, accepting emotions that gravitate towards neutrality, such as surprise and interest (e.g., Izard, 1977; Silvia, 2005; Tomkins, 1962; S. A. Turner & Silvia, 2006) although there is debate whether these emotions are really neutral (C. A. Smith & Ellsworth, 1985; Topolinski & Strack, 2015; Van de Cruys, 2017).[23] The quality of emotions has not been limited to valence, however. Anger and fear, for instance, are both negative but differ in other respects. This is referred to as emotion-specific quality.

On the quantitative side, the heat of emotions has been related to their intense character and they have also been portrayed as short-lived (Arnold, 1960; Cabanac, 2002; Ekman, 1992a, 1994, 1999a; Frijda, 2007b; Frijda et al., 1991; Mulligan & Scherer, 2012; Schwarz & Clore, 1996; Shadmehr & Ahmed, 2020; but see Tappolet, 2016). There is of course some wiggle room: Research using self-report shows that the intensity of emotions ranges from mild to strong and that they can last from a few seconds to a few days, depending on the importance of the situation and the intensity of the emotion at onset. On average, fear, anger, and joy tend to be more intense than sadness and satisfaction. Sadness tends to last

[23] Topolinski and Strack (2015) argued that the initial phase of surprise is negative (see also Van de Cruys, 2017, for an extensive treatment of the relation between surprise and affect). C. A. Smith and Ellsworth (1985) argued that interest is inherently positive.

longer than anger, and anger in turn, tends to last longer than fear (Verduyn, Delvaux, et al., 2009). Research on self-reported intensity profiles over time reveals three parameters that characterize emotions: steepness at onset, skewness or accumulation, and number of peaks (Verduyn, Van Mechelen, et al., 2009). Sadness and joy have more explosive onsets than anger and affection, with affection standing out as a late-bloomer.

The quantitative profile of emotions separates them from other phenomena with a pronounced valence, such as moods, sentiments, and personality traits. Sentiments are dispositions to have emotions, which are manifested in emotions when the conditions are right (Naar, 2018). For example, a life-long hate for an enemy may result in anger whenever the enemy is around. Personality traits are predispositions to have emotions or sentiments (Hastings et al., 2011). For example, a hot-tempered person is quick to become angry and to develop hate. Moods (e.g., grumpy mood) are milder and last longer than emotions (e.g., anger), and sentiments (e.g., hate) and personality traits (e.g., hot temper) last the longest (Goldie, 2011). This aligns with the proposals of ontologists that emotions and moods are occurrents (i.e., entities with a beginning and an end), whereas sentiments and personality traits are continuants (i.e., enduring entities) (Hastings et al., 2011).

2.1.1.4 How

The next set of properties has to do with the conditions under which emotions arise and the way in which they relate to motivation. Emotions tend to come in a sudden and uninvited way and while the person is engaged in other activities. They tend to disrupt and take over this other activity and to persist in the face of obstacles or attempts to stop them at will. We are flooded by joy, overwhelmed by sadness, gripped by fear, consumed with anger, and caught by surprise (Averill, 1982; Ekman, 1992a, 1994, 1999a; Scarantino, 2016; Zajonc, 1980). In more technical terms, emotions are said to be automatic in the sense that they are quick to arise, efficient, unintentional, and/or difficult to counteract (see Chapter 1). The fact that some authors leave room for non-automatic emotions (e.g., Barrett, Ochsner, & Gross, 2007; Kappas, 2006; Solomon, 1973) should not distract us from the typical automatic profile of emotions, which separates them from more calculated states[24] (Ekman, 1992a). On the flipside of an automatic emotion is a passive subject, who surrenders to the emotion and is a slave to its commands (Helm, 2009). Historians teach us that some phenomena that go by the name of

[24] Even if one intentionally seeks to have an emotion (or seeks to avoid one), whether or not one ends up having one still seems at least partially beyond one’s control.

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

emotions today were called passions a few centuries ago (e.g., Charland, 2010; Dixon, 2003, 2012; Rorty, 1982; Scarantino, 2016).

The disruptive and priority-claiming property of emotions is not usually counted as part of automaticity but goes by the technical term of "control precedence" (Frijda, 1986; Scarantino, 2014). This ties in with Arnold's (1960) idea that emotions take us over completely (she calls this the total stirred-up or all-over effect of emotions) and the widespread idea of emotions as emergency responses (Solomon, 1973). Whether one is reading a book or making a sandwich, when the house is on fire or bad tidings arrive, priorities will be shifted. Note that Frijda (1986) extended his notion of control precedence to also include persistence in the face of obstacles and of attempts to stop emotions at will. Note that the latter meaning has overlap with the automaticity feature of difficulty to counteract.

A final property of emotions pertains to the degree of organization between the parts of the system. Some authors claim that emotions are marked by disorganization or a breakdown of normal functioning (e.g., Young, 1943, 1949). Other authors claim exactly the opposite, that emotions are marked by an increased organization or synchronization (e.g., Leeper, 1948; Scherer, 2000). Given that there is not the slightest consensus about this property, I do not consider it as an incontrovertible desideratum.

2.1.1.5 Why

Why do we have emotions? Do they have a function or are they merely a nuisance? Emotions seem to carry an irrational flavor. They seem to cloud our thinking and make us behave counter to our best interests. The idea that emotion contrasts with reason is widespread and tenacious. You literally cannot open a newspaper without being fed the same line. No matter what goes wrong in the world, emotions have always done it. While public opinion is strongly convinced of this contrast, scholars are more divided. Some scholars confirm that irrationality is an inherent property of emotions (Dolan, 2002; Elster, 2010; Griffiths, 1997; Weber, 1947), yet an exploration of the history of the emotion concept teaches us that early philosophers distinguished between rational and irrational emotion types, with virtuous emotions such as devotion and gratitude belonging to the former type and sinful emotions such as lust and anger belonging to the latter (Dixon, 2012). This idea is revived in recent psychological work in which guilt is associated with rational effects and anger with irrational effects (Darlow & Sloman, 2010; Motro et al., 2018). Other authors argue that all emotion types can be rational or irrational depending on their fit with certain standards (Arnold, 1960). Different types of standards are considered in different types of rationality (see also Chapter 1, Section 1.3.4). Emotions have been judged for their theoretical rationality as well as for their practical rationality.

A token emotion is said to be *theoretically rational* or accurate (correct, Deonna & Teroni, 2012; Kenny, 1963; fitting, Scarantino & de Sousa, 2018; appropriate, Salmela, 2008) if its particular object instantiates the right formal object. On the premise that danger is the formal object of fear, fear of a widow spider is accurate because a widow spider is dangerous whereas fear of a house spider is inaccurate because it is harmless. An emotion that is inaccurate can still be justified, however, if it is coherent with the person's false beliefs. Thus fear of a house spider is inaccurate but justified when the person falsely believes that the house spider is dangerous. But an inaccurate emotion can also be unjustified. This is the case for emotions that are recalcitrant to reason, for instance, when the fear of the house spider persists despite the person's belief that it is not dangerous. Inaccuracy points at a gap between objective evidence and subjective apprehension. A lack of justification points at a gap between subjective apprehension and the emotion.

A token emotion is *practically rational* or adaptive if it promotes behavior that is conducive to goal satisfaction and well-being. Sometimes emotions help us achieve goals, for instance, when guilt helps us resist temptations and stick to long-term goals. Sometimes they hinder us, for instance, when sadness pushes us to indulge in temptations, a phenomenon called akrasia or weakness of will. And sometimes they lead to arational actions, seemingly useless actions such as jumping up and down out of joy, slamming the door out of anger, and kissing someone's picture out of grief (e.g., Hursthouse, 1991; Scarantino & Nielsen, 2015; M. Smith, 1998; Goldie, 2000; but see Chapter 9). Philosophers reserve the term "irrational behavior" for behavior that goes *against* goal satisfaction (i.e., that we do for the wrong reasons) while they use the term "arational behavior" for behavior that is *irrelevant* to goal satisfaction (i.e., that we do for no reason at all). It could be argued that arational behavior still qualifies as irrational behavior, however, because performing behavior that does not serve any goal still uses up energy and therefore ultimately goes against goal satisfaction. The practical rationality of emotions is a complex and thorny issue that we can start to unravel by distinguishing between a direct and an indirect influence of emotion on behavior. This can be understood against the backdrop of three pathways of behavior causation that are traditionally distinguished (Volz & Hertwig, 2016).

In a first, *purely emotional pathway*, behavior is directly caused by an emotion. Examples are anger causing aggression, fear causing flight, guilt causing reparation, and sadness causing resignation. The practical rationality of an emotion is partly determined by its theoretical rationality. An emotion that is in/accurate is likely to produce mal/adaptive behavior (Salmela, 2008). Indeed, fear of a widow spider (i.e., accurate) causes avoidance that promotes safety (i.e., adaptive) whereas fear of a house

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

spider (i.e., inaccurate) causes avoidance that hinders moving freely through the house (i.e., maladaptive). But this is not the whole story. An accurate emotion may still produce maladaptive behavior. For instance, anger about an unfair offer encourages people to engage in costly aggression, thereby sacrificing their own profit (Pillutla & Murnighan, 1996; Sanfey et al., 2003). Arational actions form another set of examples. When people jump up and down out of joy, slam the door out of anger, or kiss someone's picture out of grief, their emotion may be accurate but their behavior seems useless. Conversely, an inaccurate emotion may still produce adaptive behavior in the sense of promoting ultimate goal satisfaction. Examples are a person who works up her anger to win a fight (Greenspan, 2004) or a person who forgoes indignation over abuse to avoid more extreme abuse (Salmela, 2008).[25]

In a second, *purely cognitive pathway*, behavior is caused by a cognitive process in the absence of emotion. Two processes subsume under this pathway: a goal-directed process (e.g., choosing between a small or a large balloon) and a stimulus-driven process (e.g., the bowl of nuts triggering approach). Here, the goal-directed process counts as the more practically rational one (see Box 2.1).

In a third, *emotionally infused cognitive pathway*, finally, behavior is caused by a cognitive process, but emotion influences this process thereby exerting an indirect influence on behavior. Several types of presumed influences can be sorted under this third pathway. A first presumed influence is when emotion influences the operating conditions of processes and in this way acts as a switch between goal-directed and stimulus-driven processes. For instance, the emotion of stress is assumed to reduce attentional capacity, thereby causing the system to switch from the more rational goal-directed process to the less rational stimulus-driven process (e.g., Schwabe & Wolf, 2011; but see Buabang, Boddez, et al. 2021).

A second presumed influence is when emotion influences the representations involved in the goal-directed process. Emotions can improve or bias the accuracy of these representations, which is in fact a form of theoretical rationality. When studying this second presumed influence, researchers commonly distinguish between incidental and integral emotions (e.g., Lerner et al., 2015). *Incidental* emotions are ones that are

[25] These cases can be reinterpreted as dissociations between short-term and long-term adaptiveness. This ties in with Scarantino and de Sousa's (2018) subdivision of practical rationality into instrumental and substantive rationality. Instrumental rationality refers to whether the right strategy is chosen to reach a goal; substantive rationality refers to whether the right goal is picked to secure overall well-being. Well-being can itself be cashed out in various ways (Diener et al., 1999), thus opening the door for even more types of practical rationality.

unrelated to the decision at hand, such as lingering fear from a prior dangerous event. *Integral* emotions are ones that relate to the decision at hand, such as fear about a potential outcome of a decision. Integral emotions are often *anticipatory*, elicited by the anticipation of decision outcomes that can be either material (e.g., fear about losing money) or emotional (e.g., fear about regretting a decision). Both incidental and integral emotions must be distinguished from *anticipated* emotions, however, which are not experienced at the time of decision making but merely anticipated to be experienced in the future (e.g., expectation of regret). They are therefore more accurately called *anticipations of emotions*.

Research has shown that incidental negative emotions not only lead to more pessimistic judgments, but also to more systematic, and hence more accurate, information processing than positive emotions (Schwarz & Clore, 2003). Other studies have shown emotion-specific effects beyond valence. For instance, Lerner and Keltner (2001) found that fear leads to pessimistic but anger to optimistic risk perception. Integral, anticipatory emotions have sometimes been shown to improve decisions (Bagozzi & Pieters, 1998; Bechara et al., 1997; but see Newell & Shanks, 2014) and sometimes to impair them (Loewenstein et al., 2001). Anticipations of emotions, also called affective forecasts, are notoriously inaccurate (T. D. Wilson & Gilbert, 2003). For instance, people tend to overestimate the duration of negative emotions.

But increases and decreases in the accuracy of the representations in the decision-making process are only half of the story. This is because the ultimate practical rationality of these emotions depends on the global accuracy of the decision-making process. If an emotion improves the accuracy of a single representation (e.g., the expectancy that staying at home will lead to safety), it may not improve the accuracy of other representations (e.g., the expectancy that staying home will lead to a loss of social contacts) that are also part of the weighing process. The resulting decision may thus be out of balance, that is, not be globally accurate and therefore not be adaptive.

To summarize, the irrational property of emotions can be spelled out in different ways. Emotions can be judged to be irrational if they (a) do not represent the right formal object, (b) bias cognitive processing, or (c) promote behavior that is detrimental for the agent or for the community at large, either directly or indirectly via the cognitive biases in "(b)." Emotions often seem irrational in all three of these senses, and sometimes they do not. Theories should be able to account for this.

2.1.1.6 Alignments

Before closing the section on the intensional working definition of emotions, it is worth pointing out that theorists have assumed alignments among several of the properties of emotions listed so far. A first illustration is the clustering among intensity, automaticity, control precedence, disorganization, and irrationality. The intense, sudden, and uninvited properties of emotions are often linked to their claim for priority and the difficulty to counteract them (e.g., Ekman, 1992a). The more explosive an emotion, the harder it is to resist and the more disorganizing it can be. The automatic nature of emotions, in turn, is often tied to their irrational flavor. The latter connection may derive from the trade-off between automaticity and rationality discussed in Chapter 1 (Section 1.3.4). A similar connection has been drawn between disorganization and irrationality. If the system breaks down, rational behavior should be hindered. The clustering of these properties has led dual-system models to file "hot" emotion or affect under System 1 and "cold" reason under System 2 (see Keren & Schul, 2009). However, the legitimacy of these a priori alignments may be questioned. To give just one example, it is not because disorganization promotes irrational behavior that organization promotes rational behavior. A perfectly organized system may smoothly head towards irrational behavior.

A second illustration of property alignment is the tendency among philosophers (e.g., Deonna & Teroni, 2012; J. J. Prinz, 2004a) to narrow down the phenomenology of emotions to bodily feelings. Deonna and Teroni (2012, p. 79), for instance, argued that the phenomenology of emotions "is best captured in terms of bodily feelings" and they seem to rule out the possibility that mental entities such as images and even abstract thoughts might have qualia as well (as suggested by Block, 1995). Some scholars have resisted this alignment. Helm (2009), for instance, suggested that the positive/negative valence resulting from the process of evaluation is the central aspect of affective phenomenology. Mason and Capitanio (2012, p. 240), from their side, argued that the conventional meaning of experience should shift from sensory experience to the experience of information (see also Oyama, 2000).

2.1.2 Divisio Working Definition

Scholars differ in the precise set of emotion types they seek to explain. Restricted lists comprise prototypical or "blue-ribbon" emotions like fear, anger, sadness, joy, disgust, and surprise. More extended lists add shame, guilt, pride, gratitude, regret, disappointment, indignation, envy, jealousy, being moved, wonder, admiration, adoration, awe, shock, love, hate, affection, annoyance, irritation, envy, indignation, and many more. Despite these differences, however, there is fair consensus about the prototypical emotions, give or take a few. Emotion types can be treated as subsets in the set of emotions in the sense that they are types that comprise tokens of emotions. But emotion types have also been treated as subsets (called families) that are themselves comprised of still smaller subsets (called shades). Anger can be treated as a subset subsuming the smaller subsets of annoyance, irritation, anger, and rage, whereas fear subsumes the smaller subsets of worry, apprehension, anxiety, fear, panic, and terror (e.g., Plutchik, 2001). Such families of shades are often assembled on the basis of participants' responses in sorting tasks or similarity judgment tasks (Fehr & Russell, 1984). The upshot is that adequate theories of emotion should ideally be able to account for the prototypical emotions, and if possible, also for more subtle shadings of these emotions. A final intuition to be accounted for is that different events elicit different emotions in different people and on different occasions. For instance, upon hearing an off-color joke, Sam is amused while Sunny is shocked, and the next time Sunny hears the joke, it just makes her sad.

To wrap up, the desiderata for an adequate theory of emotion are to explain phenomena known as fear, anger, sadness, joy, and so forth, and to explain their typical and apparent properties such as their ontogenetic and phylogenetic continuity, that they are elicited by events of major importance, their mental and bodily aspects, their heat (valence and intensity), their automaticity, their control precedence, and their irrational flavor.

2.2 Stage 2: Explanation

Explanations proposed for emotions can be sorted into constitutive, causal, and mechanistic ones. A constitutive explanation specifies the components of an emotion and ideally also the interrelations among them. If Sam is sad, what are the components of his sadness and how do they hang together? A causal explanation specifies the entities that cause emotions, such as certain types of stimuli and/or certain types of mental events or representations. Let's call them *initial causal entities*. Sam is sad because he lost his cat (i.e., stimulus) or because he imagines losing his cat (i.e., representation). A mechanistic explanation specifies the mechanism, that is, the mental or neural entities that mediate between the initial causes and the emotions as well as the causal relations among these mediating entities. These can be called *mediating causal entities*. Sam losing his cat causes a suite of mental/brain processes, which in turn lead him to being sad.

Scientists are typically not after singular explanations, that is, explanations of a token emotion. Instead, they try to discover regularities that allow for generalization. In constitutive explanations, they try to figure out the components of all tokens of sadness (not just Sam's) or all tokens

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

of emotions (not just sadness). In causal explanations, they try to discover regularities between inputs and outputs. They propose, for instance, that the loss of any valued object (not just cats) and not of non-valued objects (e.g., paperclips) leads to sadness in many if not all people. The regularities specified in causal explanations, in turn, form the starting point for the development of mechanistic explanations. Once we have a clear picture of the inputs and the outputs, we can start figuring out what lies in between.

Not all existing emotion theories can gracefully be mapped on the template outlined above, however. Three observations are worth making. A first observation is that for some theories, a clean separation of constitutive and mechanistic explanations proves difficult. To create their constitutive and mechanistic explanations, theories start from the ingredients listed in the working definition. Some of these ingredients are sorted in the constitutive explanations of theories where they are assigned the role of component of emotion. Other ingredients are pushed out of the emotion and are assigned the role of mediating causal entity or consequence of emotion. There is great variety in the number of the components that theories include in their constitutive explanations. Some include only a single component whereas others include multiple components (see Section 2.2.1.3). Some even go so far as to include the maximal number of components, so that no entity is left to act as the mediating causal entity (mechanism) between the initial causal entity and the emotion. These theories do not provide an etiological mechanistic explanation, which specifies what happens *prior to* the emotion, but rather a constitutive mechanistic explanation, which specifies what happens *inside* the emotion (Craver & Kaplan, 2020). To avoid scattering mechanistic stories across two different sections (the section on constitutive explanations and the section on mechanistic explanations), all talk of mechanisms will be dealt with in the section on mechanistic explanations (Section 2.2.2).

A second observation is that some theories provide purely causal explanations (e.g., so-called descriptive appraisal theories; Clore & Ortony, 2008, 2013; see Chapter 8). They describe the commonalities and differences between events that elicit emotions or specific emotions, but remain agnostic about the mediating mechanisms. Others provide both causal and mechanistic explanations (e.g., so-called process theories of appraisal; e.g., Scherer, 2009b; see Chapter 6). Still others provide purely mechanistic explanations in that they do not specify features of initial causal entities (e.g., network theories, goal-directed theories, psychological constructionist theories). I use the umbrella term "causal-mechanistic explanations" to refer to causal and/or mechanistic explanations and discuss these together in Section 2.2.2.

A third observation is that philosophical theories have a strong focus on constitutive explanations whereas psychological theories are more in the business of crafting causal-mechanistic explanations. This is mirrored in the fact that philosophical theories are often named after the components that emotions are made of (e.g., feeling theories, judgmental theories, quasi-judgmental theories, perceptual theories, and motivational theories) whereas psychological theories are often named after mechanisms (e.g., evolutionary theories,[26] network theories, appraisal theories, the goal-directed theory, psychological constructionist theories). This has sometimes generated confusion in existing theoretical overviews (e.g., Scarantino, 2016; Scarantino & de Sousa, 2018; Tappolet, 2016, p. 12). For instance, appraisal theories (a brand of psychological theories) have sometimes been grouped together with judgmental theories (a brand of philosophical theories) under the label of "cognitive theories" although appraisal theories do not identify emotions with cognitions as judgmental theories do. This is not to say that psychologists eschew talk of components or that philosophers ignore mechanisms. Psychologists do make choices about the components to include in emotion, but they rarely provide extensive justifications for their choices as philosophers do. When it comes to mechanisms, on the other hand, psychologists tend to be more explicit than philosophers in that they embed their mechanisms in a global architecture of the mind, and not unimportantly, that they submit them to empirical testing. Philosophers invent ad hoc mechanisms or borrow them from the psychological literature. Deviations from these trends can be observed in scholars who have crossed the disciplinary boundary, such as philosophically informed psychologists (e.g., Miceli & Castelfranchi, 2015; Reisenzein, 2012) and philosophers with naturalistic inclinations (e.g., Charland, 1995; Colombetti, 2009; Griffiths, 1997; J. J. Prinz, 2004a; Scarantino, 2014). To conclude, the type of explanation that emotion theories propose or capitalize on is an additional source of variation among them (**Axis 4**).

2.2.1 Constitutive Explanations

In shaping their constitutive explanations, theories have sampled from the following list of components (see Table 2.2): (a) a cognitive component with changes in information processing,[27] (b) a motivational component

[26] Evolutionary theories propose evolutionary ancient mechanisms, known as "affect programs," and these theories have accordingly also been named affect program theories (see Moors, 2009).

[27] Here the term cognitive is taken to be broad enough to include representations in all kinds of formats, but narrow enough to exclude conative representations.

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

Table 2.2. *Traditional components*

components	content	examples
cognitive component	information processes	- stimulus evaluation - response evaluation - categorization
motivational component	concrete action tendencies	tendency to approach (distance decrease), avoid (distance increase), flee, fight
	abstract action tendencies	tendency to approach (contact increase), avoid (contact decrease)
	other modes of action readiness = highly abstract action tendencies	tendency to be active, passive
somatic component	peripheral activity (ANS)	heart rate, blood pressure, perspiration, pupil dilation
	central activity (CNS)	brain areas, brain circuits, whole-brain patterns
	hormonal activity (ES)	release of testosterone, adrenaline, cortisol, oxytocin
motor component	subtle behavior (SNS)	facial, vocal, gestural "expressions"
	coarse behavior (SNS)	approach, avoid, flee, fight give in, repair
feeling component	first-order experience: raw feelings	- feelings of other components (Intentional aspect and/or component-specific phenomenal aspect) - feelings of fear, anger, sadness, joy (emotion-specific phenomenal aspect)
	second-order experience: labeled feelings	feelings of fear, anger, sadness, joy

with changes in action tendencies, (c) a somatic component with changes in physiological responses, (d) a motor component with changes in overt behavior, and (e) a feeling component with changes in experience or feelings.

Theorists disagree about the precise content of these components (**Axis 5a**) (Section 2.2.1.1), the granularity of the components (i.e., the fineness of grain at which they are considered; **Axis 5b, FL**) (Section 2.2.1.2), and (c) the number, identity, and necessity status of the components that they ultimately include in their constitutive explanation of emotions (**Axis 5c, FL**) (Section 2.2.1.3) (see Frijda, 2007a, 2007c; Frijda & Scherer, 2009; Parrott, 2007; J. J. Prinz, 2004a). After discussing each of these sources of variety, I note problems with this traditional list of components and propose an alternative list that circumvents these problems and that enables a more in-depth comparison of theories and their explanations (Section 2.2.1.4).

2.2.1.1 Precise Content of Components

The cognitive component houses information processes. These processes differ first of all in terms of their inputs and outputs. A few examples. Stimulus evaluation theories focus on the evaluation or appraisal of stimuli in terms of danger, offense, loss, and so forth (Lazarus, 1991). The goal-directed theory focuses on the evaluation of behavior options in terms of their expected utilities (Moors, 2017a). Psychological constructionist theories focus on the categorization of bodily feelings in terms of specific emotions (Barrett, 2006b). Other differences between processes have to do with the format of the representations (e.g., propositional vs. perceptual) and the types of operations (e.g., rule-based vs. associative) involved.[28]

The motivational component contains action tendencies (e.g., tendencies to orient, flinch, approach, avoid, flee, fight, repair, give in, hug), which can be cast at higher or lower levels of abstraction (Arnold, 1960; Frijda, 1986). The more abstract an action tendency, the more abstract is the relationships that the agent tries to establish or maintain with the world, such as approach or avoidance in the sense of increasing or decreasing contact with an object. The more concrete an action tendency, the more concrete details are specified about the behavior that the agent tries to execute, such as approach and avoidance in the sense of

[28] A minority of scholars also incorporate changes in the cognitive functions of attention and memory within the cognitive component (e.g., Clore & Ortony, 2000; J. J. Prinz, 2004a). Others treat attention and memory as slave systems that support all mental components (Scherer, 2009b). The majority of scholars treat changes in attention and memory also as part of the consequences of emotions.

decreasing or increasing physical distance with an object. The distinction between abstract and concrete action tendencies is merely gradual. Both abstract and concrete action tendencies are goals understood as representations of valued future states (e.g., decreased contact or decreased distance). Abstract action tendencies can be implemented via concrete action tendencies, whether these are physical or only mental in nature. Decreasing contact, for instance, can be done by taking a step back or by redirecting attention or shutting oneself off mentally. Some authors have even included mental action tendencies that have no immediate relation with the outside world, such as reshuffling goal priorities (Cova & Deonna, 2014; Frijda, 1986). Some theorists (e.g., Deonna & Scherer, 2010) only include abstract action tendencies in the emotion and consider the concrete ones as a consequence of emotion. Others (e.g., Frijda, 1986) include abstract and concrete action tendencies with control precedence and exclude those without.

Besides action tendencies, Frijda (1986; Scarantino, 2016) included other modes of action readiness such as a mere readiness to act or change the current state (without specific direction) and a readiness to be inactive or passive (or an unreadiness to be active). Rather than treating the mere readiness to act or be passive as separate from action tendencies, however, I think they can best be considered as highly abstract action tendencies. Note that some authors have objected against inclusion of the tendency to be passive because they consider it to be an illegitimate stretch of the concept "action" tendency (e.g., Eder, 2017).

The somatic component covers (a) physiological responses (e.g., changes in heart rate, blood pressure, perspiration, pupil dilation) supported by the autonomous nervous system (ANS), a part of the peripheral nervous system (PNS), (b) brain (and spine) activity supported by the central nervous system (CNS), and (c) hormonal activity (e.g., release of testosterone, adrenaline, cortisol, oxytocin) supported by the endocrine system (ES). The activities of these systems are strongly intertwined. Hormones support the physiological responses supported by the ANS. For instance, adrenaline spikes make sure that heart rate and blood flow increase and that muscle tension builds up. It may be noted that while some theorists consider brain activity to be a subcomponent of the somatic component (e.g., Roseman, 2013), others do not mention it because they consider brain activity to be part of many if not all of the other components, situated on a lower level of analysis (e.g., Scherer, 2009b).

The motor component refers to changes in overt behavior, supported by the somatic nervous system (SNS, also part of the PNS). Behavior has traditionally been split into (a) subtle movements in the face, the voice, and gestures, often called "expressive" behavior and (b) coarse or full-body behavior (e.g., approach, avoid, fight, flee, give up; Mulligan &

Scherer, 2012; Scherer, 1984). The terms subtle and coarse are not meant to refer to a difference in intensity, but rather to the scale of the body that is involved, and related to this, the potential impact that the behavior has on the physical world. Facial and vocal behavior engage only a small area of the body whereas coarse behavior such as fighting and fleeing makes the whole body work. Because of this, the material impact of subtle movements is often restricted to the body itself whereas coarse behaviors are outward deeds with greater material impact. Both types of behavior can nevertheless have a social impact.

This characterization of the distinction between subtle and coarse behavior presents a few caveats, however. First, behavior categories such as body posture (e.g., shrugged, upright posture) defy this dichotomy because they involve the whole body while having a limited impact on the physical world (Lhommet & Marsella, 2014; Dael et al., 2012a, 2012b). Second, the characterization in terms of less vs. more impact on the physical world misses out on the fact that coarse behavior is sometimes described in highly abstract terms that do not specify the nature of the impact. For instance, approach can be physical (e.g., walking to the fridge), social (e.g., talking to someone), or purely mental (e.g., thinking of something).

Some theorists limit the motor component to subtle behavior, excluding coarse behavior. More often, however, theorists draw a distinction between "emotional" behavior, caused by their preferred mechanism of emotion causation, and instrumental behavior, caused by a goal-directed process (e.g., Scherer, 1984, 2001b, p. 374; but see Moors, 2017a). Although the contrast between emotional and instrumental speaks to the way in which the behavior was caused and therefore belongs in the section on causal-mechanistic explanations, it is nevertheless worth pointing out that it does not overlap with the contrast between subtle and coarse. Indeed, facial expressions can be purely instrumental (e.g., polite smile) and coarse behavior can be purely emotional (e.g., lashing out).

The feeling component, finally, refers to changes in conscious experience or feelings (see Dub, 2022; Frijda, 2005). This component has been split into (a) first-order experience or raw feelings, with an Intentional and/or phenomenal aspect, and (b) second-order experience or labeled feelings, consisting of the experience of having an emotion or a specific one (Block, 1995; Lambie & Marcel, 2002; Moors, 2017a; Scarantino, 2017a). Note that some authors only speak of first-order feelings (James, 1890b) and that some authors endorse the existence of unconscious feelings (e.g., J. J. Prinz, 2004a; Winkielman & Berridge, 2004).

Zooming in on first-order experience, many options are again possible. Not all theorists include an Intentional aspect (e.g., Reisenzein, 2012), but

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

for those who do, it refers to one or more of the other components in the emotional episode (cognitive, motivational, somatic, and/or motor) (e.g., Moors et al., 2013; Frijda & Parrott, 2011). Each of the components forms a potential ingredient for the Intentional aspect of emotional experience. For the phenomenal aspect of experience, two types of ingredients have been distinguished. The first type of ingredients are component-specific qualia. These can be seen as supervening on the Intentional aspects stemming from the other components. The second type of ingredients are emotion-specific qualia. These are qualia that cannot be reduced to the qualia that belong to components. The idea is, for instance, that fear feels intrinsically different from anger and that this is not (just) because fear is about danger and anger is about offense and/or because they involve different action tendencies and physiological responses (see Frijda, 2005). Theorists have sampled from these three ingredients – Intentional aspect, component-specific phenomenal aspect, and emotion-specific phenomenal aspect – to shape the content of first-order experience. Several combinations are possible: only the first two aspects, the first and the third aspects, all three aspects, only the second aspect, or only the third aspect. Many theorists, moreover, assume that a certain degree of blending takes place so that ingredients cannot easily be disentangled. Some aspects can still be foregrounded, however, if attention is focused on them (Lambie & Marcel, 2002).

2.2.1.2 Granularity of Components

Once the content of a component is chosen, theorists can treat it in a molar or a molecular way. A molar treatment describes a component with a single variable. A molecular treatment describes the component with a collection of variables that each refer to a separate aspect of the component, and the values on these variables combine to form a pattern (Moors & Scherer, 2013).

For instance, the appraisal process in the cognitive component can be described in a molar way with a single variable that has values such as danger, offense, and irrevocable loss, corresponding to what Lazarus (1991) called "core relational themes" of emotions. But it can also be described in a molecular way with multiple appraisal variables such as goal relevance (goal relevant vs. goal irrelevant), goal congruence (goal congruent vs. goal incongruent), certainty (certain vs. uncertain), expectedness (expected vs. unexpected), controllability (easy vs. difficult), agency (internal vs. external), and accountability (blame vs. credit) (C. A. Smith & Lazarus, 1993).

Likewise, action tendencies in the motivational component can be described in a molar way with a single variable that has values such as the tendency to approach, avoid, flee, fight, repair, and give in (e.g., Frijda et al., 1989). A molecular description of action tendencies, on the other hand, makes use of multiple variables such as level of activity (active vs. passive), direction of movement (towards vs. away from the eliciting stimulus), direction of adaptation (fit stimulus to self vs. fit self to stimulus), and target (self vs. other vs. inanimate object; Roseman, 2001).

The peripheral part of the somatic component has sometimes been described in a molar way, with values like shivering, boiling, blushing, and butterflies in the stomach, but more often in a molecular way with multiple variables such as heart rate, blood pressure, galvanic skin response, and muscle tension (Bauer, 1998; Kreibig, 2010; Lench et al., 2011; Mauss & Robinson, 2009; Stephens et al., 2010). The central part of the somatic component (brain activity) has been described in terms of brain areas (e.g., amygdala, anterior cingulate cortex, hypothalamus), brain circuits (e.g., thalamo-amygdala route, thalamo-cortico-amygdala route; LeDoux, 1996), and brain networks (e.g., Kragel & LaBar, 2016; Nummenmaa & Saarimäki, 2019). Single brain areas correspond more to a molar approach whereas brain circuits and networks fit more with a molecular approach. Another type of molecular approach specifies the activity in brain networks in terms of non-locationist parameters, such as the degree of neural synchronization (e.g., Kragel et al., 2018).

Turning to the motor component, the molar and molecular treatment of coarse behavior parallels that of action tendencies (molar: approaching, avoiding, fleeing, fighting, repairing, giving in; molecular: activity level, direction of movement, direction of adaptation, target). The subtle behavior in facial expressions takes on molar values such as smiling face, scowling face, startled face, and pouting face, and molecular values that correspond to muscle activity in certain parts of the face (i.e., action units; Ekman & Friesen, 1978). Vocal expressions take on molar values such as screaming and laughter, and molecular values that specify pitch, tempo, rhythm, pausing, loudness, and frequency perturbations (see review in Scherer & Moors, 2019). Gestural expressions take on molar values describing movements such as arm raising, head tilting, fist clenching (see review by Witkower & Tracy, 2019), and molecular values specifying the amount, speed, force, fluency, and size of these movements (Dael et al., 2013).

Finally, the feeling component can be treated in a molar way with values such as fear, anger, sadness, and joy. Molecular treatments describe feelings with variables such as valence, arousal, and dominance (Russell, 2003) or with variables describing aspects of the other components that are supposed to be projected into the feeling component. To illustrate the latter approach, a person can feel trapped, experience the urge to lash out, feel her blood boiling, and realize that she is frowning and clenching her teeth.

In principle, each of the variables in a molar or molecular description can be treated as discrete or dimensional. While the molar–molecular dichotomy refers to the number of *variables* used to describe an entity (single vs. multiple), the discrete–dimensional dichotomy refers to the number of *values* that a variable can take (finite vs. infinite). As the examples above show, however, "molar" often combines with "discrete" whereas "molecular" often combines with "dimensional." Indeed, in the molar approach to appraisal, the values danger, offense, and loss are discrete whereas in the molecular approach to appraisal, the variables of goal relevance, goal congruence, and controllability are readily understood as dimensions. A stimulus can be more or less goal-relevant, more or less goal-congruent, and more or less controllable.

Philosophical theories tend to take a molar approach to components. This is apparent, for instance, in the fact that they spell out the formal objects of emotions in terms of danger, offense, and loss. Psychologists more often adopt a molecular approach, probably encouraged by the need to develop ways to measure the components in empirical research. The two approaches are by no means incompatible, however. Molar descriptions can be considered as summaries of molecular descriptions. For instance, the molar appraisal value "danger" can be seen as a summary or pattern of the molecular appraisal values "goal-relevant," "goal-incongruent," and "difficult to control." So considered, molar and molecular approaches address different mereological levels (Teroni, 2021).

Molar and molecular approaches are not only possible with regard to the components of emotions but also with regard to the emotion itself. A molar approach treats emotion as a single variable that can take on values such as fear, anger, sadness, joy, and so on. A molecular approach breaks down emotions into multiple components (e.g., cognitive, motivational, somatic, motor, and feelings). Here too, molar and molecular approaches can be compatible if they are understood as addressing different mereological levels.

2.2.1.3 Number, Identity, and Necessity Status of Components

Constitutive explanations of emotion theories range from narrow to broad. The narrowest constitutive explanations identify emotions with a single component. Emotions are taken to be special types of cognitions (Neu, 2000; Nussbaum, 2001; Roberts, 2003; Solomon, 1993; Tappolet, 2016), special types of motivations (Roseman, 2013; Ryle, 1949; Shand, 1914), special types of brain mechanisms (Ekman, 1999a; Tomkins, 1962, 1963), special types of bodily responses (Watson, 1919), or special types of feelings (Barrett, 2006b; Descartes, 1649/1989; Hume, 1739; LeDoux, 2012b; LeDoux & Brown, 2017; Reisenzein, 2012; Schachter, 1964;

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

Wundt, 1897/1998). Broader constitutive explanations identify emotions with compounds of two or three components (Frijda, 2009; Green, 1992; Izard, 1972; Lazarus, 1991; W. Lyons, 1980; Marks, 1982; McDougall, 1908). The broadest constitutive explanations include all five components (Clore & Ortony, 2000; Dewey, 1894, 1895; Frijda & Parrott, 2011; Scherer, 2009b). Using cooking metaphors, several authors have proposed considering appraisal and behavioral expressions as ingredients of emotions rather than as causes and consequences. Thus, Ellsworth (2006) argued that appraisal is no more the cause of emotion than eggs are the cause of a cake, and Clore and Centerbar (2004) argued that emotion is not the cause of expression in the same way as eggs are not the cause of an omelet. Most proponents of all-inclusive constitutive explanations, however, do not demand that all of their components are present in each token emotion and they vary with regard to the component they do consider to be necessary. For Damasio (2004) the essential component is a special type of brain activity; Frijda (1986; Frijda & Parrott, 2011) picked out the motivational and feeling components ("felt action tendencies"), and A. Sloman and Croucher (1981) chose the cognitive and motivational components. Pushing this idea further, Clore and Ortony (2013) suggested that emotions require the simultaneous occurrence of any two components, regardless of which they are (see Scherer, 2005, p. 697, 2022, for a similar suggestion).

The preference of theorists for narrow vs. broad constitutive explanations can be attributed in part to the meta-theoretical principles that they endorse. Philosophers – with their strong focus on constitutive explanations – have the ideal to combine as many desiderata as possible *inside* the emotion. Theories that are (or run the risk of being) criticized for being unable to accommodate an important desideratum turn to one of four possible strategies (see Scarantino & de Sousa, 2018, for the first three). A first strategy is the elastic strategy, which consists in stretching the content of the favored component. For instance, when the cognitive component in judgmental theories was found wanting, the meaning of cognition was stretched from beliefs to information processing at large (see Scarantino, 2010).

A second strategy is the add-on strategy, which consists in planting extra components inside the emotion, resulting in a hybrid theory with a broader constitutive explanation. All multi-componential theories count as examples.

A third strategy is the alternate-components strategy, which consists in swapping the favored component for a different one, or in reshaping it (Scarantino & de Sousa, 2018). This usually results in an altogether different theory. One example is the trading of feeling theories for judgmental theories, which involves the swapping of the feeling component

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

by the cognitive component. Another example is the trading of judgmental theories by perceptual theories, which involves the reshaping of the cognitive component from judgment to perception.

A fourth strategy is the causes-consequences strategy, which consists in adding causes and consequences of emotions that cover for the missing desiderata. For instance, theories that equate emotions with motivations and/or feelings often treat cognition as the cause and behavior as the consequence of emotions (Reisenzein, 2012; Roseman, 2013; Scarantino, 2017a; Schachter, 1964).

Critical voices within philosophy consider the first two strategies to be poor choices. The elastic strategy leads to overinclusive meanings of entities (like cognition) so that they become trivial (see Scarantino, 2010; Scarantino & de Sousa, 2018). The add-on strategy is considered an easy cop-out that is unparsimonious (but see Goldie, 2012)[29] and requires an account of how the components hang together in the emotion (Soteriou, 2018). The alternate-components strategy is considered more promising because it allows theories to keep a narrow (and hence parsimonious) constitutive explanation. The causes-consequences strategy, finally, shares this asset but abandons the ideal to keep all the desiderata within the confines of the emotion itself.

Psychological theories – with their strong focus on causal-mechanistic explanations – may be expected to have a preference for the causes-consequences strategy. This strategy is indeed taken by some psychological theories (e.g., Reisenzein, 2012; Roseman, 2013). Many contemporary psychological theories, however, are multi-componential. They adopt an extreme add-on strategy, seemingly unhindered by a concern for parsimony. In their defense, however, it should be mentioned that they do not consider all of the included components to be strictly necessary and they do provide an account of how the various components hang together. Indeed, they do specify the causal relations between the components inside the emotion (thereby providing a constitutive mechanistic explanation, see Chapter 1).

For those theories that do not include all components but that push some of the relevant entities to the causes or consequences of the emotion, it makes sense to distinguish between the emotion proper and the broader episode in which the emotion occurs (Moors, 2007; Scarantino & de Sousa, 2018). Given that there is more consensus about the components to include in the broad episode than in the emotion proper, and the

[29] Goldie (2012) argued that instead of striving for parsimony, which dictates to not include any entity beyond necessity, one could also follow the opposite rule, which is to not exclude any entity beyond necessity, in line with Bishop Butler's wisdom that everything is what it is and not another thing (see Soteriou, 2018).

impression held by many psychologists that the demarcation of the emotion proper is an arbitrary matter, Russell (2003) proposed replacing talk of emotions with talk of emotional episodes and in this way to "reconstitute the explanandum" (Bechtel, 2008).

2.2.1.4 Problems with Traditional Components and a Proposal for Alternative Components

The variety across theories regarding the precise content and granularity of components poses a complication for the comparison of these theories. Other obstacles are that the components cannot be distinguished according to a single organizing principle, despite what some authors have suggested, and that they are not mutually exclusive on most readings. Scherer (2005) argued that all components have a different *function*, but it may be too restrictive to demand that all components have a function. For instance, the feeling component could be endowed with a monitoring function, but some authors prefer treating feelings as epiphenomena instead.

Frijda (2007b) argued that all components involve different *processes*, but not all components contain processes, at least if we stick to the idea that something counts as a process if it relates an input to an output (see Chapter 1). The cognitive component contains processes (linking stimuli to afferent representations) but the motivational component does not (it merely contains efferent representations).

Finally, Clore and Ortony (2000) argued that the different components correspond to different *ontological categories*, but this is not straightforward. A first problem is that the separation between mental components (cognitive, motivational, and feeling) and bodily components (somatic and motor) is ultimately artificial because a common way to close the mind–body gap has been to argue that mental components are implemented in the part of the body called the brain. To solve this problem, some authors have proposed to discard the somatic component altogether (e.g., Parrott, 2007) while many others have restricted the somatic component to the peripheral part of the somatic component (e.g., Scherer, 2009b; see above). Another problem is that the mental components do not admit of a clean separation among themselves. In their capacity as mental components, they all involve representations. The cognitive and motivational components can be distinguished on the basis of the format of their representations (afferent in the cognitive component, efferent in the motivational component), but they are otherwise not comparable given that the cognitive component contains processes (i.e., transitions between inputs and outputs) whereas the motivational component does not (i.e., just efferent representations). The feeling component, moreover, is often treated as the conscious part

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

Table 2.3. *Novel components*

<table>
<tr><th>observable input</th><th colspan="2">mental representations: unconscious (unfelt) or conscious (felt)</th><th>observable output</th></tr>
<tr><td>S: raw stimulus</td><td>[S]: afferent/stimulus representation</td><td>[O/R]: efferent representation</td><td>R: overt response</td></tr>
<tr><td rowspan="3">eS: raw external stimulus</td><td rowspan="2">[ecS]: external concrete stimulus representation</td><td>[O^{v}]: non-behavioral goal</td><td>sR: somatic response</td></tr>
<tr><td rowspan="2">[R]: abstract behavioral goal or behavior representation</td><td>mR: motor response</td></tr>
<tr><td>[eaS]: external abstract stimulus representation</td><td>smR: subtle motor response</td></tr>
<tr><td rowspan="2">iS: raw internal stimulus</td><td>[icS]: internal concrete stimulus representation</td><td>[r]: concrete behavioral goal or behavior representation</td><td>cmR: coarse motor response</td></tr>
<tr><td>[iaS]: internal abstract stimulus representation</td><td></td><td></td></tr>
</table>

of the other two mental components and hence as a subcomponent of these other components (see Moors, 2017a; see also Colombetti & Thompson, 2007, p. 59). Taken together, a clean separation in terms of ontological categories seems difficult to maintain.

Because of the problems plaguing the traditional list of components, I believe it is useful to propose an alternative list of components that fare better in terms of mutual exclusivity (see Table 2.3). However, given that the traditional components are so widespread in the literature, it will be hard to completely disconnect from them without also disconnecting from the literature. I will therefore tie the novel components back to the traditional ones and continue to bring the latter up in the next chapters. Each of the novel components listed below is accompanied by an acronym. As a general rule, mental components have squared brackets whereas overt components do not.

The first novel component is raw stimulus input (S): unprocessed physical stimuli that are (a) external (eS) or (b) internal to the body (iS). External stimuli are generated in the outside world. Internal stimuli have their source in the body of the subject, referring to the immediate outcomes or effects of the subject's own somatic and motor activity. Immediate outcomes may be resident in organs or muscles, in which case they can only be perceived via interoception or proprioception. But immediate outcomes can also be visual or auditory in nature. Strictly

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

speaking, these visual and auditory outcomes belong to the outside world again, and therefore come under the heading of external stimuli.

The second novel component is an afferent representation ([S]), which has a mind-to-world direction of fit. Afferent representations try to accommodate themselves to the physical world by extracting information from the world. The content of afferent representations can be characterized along the dimensions external-internal and concrete-abstract, resulting in the following four subtypes: (a) representations of *external* stimuli in terms of *concrete* sensory or perceptual features ([ecS]) (e.g., the coiling form of a snake, a snake); (b) representations of *internal* stimuli (somatic and motor activity) in terms of *concrete* sensory or perceptual features ([icS]) (e.g., throbbing sensation, muscle tension); (c) representations of external stimuli in terms of *abstract* features or categories ([eaS]) (e.g., danger, valence, goal congruence, controllability); (d) representations of *internal* stimuli (somatic and motor activity) in terms of *abstract* features or categories ([iaS]) (e.g., valence, emotion types).

The third novel component is an efferent representation ([O/R]), which has a world-to-mind direction of fit. Efferent representations try to assimilate the physical world to themselves by spurring the organism to overt behavior in the physical world. Efferent representations are goals, that is, representations of valued outcome. Goals can be organized in a goal hierarchy in which superordinate goals reside at the top (e.g., survival, autonomy, control, connectedness, happiness, identity) and branch out in ever more subordinate goals at intermediate levels (e.g., safety, food, partner) and further down in behavioral goals, also called action tendencies or intentions (e.g., approach, avoidance).[30] The level of action tendencies can again be split into various sublevels, ranging from more abstract action tendencies (e.g., tendency to avoid) to more concrete action tendencies or motor programs (e.g., tendency to flee or run away). In sum, subtypes of efferent representations are non-behavioral goals ([O^v]), (b) abstract behavioral goals ([R]), and (c) concrete behavioral goals or motor programs ([r]).[31]

[30] I do not follow Frijda's (1986; Frijda & Parrott, 2011) proposal to distinguish between (a) action tendencies, which are reserved for emotions and which are endowed with control precedence (i.e., they take priority over other goals and persist in the face of obstacles), and (b) intentions to act, which are reserved for non-emotional, instrumental action. Regardless of how they are caused, action tendencies and intentions are both goals to act. This means that they both have dynamic properties such as persistence in the face of obstacles (see Chapter 1). Whether a goal takes priority over other goals depends on its value, but this is a matter of degree, which obviates the need to use two different terms.

[31] The difference between [R] and [r] is fluid. I will only use the [r]-symbol when I need it to mark the distinction with [R].

Goals at lower levels have a means–end relation to goals at higher levels if they belong to the same branch. For instance, running away serves avoidance, avoidance serves safety, and safety serves survival. Other assumptions are that goals at higher levels are valued more, are more abstract, and are more likely to be innate and hence universal, whereas goals at lower levels are valued less, are more concrete, and are shaped by learning and hence more culturally specific and idiosyncratic. The idea that subordinate goals are valued less squares with the observation that they are more easily replaced by other goals. For instance, if safety cannot be reached via the tendency to run away, it may be replaced by the tendency to hide, whereas the goal for safety itself is harder to replace by another goal.

As discussed in Chapter 1, the common-coding hypothesis (W. Prinz, 1997) states that when the content of a representation is a behavior, it does not matter whether the representation is afferent or efferent. Thinking about or observing a behavior (i.e., afferent representation) activates the same brain regions as preparing to engage in the behavior (i.e., efferent representation). In other words, if the content of a representation is a behavior, the difference between afference and efference falls into oblivion. The implication is that [R] can be seen as a subtype of [aS] and [r] as a subtype of [cS].

The fourth novel component contains somatic responses (sR), which I restrict to peripheral autonomous responses and the central and endocrine activity that supports these responses. Needless to say, each of the other components has its own supporting central and endocrine activity.

The fifth novel component, finally, contains overt motor responses (mR), consisting of both (a) subtle behavior in the face, the voice, and gestures (smR) and (b) coarse behavior as in full-body behavior (cmR) (see Chapter 9, for a more fine-grained taxonomy of behavior).

Linking the five traditional components to the five novel ones can be done as follows. The information processes in the cognitive component translate raw stimulus input (S) into afferent mental representations ([S]). The motivational component houses efferent representations, and in particular, abstract ([R]) and concrete ([r]) behavioral goals. The (peripheral part of the) somatic component aligns with the somatic responses (sR) and the motor component with the motor responses (mR). The feeling component can easily be matched with those parts of the afferent and efferent representations ([S], [R]) that pervade consciousness. If it is assumed that only representations can become conscious (Moors, 2017a), raw external and internal stimuli (eS, iS) can only enter consciousness after they have been processed and have become the content of an afferent representation ([eS], [iS]). As mentioned, not all scholars subscribe to this view. Some have decoupled consciousness from

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

representations considering feelings to be a purely phenomenal affair (e. g., Reisenzein, 2012), and some allow feelings to be unconscious (e.g., J. J. Prinz, 2004a). For ease of communication, I will continue to use the term "components" to refer to the traditional components, unless specified otherwise or when there is a risk of confusion.

2.2.2 Causal-Mechanistic Explanations

The next source of variation among emotion theories concerns the causal-mechanistic stories that they present (**Axis 6**). On a coarse level, theories vary in the order in which they place the components within the emotional episode (**Axis 6a**).[32] Inspection of the proposed orders already teaches us a few things. One thing that stands out is that theories vary in the components that they ultimately or primarily want to explain, that is, their *primary explanandum* (**Axis 6a1**). Some theories work more towards explaining bodily (somatic and motor) components (e.g., evolutionary theories) whereas others focus more on feelings (e.g., psychological constructionist theories). The distinction is not a hard one. Most theories also provide an account of their non-primary components.

Another thing we can learn is that theories vary in the component(s) they take to be the *explanans* (or *explanantia*), that is, the component(s) in which they place their proposed mechanism (**Axis 6a2**). Mechanisms to explain bodily components have typically been placed in the connection between the cognitive and the motivational component. Mechanisms to explain the feeling component, on the other hand, show more variety. A noticeable issue of disagreement is whether feelings are generated (a) centrally, that is, directly by the mental components (i.e., cognitive and/or motivational), (b) peripherally, via feedback from the bodily components (i.e., somatic and motor), or (c) both centrally and peripherally (**Axis 6a2*, FL**).

Besides analyzing mechanisms on a coarse level in terms of "traditional" components, we can also analyze them on a more fine-grained level in terms of "novel" components (**Axis 6b**). Let us briefly consider the explanation of the bodily components via the cognitive component in connection with the motivational component. This path can be split into a cognition step and a transition step from cognition to motivation. In both steps, mechanisms proposed by theories differ regarding the content of their inputs and outputs.

[32] Ellsworth (1994, p. 228) wrote that "debates about the primacy of cognition, bodily responses, or feelings make little sense when emotions are considered as a stream." I would argue that even if components follow each other in close temporal succession, there is still merit in figuring out their logical order.

In the cognition step, options are (a) a perceptual process, which links a raw stimulus input to a representation of concrete stimulus features (S→[cS]) (e.g., James, 1890b; certain evolutionary theories), or (b) an evaluation process, which links a raw or perceptually preprocessed stimulus to abstract stimulus features such as appraisals (S/[S]→[aS]) (e.g., certain stimulus evaluation theories) (**Axis 6b1**).

In the transition step, options are (a) a stimulus-driven process, which consists of the activation of an association between an afferent representation and an efferent representation ([S–R]) (e.g., the majority of emotion theories), and (b) a goal-directed process, in which the expected utilities of one or more action options are weighed up and the action option with the highest expected utility activates its corresponding action tendency ($[S{:}R_1\text{–}O^v/R_2\text{–}O^v] \rightarrow [R]$) (e.g., the goal-directed theory). The contrast between stimulus-driven and goal-directed processes is one of the central, if not *the* most central contrast between theories that try to explain behavior-related components of emotional episodes (**Axis 6b2, FL**). I therefore describe it at length in Box 2.1.

In addition to the contents of inputs and outputs, mechanisms can also vary in terms of the format of the representations (**Axis 6c**), the operations handling these representations (**Axis 6d**), and/or the conditions under which the mechanisms occur (**Axis 6e**). Options for formats discussed so far are perceptual vs. conceptual vs. associative vs. propositional, and afferent vs. efferent; options for operations are associative vs. rule-based, simple vs. complex, and/or primary vs. secondary vs. tertiary; and options for operating conditions are poor vs. ample, related to automaticity vs. non-automaticity. A final source of variation worth pointing out is whether theories propose a special-purpose mechanism dedicated to emotions or whether they believe that emotions are brought about by general-purpose mechanisms and therefore can be reduced to other entities (**Axis 6f, FL**).

BOX 2.1 Stimulus-Driven vs. Goal-Directed Processes

Behavior theories distinguish between two processes of behavior causation: stimulus-driven and goal-directed processes. The behavior caused by a stimulus-driven process is said to be a "reaction"; the behavior caused by a goal-directed process is said to be an "instrumental action." This box clarifies (a) my usage of the terms stimulus-driven and goal-directed process, (b) common routes to install both processes, (c) ways to diagnose them, and (d) ways to understand the interplay between them (Balleine & Dickinson, 1998; Dickinson & Balleine, 1994; Heyes & Dickinson, 1990; Moors et al., 2017; T. W. Robbins & Costa, 2017).

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

What

A stimulus-driven process is one in which a stimulus activates the association between the representation of a stimulus (with concrete or abstract stimulus features) and the representation of a response. The latter representation then translates in overt behavior. This process can be depicted as follows:

$$S \rightarrow [S\text{–}R] \rightarrow R$$

A goal-directed process is one in which an organism selects a response based on the expected utilities of the available response options. The expected utility of one response option depends on the value of the outcome of this response option and the expectancy that this outcome will be attained. The minimal requirement for a process to count as goal-directed is that the expected utility of one response option is processed. If more response options are available, the response option with the highest expected utility is selected. Once a response option is selected, it activates its corresponding action tendency or intention, which subsequently translates in overt behavior. This process can be depicted as follows:

$$\begin{bmatrix} S{:}R_1 — O^v \\ S{:}R_2 \diagup \end{bmatrix} \longrightarrow [R_1] \longrightarrow R_1$$

Moors et al. (2017) recently proposed to embed this goal-directed process in a broader goal-directed cycle. This cycle can be split into three large phases: (a) the detection of a discrepancy between a stimulus and a goal, (b) the selection of a strategy or behavior to undo the discrepancy, and (c) the feedback of the outcome of the behavior to the start of the cycle. The detection and feedback phases are drawn from feedback theories of motivation (also called cybernetic theories or control theories; Carver & Scheier, 1985, 2000, 2002; MacKay, 1951; Miller et al., 1960; Powers, 1973a; Wiener, 1948). The selection phase in the middle is drawn from expectancy-value theories of motivation (Tolman, 1932), which are similar to expected utility theories of decision making in behavioral economics (see Camerer & Loewenstein, 2004) and belief-desire theories of intentional action in philosophy (Davidson, 1985a). A full description of the goal-directed cycle will be deferred until Chapter 7 (see also Figure 7.1).

In keeping with James's (1890b) and W. Prinz's (1997) idea that response representations are always efferent or dynamic (see Chapter 1), the [R] in the [S–R] association is not different from an action tendency or intention to act. From this, it follows that behavior caused by a stimulus-driven process is also intentional behavior, as it is caused by the intention to engage in the behavior. Apart from noting that this intention need not be a conscious intention, it is important to stress that intentional behavior is not the same as goal-directed behavior, at least not according to the strict usage of goal-directed adopted here. Only if the intention that causes the behavior is itself caused by representations of values and expectancies of the outcomes of the behavior, does the behavior qualify as goal-directed. If the intention is caused by the mere representation of a stimulus, the behavior qualifies as stimulus-driven instead.

Routes for Installation

The [S:R–O^v] representations in goal-directed processes can be computed online (i.e., tertiary process). Alternatively, they can be retrieved from memory if they were previously installed by learning (i.e., secondary process), more in particular, a "moderate" operant conditioning procedure in which execution of a response in the presence of a stimulus is followed by an outcome on a moderate number of occasions. A few examples: Typing in one's password when the opening screen is on gives access to the computer; making jokes at a party breeds popularity; lending a helping hand to a person in need leads to social approval; and drinking a beer after work leads to feeling relaxed. In addition to learning through direct experience of the outcomes, the contingencies may also be picked up via observation, instruction, imagination, or logical inference.

The [S–R] representations in stimulus-driven processes can be innate (i.e., primary process) or learned (i.e., secondary process). Innate connections between unconditioned stimuli (USs) and unconditioned responses (URs) used to be called instincts (Darwin, 1872; James, 1890b), but now go by the names of reflexes or fixed reaction patterns (LeDoux & Daw, 2018). A few examples are loud noises leading to a startle response, the sudden loss of support leading to contraction of the limbs, and seeing a predator leading to a flight response. Learned [S–R] connections are called habits. They can be installed via three learning procedures. The first is an "extensive" operant conditioning procedure in which execution of a response in the presence of a stimulus is followed by an outcome on an extensive number of occasions. The idea is that as long as an S:R–O^v contingency is trained on a moderate number of occasions, the [S:R–O^v] link remains intact, but that once the number of repetitions becomes extensive, the outcome representation is discarded or no longer activated ([S:R–$Ø^v$]) so that only an [S–R] association is left and stamped in (i.e., Thorndike's, 1911, law of effect; Wood & Rünger, 2016). For instance, if typing in one's password upon seeing the opening screen of the computer was repeatedly followed by access to the computer, the mere sight of the opening screen should come to trigger the tendency to type in the password without envisaging access to the computer. Likewise, if drinking beer after work always leads to relaxation, coming home from work should trigger the tendency to drink beer without envisaging relaxation.

The second learning procedure is a purely associative or Hebbian conditioning procedure in which a stimulus is repeatedly followed by a response without being followed by an outcome. The mere co-occurrence may suffice to forge a link between an [S] and an [R], albeit a weaker link than if it were followed by an outcome.

A third learning procedure is a classical or Pavlovian conditioning procedure in which a first stimulus that did not originally lead to a response is consistently paired with a second stimulus (US) that did already produce a response (UR) via an [S–R] connection.[1] After this procedure, the mere presentation of the first stimulus, which is now a conditioned stimulus (CS) also produces a response, called a conditioned response (CR). For instance, if loud noises (US) naturally lead to fleeing (UR) in wild cattle, the regular pairing of hunters with gun shots may over time cause the animals to flee (CR) upon the mere sight of the hunters (CS). The CR does not have to be similar to the UR, like in the wild cattle example, but it can also be anticipatory, opponent to, or just different from the UR (see Moors, 2017b). Finally, each of the three conditioning procedures involved in establishing [S–R] associations could happen via direct experience, observation, instruction, or even inference.

Diagnosis

Goal-directed processes can flexibly adapt to changes in the outcomes of behavior whereas stimulus-driven processes cannot. If the outcome of a response is no longer valued or a response is no longer followed by the same outcome, [S:R–O^v] links get updated, thus enabling a flexible adjustment of the behavior. For instance, a person who learned that making jokes leads to social approval from her colleagues may stop making jokes when she discovers that her new colleagues have no sense of humor. If only an [S–R] link is put in place, changes in actual outcome values and actual response-outcome contingencies will not affect this link, thus producing maladaptive behavior, called "action slips" or "bad habits." The typing in of one's old password after it has expired is an example of an action slip. The persistent drinking of alcohol when it only leads to bad outcomes is an example of a bad habit.

Based on this rationale, the most common procedures to diagnose whether a behavior is caused by a goal-directed or stimulus-driven process are to devalue the outcome of the behavior (i.e., devaluation method; Adams & Dickinson, 1981) or to degrade the behavior-outcome contingency (i.e., contingency degradation method; Balleine & Dickinson, 1998). If the behavior remains the same, it is inferred that outcome knowledge was not represented so that it could not get updated, and hence that the behavior was caused by a stimulus-driven process. If the behavior decreases, it is inferred that outcome knowledge was represented and updated, and hence that the behavior was caused by a goal-directed process.

Using devaluation and contingency degradation methods, animal research shows that reliance on habits is not ubiquitous but instead depends on a host of conditions (e.g., extensive training, absence of a choice procedure, interval reward schedule) and is short-lived (Adams, 1982; Colwill & Rescorla, 1985; Dickinson, 1985; T. W. Robbins & Costa, 2017). In humans, moreover, evidence for the role of habits is hard to find (e.g., failures reported by de Wit et al., 2018) and studies that do report habits (Hardwick et al., 2019; Schwabe & Wolf, 2011) remain open to alternative, goal-directed explanations (Buabang, Boddez, et al., 2021; Buabang, Köster, et al., 2021; see below).

Interplay

Traditional dual-process models propose a default-interventionist architecture with stimulus-driven processes as the default and goal-directed processes as the intervenor (Hofmann et al., 2009; Wood & Neal, 2007; Wood & Rünger, 2016). This is rooted in the trade-off between automaticity and adaptiveness discussed in Chapter 1. Stimulus-driven processes are seen as simple and therefore

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

automatic, which means that they can operate under poor conditions and therefore run by default. Goal-directed processes, on the other hand, are seen as more complex and therefore non-automatic, which means that they can only operate under ample conditions. At the same time, the simplicity of stimulus-driven processes also makes them more rigid, which in turn makes them more likely to produce maladaptive behavior. This is why the more complex and therefore more flexible goal-directed process must sometimes intervene to correct the course of action and make it more adaptive, but this is only possible when operating conditions are ample. Thus, if operating conditions are poor, the agent has no other choice but to switch to the stimulus-driven process. Note that the alignment between all these dichotomies turns these dual-process models into dual-system models.

Let us pause to closely analyze the step from the rigidity of stimulus-driven processes to maladaptive behavior. As explained in Chapter 1 (Section 1.3.4), a process of behavior causation is practically irrational in the *process-sense* if it does not take into account (objectively accessible) information about the outcomes of behavior options. Given that a stimulus-driven process does not consider any outcome of any behavior option, it can safely be regarded as irrational in this sense. That being said, the behavior produced by a stimulus-driven process may still satisfy a person's goals, and hence still qualify as rational in the *output-sense*. This may be the case, for instance, when the stimulus-driven process grew out of a previous goal-directed process and the reward structure of the environment did not change. For instance, if typing in one's password is caused by a stimulus-driven process, it remains adaptive as long as the password does not expire.

Recently, Moors et al. (2017; Moors, 2017a, 2017b, 2019; Moors & Boddez, 2017; Moors & Fischer, 2019) proposed a dual-process model with a parallel-competitive architecture in which both processes operate in parallel and compete with each other, and in which the goal-directed process often wins the competition. The root of this model is the idea that not only stimulus-driven but also goal-directed processes can be automatic (see Moors et al., 2017). If both processes are automatic, there should be constant competition between them, also under poor operating conditions. The model, moreover, assumes that this competition is often won by the goal-directed process because this process combines more benefits (automaticity and adaptiveness) than stimulus-driven processes do (only automaticity), and the system should prioritize the process with the most benefits (Moors et al., 2017). The implication is that goal-directed processes should be the default determinant of behavior while the stimulus-driven process should determine behavior only in exceptional cases (see Chapter 7). This is exactly opposite to what the default-interventionist model suggests.

Much of the competition in the parallel-competitive model, however, will be between different goal-directed processes serving different goals. One example is the competition involved in action slips, such as when typing in an old password competes with typing in a new password. This may not reflect the competition between a stimulus-driven tendency to type in the old password and a goal-directed tendency to type in the new password. Instead, it may reflect the competition between two goal-directed processes, one based on the old

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

contingency and the other based on the new contingency (Buabang, Köster, et al., 2021).[2] Another example is the struggle with temptation involved in bad habits. This may not reflect the competition between a stimulus-driven tendency to indulge (e.g., drinking alcohol) and a goal-directed tendency to suppress indulgence (e.g., abstain). Instead, it may reflect the competition between two goal-directed processes, one based on the contingency between indulgence and a short-term outcome (e.g., relax) and the other based on the contingency between abstinence and the preservation of a long-term outcome (e.g., health).

The flexibility of goal-directed processes to adapt to changes in current outcome values makes them more practically rational in the process-sense than stimulus-driven processes, which cannot adapt in this way. This does not mean that goal-directed processes are always flawless. The fact that they rely on subjective values and expectancies makes them vulnerable to errors. This is one of the reasons why they do not always lead to goal satisfaction, and hence are not always rational in the output-sense. Finally, it may be good to keep in mind that even the most Olympian rational decision may be thwarted by unforeseeable outcomes caused by external obstacles.

As I will try to demonstrate in this book, most emotion theories, with the clear exception of the goal-directed theory (Moors, 2017a), endorse a dual-system model with a default-interventionist architecture.[3] In particular, they put forward (a) a stimulus-driven process to explain emotions and (b) a goal-directed process to explain non-emotional instrumental behavior and emotion regulation. The goal-directed theory of emotion, on the other hand, advocates a dual-process model with a parallel-competitive architecture in which (a) the goal-directed process is the default determinant of non-emotional as well emotional behavior, and (b) the stimulus-driven process is the exception.

[1] Note, however, that a classical conditioning procedure may also connect the first stimulus to a second stimulus that is involved in an [S:R–O^v] connection rather than in an [S–R] connection (Moors, 2017b).

[2] Note that this interpretation defies the assumption that extensive operant conditioning leads to the discarding of the outcome representation in an [S:R–O^v] link (so that it turns into an [S–R] link). The alternative interpretation is that the extensive training of a given [S:R_1–O^v] link makes the [R_1–O^v] connection grow stronger instead of weaker. If the reward structure of the environment changes so that a different [R_2] now leads to the [O^v], the old (now non-valid) [S:R_1–O^v] link has to compete with the new (now valid) [S:R_2–O^v] link. Action slips occur when the competition is won by the old [S:R_1–O^v] link. Support for this interpretation was recently obtained in our lab (Buabang, Köster, et al., 2021).

[3] Some traditional dual-process models propose a parallel-competitive architecture in which stimulus-driven and goal-directed processes start to operate in parallel, but in which the stimulus-driven process often wins the competition because the goal-directed process is too slow. The net result is the same as in the default-interventionist architecture, however: The stimulus-driven process is the default determinant of behavior and the goal-directed process can only occasionally change its course (see Wood & Rünger, 2016). I will therefore group it together with the default-interventionist view.

2.3 Stage 3: Validation

In the validation stage, researchers carry out empirical studies to test the explanatory success of their explanations, that is, whether the explanantia in their explanations can indeed be linked to the explanandum specified in the working definition. Returning to the water example from Chapter 1, empirical research can take samples of water and check whether the molecular structure of these samples is indeed H_2O or it can combine two hydrogen molecules and one oxygen molecule and check whether the product indeed looks like water. During this stage, researchers may also pause to evaluate the internal and/or external consistency of the explanation. In the water example, this would amount to asking whether it makes sense at all to hypothesize that two hydrogen molecules and one oxygen molecule combine into the molecule H_2O based on the rules of combining atoms and what we know from previous research. Empirical research in the emotion domain is carried out by psychologists and neuroscientists. This research is mainly focused on testing causal-mechanistic explanations rather than constitutive explanations.

Finally, it is worth pointing out that empirical research is not restricted to theory testing but may also foster theory building. Empirical research programs not only serve to test the explanatory success of the proposed mechanistic explanations (i.e., falsification or confirmation) but also to fill in the gaps in these explanations (i.e., exploration). Most theories present a mechanistic sketch for which the details remain to be worked out. For instance, evolutionary theories propose that emotions have dedicated brain mechanisms, but their precise implementation remains to be fleshed out. Similarly, appraisal theories emphasize the role of appraisal in emotion causation but hypotheses about precise links between appraisals and emotions remain open to empirical adjustment. Box 2.2 describes the most commonly used methods in emotion research.

2.4 Stage 4: Scientific Definitions

The explanations that theories propose for a phenomenon provide the input for candidate scientific definitions of this phenomenon. These explanations deliver criteria for both the demarcation and the partitioning of the set under consideration. In the water example of Chapter 1, the constitutive explanans of H_2O allowed us to replace the working definition of water as transparent, odorless fluid by a scientific definition that equates water with H_2O. It further makes sense to divide the set of water into gasiform, fluid, and solid subsets based on the criterion of temperature. Applied to the case of emotion, the constitutive and/or mechanistic explanations of emotions deliver criteria for the demarcation of the set of emotions as well as criteria for the set's partitioning.

Ideally, the proposed explanations should first have been thoroughly validated in Stage 3 and only those that earned wide consensus should be

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

allowed to infuse the scientific definitions of emotion. The reality, however, is that psychologists and philosophers are easily convinced by the sometimes weak empirical evidence for their theories, and that consensus is still a far dream. The field is characterized by an enormous multiplicity of theories that vie for dominance. This is partly because psychology and naturalistic philosophy are relatively young disciplines but also because the topic of investigation poses exceptional challenges for measurement: mental states that are not directly observable (see Box 2.2). Against this background, emotion theorists rush to put their scientific definitions out in the hope to already inspire practitioners and policy makers before theorists with less luminous insights do (Fang & Casadevall, 2015).

2.4.1 Intensional Scientific Definition

2.4.1.1 Demarcation

An intensional definition of emotion specifies what the members of the set of emotion have in common (i.e., necessity) and in which way they differ from the members of others sets (i.e., sufficiency). As emotion theories propose different constitutive and causal-mechanistic explanations, they naturally differ in the criteria for demarcation that they put forward (**Axis 7**). Constitutive explanations yield as criteria the presence and nature of certain components. Causal-mechanistic explanations yield as criteria causal relations among certain components. A more extensive treatment of these definitions will be deferred until the next chapters.

2.4.1.2 Adequacy

To determine the adequacy of a candidate scientific intensional definition, the criteria for demarcation must be held against the light of the meta-criteria of similarity and fruitfulness (Carnap, 1950; see Chapter 1). To reiterate, the similarity meta-criterion requires the candidate scientific definition to have a high degree of overlap with the working or folk definition, carving out a similar extension. This can be evaluated with the help of the properties of emotions listed as desiderata in the working definition. However, theories are free to regiment or trim their scientific definitions and reject certain properties as only apparent but not real. Ideally, however, they should still be able to account for these apparent properties, that is, explain why people associate them with emotions. This can be captured with the term "apparent-similarity meta-criterion." This meta-criterion is a first filter. If a theory cannot account for the apparent properties of emotions, it does not qualify as a theory of emotion in the first place.

The fruitfulness (annex simplicity) meta-criterion requires scientific extrapolation in the long run, which dictates that a deep (and preferably simple) source of homogeneity should be discovered among the members of the set of emotions in the present. This meta-criterion entails evaluating whether the explanantia proposed by a theory can provide such a common denominator.

Unlike the working definition of water, however, the extension of the working definition of emotion turns out to be heterogeneous (Russell, 2003). The phenomena that people call emotions range from acute states like terror and rage to pensive states like regret and dreamy admiration (see Section 2.1). As a consequence, a trade-off between the meta-criteria of (real) similarity and fruitfulness seems inevitable. Different theories have proposed different trade-offs (**Axis 8, FL**). Some maximize similarity at the cost of fruitfulness whereas others do the reverse.

Theorists who choose to prioritize similarity over fruitfulness give up the quest for a classic intensional definition and turn to a cluster-type or syndrome definition in which no condition is individually necessary, but some subcollection is sufficient (e.g., Averill, 1980; Clore & Ortony, 2000; Parkinson, 1995; Scherer, 2005, 2022; Shaver et al., 1987). The members in a cluster set share a sufficient number of conditions with the other exemplars or with a prototype (Boyd, 2010). Because cluster-type definitions are less fruitful than their classic counterparts, the theories discussed in the following chapters all somehow strive for a classic definition, even if several individual theorists at some point have resigned themselves to a cluster-type definition.

Theorists who choose to prioritize fruitfulness over similarity have turned to one of the following four options, which range from subtle to more radical.[33] A first option is to relax the similarity criterion but to safeguard the initial set by trimming or pruning a small part of the initial heterogeneous set so that it becomes more homogeneous (cf. "fish" in Chapter 1). For instance, some theorists toss out surprise and interest because these have no marked valence, and some theorists discard disgust because it does not seem to have a rich enough Intentionality. This "regimenting" of vernacular terms is so common that it is taken for granted.

A second option is to again make changes to the extension of the folk set but to replace it with a novel set. For instance, Charland (2005b) replaced the vernacular set of emotion with a novel set labeled with the capitalized neologism EMOTION and demarcated it from other sets such as cognition by pointing at the unique property of valence. This indicates, on the one hand, that he denied scientific status to the original folk set labeled as "emotion." Yet the kinship between the two labels suggests that he was still after some similarity and that rather than eliminating "emotion," he tried to reconstitute it (see Bechtel, 2008).

[33] Note that some scholars not merely relax the similarity criterion but reject it outright. For instance, Skinner (1945, 1977) famously argued that natural language and common sense are always misleading and that science should free itself from their shackles. Rather than starting from a folk set, Skinner's (1945, 1977) ideal starting point is language-independent phenomena, labeled with neologisms, that are formed with the sole purpose of being fruitful.

A third option is to split the heterogeneous folk set into several smaller sets and to provide a separate intensional definition for each (cf. "jade" in Chapter 1). One example is Panksepp's (2012) proposal to shift the focus from the set of emotions as a whole to individual emotions labeled with the capitalized neologisms of FEAR, RAGE, PANIC/GRIEF, LUST, CARE, PLAY, and SEEKING. Another example is Griffiths' (1997, 2004a) proposal to split the set of emotions into primary emotions (i.e., those that have a dedicated brain basis) and secondary emotions (i.e., those without such a basis). Although both types of emotions can be mixed, in their pure forms they have nothing in common (i.e., disunity thesis; see Hutto et al., 2018).

A fourth, eliminativist, option is to abandon the set of emotions altogether (cf. "air" in Chapter 1; e.g., Duffy, 1941a, 1941b; Fridlund, 1994; Meyer, 1933; Moors, 2017a; Russell, 2003). Just like chemists abandoned the set of air and focused on the molecules in tokens of air (e.g., oxygen, nitrogen, carbonite, and argon), some emotion scholars have abandoned the set of emotions and focus on the components in emotions. Note that this move is not the mere shift to a lower level of analysis. The molecules in air also occur in things that are not air (e.g., oxygen is everywhere and nitrogen is present in soils and food protein) and so it is for the components in emotions. Evaluation, action tendencies, responses, and feelings also occur in non-emotional episodes. If the folk set of emotion cannot attain a scientific status, it could still be worthwhile to study people's usage of the term, but that is an entirely different enterprise (see Russell, 2003; Scarantino, 2012b).

2.4.2 Divisio Scientific Definition

2.4.2.1 Partitioning

Traditionally, emotion theories have been contrasted on the axis discrete-dimensional (**Axis 9, FL**), an axis that pertains to the divisio definition of emotions. For instance, evolutionary theories endorse a discrete view; network theories and appraisal theories come in a discrete and a dimensional variant. Psychological constructionist theories have long been portrayed as dimensional because they go against a purely discrete view and because they foregrounded the dimensions of valence and arousal. In reality, however, some of them combine both views (e.g., Barrett et al., 2009, p. 433), whereas others make no commitments regarding this axis (e.g., Russell, 2003).

As explained in Section 2.2.1.2, the dichotomy between discrete and dimensional refers to the number of values that a certain variable can have: discrete variables have a countable number, dimensional variables an uncountable number. But the labels discrete and dimensional as they are used in relation to the divisio definition of emotion conceal other differences. They are usually conflated with molar and molecular

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

approaches to emotion, which refers to the number of variables used to describe the emotion: a single one or multiple ones. Thus, proponents of a discrete view place emotions on a single discrete variable (combining molar and discrete) whereas proponents of a dimensional view place them in a space shaped by multiple dimensional variables (combining molecular and dimensional).[34] To illustrate, discrete theories such as evolutionary theories assume that the variety between emotions can ultimately be tied to a limited number of dedicated brain mechanisms. Dimensional theories such as certain appraisal theories, on the other hand, believe that each conceivable emotion type has its own position in a multidimensional space made up of dimensional appraisal variables (e.g., goal relevance, goal congruence, certainty, expectedness, controllability). Given the infinite variety within this space, the most parsimonious way to describe this variety is by identifying the contributing dimensions.

It is worth pointing out that a discrete view does not preclude acknowledging infinite variety. In fact, it will be hard to find a single discrete theorist who believes that the chosen discrete emotion types capture all the variety that exists in emotion land (see Ekman, 2016; J. Lange & Zickfeld, 2021). Most discrete theorists agree that the variety is infinite, yet they believe that there is at least one principle that allows organizing this variety into separate bins, and that these bins happen to correspond (more or less) to vernacular emotion types. In addition, they often believe that this grouping variable hangs together with other variables so that the overall variety within subsets is smaller than between subsets, all other things being equal. Conversely, a dimensional view does not preclude identifying discrete subsets either. Dimensional theories could easily create "fluid" bins on the basis of each of the variables that they consider.

Nevertheless, I believe that the above considerations do not erase the distinction between (a) a discrete view, which posits that the variety in the set of emotions is best captured with discrete emotion subsets, and (b) a dimensional view, which posits that this variety is most parsimoniously described with dimensions. If the distinction is understood in this way, most theories fall on one side of the divide. Still, some theories integrate both views in that they consider the preference for discrete subsets and dimensions itself as a relative matter (i.e., as depending on the component under consideration; e.g., Barrett et al., 2009; see Chapter 8).

2.4.2.2 Adequacy

To determine the adequacy of a candidate scientific divisio definition, the criteria for organizing the set of emotions may be held up against the meta-criteria of similarity and fruitfulness again. Here, similarity is high if

[34] Other options, such as (a) a single dimensional variable (combining molar and dimensional) and (b) multiple discrete variables (combining molecular and discrete) are less common.

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

a theory's explanation succeeds in generating the prototypical emotion subsets from the working definition. This is straightforward for discrete theories. Dimensional theories do not naturally generate these prototypical subsets, though they may still be able to make sense of them, that is, pass the apparent-similarity test.

Fruitfulness is high if the principles for partitioning the set of emotions allow for meaningful extrapolations. In the water example, the division into solid, fluid, and gasiform H_2O seems fruitful. Predictions for one member of the subset of solid H_2O probably generalize to other members of this subset. As this example reveals, this also requires that the subsets themselves are homogeneous. Turning to emotions, discrete theories maintain that fruitfulness is covered by discrete emotion subsets, whereas dimensional theories reckon that fruitfulness is much higher if dimensions are used as the organizing principle. Still, discrete theories may experience a tension between similarity and fruitfulness for certain emotion subsets. To solve this tension, they may again revert to the options stipulated above: prioritize similarity at the cost of fruitfulness or prioritize fruitfulness at a cost of similarity (**Axis 10**). To illustrate the latter strategy, let us consider Cova and Deonna's (2014) quest for a scientific definition of the emotion of "being moved." The authors started from a folk set in which the members – collected from a survey – included emotions elicited by acts of generosity, births, weddings, love declarations, a victory against all odds, a lost son returning home, mass manifestations, and so on. The researchers discovered that what connected most members was that an important positive value (e.g., friendship, honesty, courage) stood out, either because it emerged against the background of negative stimuli (e.g., the winning against all odds, a lost son returning home), because it was extraordinary (e.g., acts of generosity, births, love declarations), or because a ceremony was put in place to celebrate it (e.g., weddings). Once the researchers settled on this common denominator, taking it as the basis for the scientific definition of the set, they went on to exclude members – listed by some participants of their survey – that did not conform to this definition, such as emotions elicited by negative events such as the sight of a homeless person, a famished child, or other tragic news. This means that the researchers decided to trim the initial heterogeneous set, sacrificing part of the similarity criterion to the benefit of homogeneity and hence fruitfulness.[35] Table 2.4 provides a summary of axes, values, and fault lines of emotion theories.

[35] Thus, even though empirical research was conducted to keep "conceptual and phenomenological exploration from the armchair [. . .] in check" (Cova & Deonna, 2014, p. 449), it seems that the armchair did have the final say in how the set was eventually demarcated.

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

Table 2.4. *Axes, values, and fault lines*

axes, fault lines		values on axes
Axis 1, FL	prospect about whether the folk set of emotion will turn into a scientific set	vindicator, skeptic
Axis 2, FL	prospect about whether vernacular emotion subsets will turn into scientific subsets	vindicator, skeptic
Axis 3	working definition: desiderata	
Axis 3a	intensional working definition	see Table 2.1
Axis 3b	divisio working definition	restricted vs. extended list
Axis 4	types of explanations provided	constitutive, causal, constitutive mechanistic, etiological mechanistic
Axis 5	constitutive explanations	
Axis 5a	content of components	see Tables 2.2 and 2.3
Axis 5b, FL	granularity of components	molar, molecular
Axis 5c, FL	number, identity, and necessity status of components	narrow, broad
Axis 6	causal-mechanistic explanations	
Axis 6a	order of components in the emotional episode	
Axis 6a1	components as primary explananda	bodily components, feeling component
Axis 6a2	components as explanantia	
Axis 6a2*, FL	components as explanantia of the feeling component	central, peripheral
Axis 6b	novel components as explanantia	
Axis 6b1	novel components involved in the cognitive component	
Axis 6b2, FL	novel components involved in the transition from the cognitive to the motivational component	stimulus-driven, goal-directed
Axis 6c	formats of representations	- perceptual, conceptual, associative, propositional - afferent, efferent

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

Table 2.4. (*cont.*)

axes, fault lines		values on axes
Axis 6d	operations	- associative, rule-based - simple, complex - primary, secondary, tertiary
Axis 6e	operating conditions	poor ~ automatic, ample ~ non-automatic
Axis 6f, FL	mechanisms dedicated to emotion or not	special-purpose, general-purpose
Axis 7	intensional scientific definition: criteria for demarcation	
Axis 8, FL	intensional scientific definition: trade-off between meta-criteria of similarity and fruitfulness	- similarity over fruitfulness: cluster definition - fruitfulness over similarity: regimenting, replace set by subset, split set into more subsets, eliminativist
Axis 9, FL	divisio scientific definition: criteria for partitioning	discrete, dimensional
Axis 10	divisio scientific definition: if Axis 9 = discrete: trade-off between meta-criteria of similarity and fruitfulness	similarity over fruitfulness, fruitfulness over similarity

BOX 2.2 Research Methods

Empirical research on emotions typically examines the influence of some components on other components (see validation sections in Chapters 4 to 8). This requires that some components are manipulated and others are measured. The present box lists methods for the manipulation and measurement of components and draws attention to the advantages and disadvantages of these methods (see also reviews in Mauss & Robinson, 2009; Moors & Scherer, 2013; Parkinson, 1997; Quigley et al., 2014; Roseman & Evdokas, 2004; Scherer, 1988; Scherer & Moors, 2019).

Manipulation

The first distinction of interest is that between overt (bodily) and covert (mental) components. Overt components are the somatic and motor components. Covert components include the cognitive, motivation, and feeling components. For the cognitive component, I consider the manipulation of (a) stimulus evaluation or appraisal (as per appraisal theories) and (b) the categorization of bodily feelings in terms of specific emotion categories (as per psychological constructionist theories) (see Table 2.5).

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

The manipulation of overt components can be done in a *direct* or *indirect* way. Peripheral physiological responses (somatic component) can be manipulated directly by administering chemical substances (e.g., epinephrine; Schachter & Singer, 1962). They can also be manipulated indirectly via the manipulation of behavior (e.g., relaxation and physical effort influence heart rate). Behavior (motor component), in turn, can be manipulated directly. An experimenter can raise the participant's arm or stimulate facial muscle contractions via electrical currents (e.g., Duchenne, 1862/1990). More often, however, behavior is manipulated indirectly via instructions. For instance, in the approach-avoidance task, participants have to pull or push a joystick (Cacioppo et al., 1993). In the amplification-suppression task, they are asked to amplify or suppress their spontaneous facial expressions (Lanzetta et al., 1976). In the muscle-to-muscle instruction task, they are asked to move different parts of their faces in specific ways (e.g., squeeze eyelids, raise nose, lift mouth corners; Duclos et al., 1989). Here, the instructions activate an intention or action tendency, which in turn causes the behavior. The less the instruction itself is interpreted in emotional terms, the more unobtrusive the manipulation is. For instance, the instruction to put on a happy face is more transparent than the instruction to hold a pen between the teeth (Strack et al., 1988).

The manipulation of covert components can only be done in an *indirect* way, but some methods are *remote* and others are *proximal*. The remote way to manipulate appraisals is via the manipulation of to-be-appraised stimuli (see Chapter 6). These can be either real or representational stimuli (Robinson & Clore, 2001). Examples of real stimuli are a real spider for spider phobics (Morina et al., 2015), social exclusion (K. D. Williams & Jarvis, 2006), other-induced point subtraction (Cherek, 1981; McCloskey et al., 2005), un/controllable electric shocks (Seligman, 1968; Wiech et al., 2006), and performance feedback (M. Lewis et al., 1990; Siemer et al., 2007; C. A. Smith & Kirby, 2009; see reviews by Nummenmaa & Niemi, 2004; Pedersen et al., 2002). Examples of representational stimuli are verbal scenarios describing real stimuli that people are asked to imagine or to recall (e.g., Kuppens et al., 2007; C. A. Smith & Ellsworth, 1985; C. A. Smith & Kirby, 2009; C. A. Smith & Lazarus, 1993), pictures (e.g., from the International Affective Picture Set; Lang et al., 2005), film clips (Kreibig, 2010), and a spider in virtual reality (Morina et al., 2015).

The proximal method manipulates appraisals via appraisal words, which are verbal representational stimuli referring to appraisals. These words can be embedded in an instruction, such as the instruction to recall or imagine a goal-incongruent event that was difficult to control (Fast & Chen, 2009; Galinsky et al., 2003; Kuppens et al., 2003; Lammers et al., 2008). The words can also be presented as primes in a priming procedure. For instance, P. K. Smith and Bargh (2008, Experiment 2) presented words referring to high vs. low power (and they measured the influence of these appraisals on the tendencies to approach vs. avoid). Neumann (2000) used a procedural priming procedure, in which participants were trained to make self- vs. other-attributions (and he measured the influence of these attributions on feelings of guilt vs. anger).

In choosing a method for the manipulation of appraisals, at least three issues are important. A first issue is that of control over confounding variables. The manipulation of appraisal via to-be-appraised stimuli (i.e., the remote method) ensures more control over the objective stimulus features whereas the manipulation of appraisals via appraisal words (i.e., the proximal method) grants more control over the appraisals that participants make. A second issue is the extent to which the processes induced by the manipulation resemble the processes induced by real emotion-eliciting stimuli outside the lab. It seems fair to assume that methods that manipulate appraisal via real stimuli approximate this process better than methods that make use of representational stimuli. This is not to say that representational stimuli are incapable of eliciting emotions, as is illustrated by emotions elicited by fiction (e.g., stories, films). As long as it is not clear whether representational stimuli elicit emotions via the same or a different process than real stimuli, however, the use of real stimuli is preferable. In case the representational stimuli are also appraisal words – and their relation with feelings is measured, for instance – there is an additional risk that semantic knowledge about relations between appraisals and feelings is activated in the participants' minds, and that this knowledge is responsible for the effects rather than any actual relations between appraisals and feelings in the world (Moors & Scherer, 2013; Parkinson, 1997). A third consideration is that methods that make use of appraisal words are more transparent, which makes them more vulnerable to self-presentation and demand effects (see Howle & Eklund, 2013).

Researchers studying the role of emotion categories have manipulated the accessibility of these categories in a remote or proximal way (see Chapter 9). The remote method involves the presentation of real or representational stimuli that are supposed to activate different emotion categories. Real stimuli were used in a study by Schachter and Singer (1962) in which confederates of the experimenter behaved in an angry or elated way. Representational stimuli were used in a study by Lindquist and Barrett (2008) in which they showed a picture of an interaction between an angry and a fearful man and in which they asked participants to tell a story about one of them. The proximal method involves the presentation of emotion words, which are verbal representational stimuli referring to emotion categories. By repeating these words either a few times or extensively, the accessibility of their corresponding emotion category is supposed to be increased (i.e., primed) or decreased (i.e., satiated), respectively (Lindquist et al., 2006).

Action tendencies are also mental entities and therefore require indirect methods for manipulation. This is typically done with a proximal method. This method consists of the manipulation of words that refer to actions or action tendencies. These words can be embedded in an instruction to imagine or recall an event that led to an action tendency (e.g., an urge to flee), but they can also be embedded in an instruction to execute the action, as is the case in stimulus-response compatibility tasks (e.g., approach/avoid: Kozlik et al., 2015; fight/flee: Fischer et al., 2022; see Chapter 6). It could in principle also be done with a remote method, involving instructions to engage in other actions than the ones under study. For instance, participants could be asked to take a slumped vs. upright position to induce a tendency to give up or be persistent (see Riskind, 1984).

Methods involving action tendency words (i.e., the proximal method) present similar advantages and disadvantages as methods involving appraisal words: There is more control about the action tendency that is activated, but also more risk for tapping into semantic knowledge, for self-presentation, and for demand effects.

The final mental component, the feeling component, must also be manipulated via indirect methods. Again, some methods are remote and others are proximal. Remote methods involve the presentation of real stimuli (e.g., announcing that participants will have to give a public speech) and representational stimuli (e.g., having participants watch film clips or listen to music). A proximal method is when words are presented that refer to specific feelings. These words can be embedded in an instruction to imagine or recall an instance of this feeling and to try to relive it. The words can also be presented without such instructions as when people are simply asked to read aloud sentences that evolve from neutral to positive (to induce positive feelings) or from neutral to negative (to induce negative feelings) (Velten, 1968). Both remote and proximal methods to induce feelings presumably rely on the mediation of other mental and/or bodily components.

Table 2.5. *Methods for the manipulation of emotional components*

		DIRECT	INDIRECT	
overt components	somatic	chemical substances	via instructions to behave	
	motor	direct manipulation or electric stimulation	via instructions to behave	
			REMOTE	PROXIMAL
covert components	cognitive: appraisal		to-be-appraised stimuli: 1. real stimuli 2. representational stimuli	representational stimuli (words) literally referring to appraisals 1. embedded in instructions for imagination or recall 2. presented as primes in priming procedure

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

Table 2.5. (*cont.*)

		DIRECT	INDIRECT	
			REMOTE	PROXIMAL
	cognitive: categorization		stimuli suggesting emotion categories: 1. real stimuli 2. representational stimuli	representational stimuli (words) literally referring to emotion categories presented a few/many times to increase/decrease accessibility
	motivational		instructions to execute actions other than the ones under study but that are supposed to cause the action tendencies under study	representational stimuli (words) literally referring to actions or action tendencies 1. embedded in instructions for imagination or recall 2. embedded in instructions to execute the actions
	feeling		stimuli that are supposed to cause feelings presumably via the mediation of other components	representational stimuli (words) literally referring to feelings 1. embedded in instructions for imagination or recall 2. embedded in instructions to read the words

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

Measurement

Methods for the measurement of components can be categorized as objective vs. subjective, and as direct vs. indirect (see Table 2.6). In *objective* methods, the measurement output is verifiable by others. This is the case for measures generating physiological and behavioral responses as their output. For instance, everybody can verify the numbers on a heart-rate monitor and everybody can register approach/avoidance responses. In *subjective* methods, by contrast, the measurement output is not verifiable by others. This is the case for verbal self-report. There is no way for others to verify whether a participant's rating on a scale is accurate.

In *direct* methods, the researcher uses the measurement output as a direct readout of the values of the to-be-measured variable/component. For instance, a heart-rate monitor directly delivers the values for the heart-rate variable. Self-reported action tendencies directly deliver the values for the action tendency component. In *indirect* methods, on the other hand, the researcher derives the values of the to-be-measured variable/component from another variable/component that is supposed to be influenced by the to-be-measured variable/component (De Houwer & Moors, 2010). For instance, behavior is a direct measure of the behavior component, but an indirect measure of the motivational component based on the assumption that action tendencies influence behavior. Note that the validity of an indirect measure hinges on the validity of the assumption that the to-be-measured component does indeed influence the measured component. It is worth emphasizing that "direct" in this context does not mean that these methods yield direct access to the to-be-measured variable/component. Saying that self-report of action tendencies is a direct measure of action tendencies does not mean that self-report yields direct access to this component. It is possible that the *participant* has to make several interpretive steps before producing a measurement output and this output can even be false. What it does mean is that the *researcher* does not make any interpretive steps going from the measurement output to the values of the variable/component. In the case of an indirect method, on the other hand, the researcher does have to take interpretive steps when inferring the values of the to-be-measured variable/component from the measurement output.

If both dichotomies are combined, four types of methods result: objective and subjective direct methods and objective and subjective indirect methods. Overt components can be measured with all four methods, but the preferred methods are *direct objective* ones. Examples are the measurement of peripheral somatic responses such as heart rate, blood pressure, and skin conductance response (SCR) with proper technical devices (see Kreibig, 2010); the measurement of central somatic responses with imaging techniques such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) (Lindquist et al., 2012; Sander et al., 2018); the measurement of coarse behavior in choice tasks or by registering spontaneous behavior (Fischer, Kuppens, & Moors, 2020); the measurement of facial expressions with the judgment method (e.g., Ekman, 1973), the facial action coding system (FACS; Ekman & Friesen, 1978), and electromyography (EMG) (e.g., Tassinary & Cacioppo, 1992; see Barrett et al., 2019); the measurement of vocal expressions via acoustic analysis (Juslin & Scherer, 2005); and the measurement of gestural expressions with the body action and posture coding system (BAP; Dael et al., 2012a, 2012b).

Covert components can in principle be measured with all but direct objective methods. Appraisals, categorizations, and feelings are most often measured via *direct subjective* methods, that is, self-report. Self-reports can be collected long after the fact on a single occasion (i.e., retrospective questionnaire) or they can be collected online on multiple occasions (i.e., experience sampling method, ESM; Csikszentmihalyi & Larson, 2014). Appraisals and feelings have also been measured via *indirect objective* methods. Common indirect objective methods to measure the positive or negative evaluation of stimuli, for instance, are compatibility tasks such as the implicit association task (IAT; Greenwald et al., 1998) and the evaluative priming task (EPT; Fazio & Olson, 2003; Spruyt et al., 2011). In a typical EPT, participants are asked to classify positive and negative target stimuli that were briefly preceded by prime stimuli. If a given prime stimulus leads to faster reaction times (RTs) on subsequent positive/negative targets than negative/positive targets, researchers infer that the participant evaluated the prime stimulus as positive/negative. It may be noted that although the IAT and EPT are supposed to measure the appraisal of stimuli, they could also be used to measure the feelings that the stimuli evoke.

A well-known disadvantage of self-report is that it opens the door for self-presentation strategies and demand effects (Schwarz & Strack, 1991). This problem should be reduced (but not completely solved) by using compatibility tasks (e.g., IAT and EPT; see Schnabel et al., 2008). In both self-report and compatibility tasks, however, there is a risk of tapping into semantic knowledge about conceptual relations between whatever is manipulated and the measured appraisals or feelings. This problem might be somewhat reduced for the online self-report method (ESM). Because in this method, less time has elapsed between the emotional episode and the measurement, the hope is that it will tap into episodic rather than semantic knowledge (Robinson & Clore, 2002). It should be kept in mind, however, that even online measures are necessarily retrospective and that people's episodic knowledge is influenced by their semantic knowledge (Lambie & Marcel, 2002).

Action tendencies can likewise be measured with direct and indirect subjective methods. A *direct subjective* method asks participants to self-report on their action tendencies (Frijda et al., 1989). An *indirect subjective* method asks participants to self-report on their behavior, which is then used to infer the action tendencies causing the behavior. Action tendencies can also be measured with *indirect objective* methods. Indeed, they can be inferred from overt behavior. A first indirect objective method is to simply observe the participant's behavior (e.g., a choice in a single-trial study or choice frequencies in a multiple-trial study; e.g., Bossuyt et al., 2014a, 2014b). A second indirect objective method is to infer action tendencies (i.e., the tendencies to engage

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

in coarse behavior) from facial expressions (i.e., subtle behavior) (Frijda & Tcherkassof, 1997). For instance, lip biting may reveal the presence of a tendency to attack. A third indirect objective method is to assess action tendencies via compatibility tasks (e.g., Bossuyt et al., 2014a; Krieglmeyer et al., 2013). In approach-avoidance tasks, for instance, participants are instructed to approach positive stimuli and avoid negative stimuli in a "compatible" block and approach negative stimuli and avoid positive stimuli in an "incompatible" block. Shorter RTs in the compatible than the incompatible block suggest that the positive/negative stimuli elicited a tendency to approach/avoid, so that when participants were instructed to approach/avoid (in the compatible block), execution was facilitated compared to when they were instructed to avoid/approach (in the incompatible block). Here the presence of a particular action tendency is inferred from the relative speed with which it can be executed when instructed (see review by Kozlik et al., 2015). A fourth indirect objective method, recently developed at our lab, makes use of motor-evoked potentials (MEPs) enhanced by transcranial magnetic stimulation (TMS) (Fini et al., 2020; Fischer, Fini, et al., 2020; Moors et al., 2019). In one study (Fischer, Fini, et al., 2020), we first trained participants to use their right index finger to approach and their thumb to avoid. After that, participants were presented with positive and negative pictures that they had to observe. On each trial, a single TMS pulse was administered to the left primary motor cortex (M1) at 300 ms post-stimulus onset in order to boost any action tendency that was already elicited by the picture so that it would become measurable by MEPs on the right (i.e., contralateral) hand. The comparison of MEP amplitudes on both fingers allowed us to infer which action tendency was elicited by certain stimuli: Higher MEP amplitudes on the approach/avoid finger than on the avoid/approach finger suggested that participants had a tendency to approach/avoid. A fifth, and final indirect objective method makes use of EEG, more in particular lateralized readiness potentials (LRPs). LRPs index the motor preparation for a left- or right-hand response via the neural activity in the contralateral motor cortex. Participants could be trained to use their left hand for one response (e.g., approach) and their right hand for another response (e.g., avoid). Comparison of the amplitudes of the LRPs for both responses would allow to infer which of both responses is prepared for more strongly in the presence of certain stimuli (see e.g., Eder et al., 2012).

The latter four indirect methods provide a window into action tendencies that are not (yet) translated into coarse overt behavior. The neuroscientific methods, moreover, present a clear temporal advantage and are therefore less likely to be "contaminated" by participants' attempts to regulate action tendencies in order to improve self-presentation or to meet the expectations of the experimenter (i.e., demand effect). They are therefore said to be the most "implicit" methods.

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

Table 2.6. *Methods for the measurement of emotional components*

			DIRECT	INDIRECT
OBJECTIVE	overt compo-nents	somatic	1. peripheral: e.g., heart rate, SCR, blood pressure 2. central: e.g., EEG, fMRi	
		motor	1. aspects of coarse behavior: e.g., single choice, choice frequency 2. facial behavior: e.g., judgment method, FACs, EMG 3. vocal behavior: e.g., acoustic analysis 4. gestural behavior: e.g., BAP	
	covert compo-nents	cognitive: appraisal		inferred from RTs in compatibility tasks
		cognitive: categorization		could in principle be inferred from RTs in compatibility tasks
		feeling		inferred from RTs in compatibility tasks
		motivational		inferred from 1. coarse behavior 2. subtle behavior: e.g., facial, vocal, gestural behavior 3. RTs of instructed behavior (in compatibility tasks) 4. TMS enhanced MEPs 5. LRP

Table 2.6. (*cont.*)

			DIRECT	INDIRECT
SUBJECTIVE	overt components	somatic	self-reported somatic responses (only peripheral)	
		motor	self-reported behavior	
	covert components	cognitive: appraisal	self-reported appraisal	inferred from self-reported feelings
		cognitive: categorization	self-reported emotion categories	inferred from self-reported feelings
		feeling	self-reported feelings	
		motivational	self-reported action tendencies	inferred from self-reported behavior

https://doi.org/10.1017/9781107588882.005 Published online by Cambridge University Press

PART II

Emotion Theories One by One

In Chapter 2, I listed a number of axes on which emotion theories can differ. These differences could arise across the four stages of the demarcation-explanation cycle: working definitions in Stage 1, explanations in Stage 2, validation in Stage 3, and scientific definitions in Stage 4. Chapter 2 already worked out the details for the working definitions in Stage 1. I listed the desiderata, that is, the typical and apparent properties for which there is fair consensus that emotion theories need to account for them. In the coming chapters, I discuss the next three stages for a range of emotion theories. To structure the discussion, I will group emotion theories into families based on the causal-mechanistic stories that they peddle (with a clear emphasis on mental rather than neural mechanisms; i.e., **Axis 6**). My choice for this axis is prompted by a radically mechanistic approach. I apply this approach not only to psychological emotion theories, which are already mechanistically oriented at heart, but I will also try to "squeeze" philosophical emotion theories into the mechanistic mold. This will inevitably come at the cost of not doing full justice to all the questions that are traditionally posed within the philosophy of emotion (e.g., epistemology, phenomenology, rationality, ethics). Yet it may also bring a hitherto underexplored benefit, namely that answers to mechanistic questions could eventually also throw some light on these non-mechanistic questions. The analysis presented here is first and foremost an exercise. It is an honest attempt to make – sometimes hidden – assumptions about mechanisms more explicit. As such, it is an invitation for the reader to participate in what will hopefully become a fruitful discussion in years to come.

The order of appearance of emotion theories is as follows. I start with general precursors in Chapter 3 (Darwin, 1872; James, 1890b), followed by evolutionary theories in Chapter 4 (Ekman, 1992a, 1999a; Izard, 2011; Keltner et al., 2016; Panksepp, 1998; Scarantino, 2014; Tomkins, 1962) and network theories in Chapter 5 (Bower, 1981; Lang, 1994; Leventhal, 1984; M. D. Lewis, 2005; Teasdale, 1999). After that, I discuss stimulus evaluation theories in Chapter 6. These are split into two brands. The first brand, called evaluation-first theories, includes appraisal theories (Ellsworth,

https://doi.org/10.1017/9781107588882.006 Published online by Cambridge University Press

2013; Frijda, 1986; Lazarus, 1991; Roseman, 2013; Scherer, 2009b), judgmental theories (Gordon, 1987; Green, 1992; W. Lyons, 1980; Marks, 1982; Neu, 2000; Nussbaum, 2001; Reisenzein, 2012; Solomon, 1993), quasi-judgmental theories (Armon-Jones, 1991; Greenspan, 1988), and perceptual theories (de Sousa, 1987; Döring, 2007; Roberts, 2013; Tappolet, 2016). The second brand, called embodied theories, includes J. J. Prinz's (2004a) embodied appraisal theory, Deonna and Teroni's (2012) attitudinal theory, Griffiths's (2004b) affordance theory, and several enactivist theories (Colombetti, 2014; Hutto, 2012; Shargel & Prinz, 2018). Chapter 7 discusses response evaluation theories, in particular my own goal-directed theory (Moors, 2017a; Moors et al., 2017). Chapter 8 zooms in on psychological constructionist theories (Barrett, 2006b, 2012, 2017b; Russell, 2003, 2012; Schachter, 1964). Chapter 9 closes with social theories (Averill, 1980; Fridlund, 1994; Mesquita & Boiger, 2014; Parkinson, 1995). The latter family is the odd one out, as its members have mechanistic affinity with one or more of the personal emotion theories discussed in the other chapters, while opening their scope to the social dimension of emotions.

The discussion of each theoretical family (except the social theories in Chapter 9) is organized into four sections. The first section discusses the precursors specific to the theory at hand. The second section presents the constitutive and causal-mechanistic explanations of the theory (i.e., Stage 2 in the demarcation-explanation cycle; **Axes 5** and **6**). I start by sketching the rudimentary order in which the traditional components listed in Table 2.2 are placed within the emotional episode (**Axis 6a**). I also indicate the component(s) that is (are) singled out as the emotion proper and that form the constitutive explanation of emotion in the narrow sense (**Axis 5c**). After that, I break down the sequence into smaller steps and I analyze each step in terms of the novel components listed in Table 2.3 (**Axis 6b**). Information about the content of these novel components will be complemented with information about representational format (**Axis 6c**), operations (**Axis 6d**), and operating conditions (**Axis 6e**) insofar as this is available. In this second section, I also identify a number of axes that help organize the within-family variety. I will label these axes with capital letters instead of numbers. Some of these axes correspond to those that were already discussed (see list in Table 2.4) and this will be indicated.

The third section presents the intensional and divisio scientific definitions that flow from the theory's explanations (i.e., Stage 4 in the demarcation-explanation cycle; **Axes 7** and **9**). In the subsection on the intensional definition, I start by listing the theory's criteria for demarcation of the set of emotion, after which I evaluate the adequacy of this definition in terms of the meta-criteria of apparent-similarity and fruitfulness. In the subsection on the divisio definition, I first list the theory's criteria for partitioning of the set of emotion, after which I also evaluate this divisio definition in terms of the meta-criteria of apparent-similarity and fruitfulness.

The fourth section focuses on validation (i.e., Stage 3 in the demarcation-explanation cycle). This section provides an overview of research lines designed to evaluate the theory's causal-mechanistic explanations on empirical grounds. For some theories, the section also touches upon lines of research that serve theory development rather than validation. Finally, for some theories, the section also handles issues that have come up with regard to the internal/external consistency of the explanations.

As the above indicates, I switched the order of Stages 3 (i.e., validation) and 4 (i.e., scientific definitions). One reason for this is that I follow the order in which theorists have de facto proceeded. As mentioned in Chapter 2, theorists have put up their scientific definitions before completing the validation stage. Another reason is for communicative purposes. By laying all the cards of a theory on the table, we are better equipped to evaluate the type of empirical evidence that is required for validating it.

CHAPTER 3

General Precursors

I start with the theories of Darwin (1872) and James (1890b) because they can be considered as general precursors for many of the theories discussed later. Darwin (1872) preceded James (1890b) and inspired evolutionary theories, but it is James (1890b) in particular, with his minimalist and counterintuitive theory, who kick-started the development of several contemporary theories. Many other precursors deserve a place in the guard of honor. As some of these precursors can more easily be attached to a particular theoretical family, they will be discussed in the chapters dedicated to the respective family (for historical reviews, see Arnold, 1960; Gendron & Barrett, 2009; Reisenzein, 2012; Scarantino, 2016).

3.1 Darwin

3.1.1 Explanations

Darwin's (1872) sequence of components in an emotional episode echoes the sequence endorsed in folk psychology. A stimulus gives rise in some unspecified way to an emotion or feeling. As Darwin (1872) used the terms emotion and feelings interchangeably, we may infer that he equated emotion with the feeling component (i.e., constitutive explanation). The feelings subsequently produce bodily reactions (somatic and motor components). For example, the perception of a crouching tiger produces fear, which in turn produces palpitations and trembling, a rise in heart beat and sweating palms, and fleeing behavior. In this sequence of events, the first phase is the transition from the stimulus to the feeling component. The second phase is the transition from the feeling component to the bodily components.

As Darwin (1872) does not provide a causal-mechanistic account of the first phase, and hence does not explain the emotion proper, he cannot be credited with a theory of emotion per se. Rather than taking emotion as the explanandum, Darwin (1872) took emotional behavior as the explanandum once the emotion was already in place. The explanations that he provided for this second phase were not merely mechanistic but also

https://doi.org/10.1017/9781107588882.007 Published online by Cambridge University Press

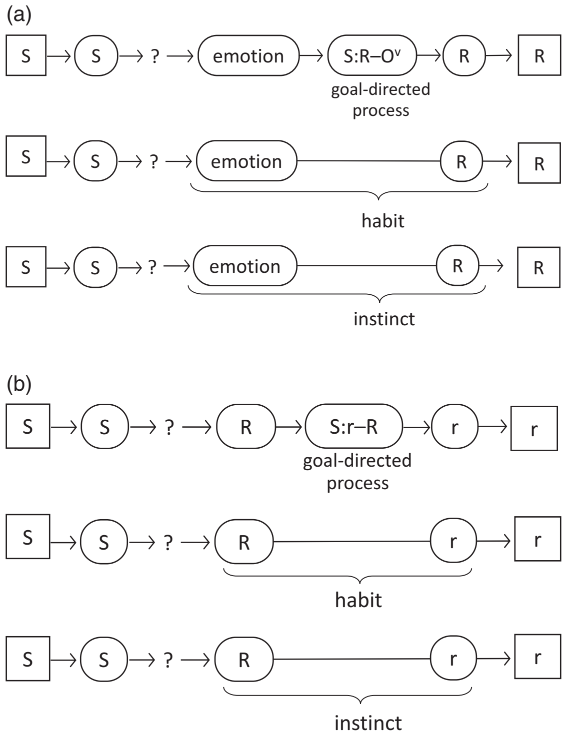


Figure 3.1 Darwin's theory: (a) Darwin's transition from goal-directed process to habit to instinct: first interpretation; (b) Darwin's transition from goal-directed process to habit to instinct: second interpretation
Note: Observable entities are in squared boxes, mental entities in rounded boxes.

evolutionary. He provided an evolutionary story for the mechanisms that elicit emotional behavior today.

Darwin (1872) started from the observation that much emotional behavior, and especially facial expressions, have a compelling and universal character. It is this observation for which Darwin (1872) proposed an evolutionary, and eventually also a mechanistic explanation. The evolutionary story (similar to that proposed by Spencer, 1855) goes somewhat like this (see Figure 3.1(a)). In our evolutionary past, certain mental states including emotions or feelings such as anger or dislike, but also motivations such as hunger, led to voluntary, so-called "serviceable" behaviors, which I take to be instrumental behaviors caused by goal-directed processes. For instance, if a dog is angry and seeks to attack an opponent, he selects the action tendency to bite the opponent but at the same time to put his ears flat in his neck to protect them from being bitten off. If this goal-directed process is frequently repeated in the individual's life, it transforms into a so-called "habit." This is a process in which the emotion directly activates the action tendency even if the action is no longer instrumental in the current context. Note that Darwin's (1872) usage of the term habit as learned [emotion–R] links (or more generally, [mental-state–R] links) deviates from the contemporary usage, which reserves the term habit for learned [S–R] links (see Box 2.1). Across several generations, the [emotion–R] links become hereditary at which point they are called "instincts."[36] Instincts are innate and hardwired in the brain. Here as well, the behavior may no longer be of use, which is why it has grown into a weaker version, that is, a remnant or vestige of the once useful behavior. The mechanistic explanation of current weak or subtle emotional behavior then, not just of dogs but also of other species such as humans, is explained by Darwin (1872) as the automatic activation of an innate [emotion–R] link. For instance, when we are angry at an opponent, we clench our teeth as a remnant of goal-directed biting behavior, and when we are afraid of a threat, we widen our eyes as a remnant of the goal-directed behavior to better examine the source of the threat. Darwin (1872) called this subtle emotional behavior "expressive" behavior.

In addition to this evolutionary mechanism, known as the principle of serviceable associated habits – short, the teleological principle – Darwin (1872) also proposed two other mechanisms for the evolution of instincts: the principle of antithesis and the principle of direct activation of the nervous system – short, the idiopathic principle. In brief, the principle of antithesis explains the genesis of behavior that was never serviceable itself but that is opposite in form to behavior that was. Examples are opening the hand palms and raising the eyebrows in impotence as the opposites of clenching the fists and lowering the eyebrows in anger. The idiopathic principle explains responses that are not serviceable themselves but that are a byproduct of surplus energy aroused by certain serviceable behaviors that seeks its way out via other bodily channels, both muscular and visceral ones. Examples are trembling of the muscles, sweating of the skin, blushing, dilation of the nostrils, and movements in joy.[37]

[36] Darwin (1872; see also James, 1890b) sometimes used the term instincts and habits to refer to the *responses* caused by innate and habitual [emotion–R] links. For ease of communication, however, I will use these terms to refer to the [emotion–R] *links* themselves, and I will call the responses caused by these links instinctual and habitual responses.

[37] Darwin (1872) considered these responses to be useless. Today, however, several of them are considered useful. For instance, trembling and sweating are seen as useful for controlling body temperature.

In sum, instincts grow out of the habits of our ancestors (teleological principle), the opposites of these habits (antithesis principle), and from an overflow of nervous discharge (idiopathic principle). But for Darwin (1872), habits or learning also remained important in present-day life: Some instincts require current habits (i.e., learning) in order to develop in the individual's life and many of our instincts are quickly modified by current habits. At the same time, however, Darwin (1872) insisted that only expressions that stem from instincts "deserve to rank as true expressions" (p. 50). Although it is hard to tell in what sense he used the word "true" here, it is possible that he meant "truly emotional." If so, it would be a small step to credit him with the view that it is not just the [emotion–R] links that are hardwired but also the emotions themselves. To discover these true expressions, Darwin (1872) recommended turning to the observation of the expressions of animals, infants, the insane, and people that are cut off from Western society. He did not trust the expressions of adults as they are often too subtle and there is a high risk that researchers are biased in their interpretation of them. A final method he recommended was to artificially produce facial expressions in adults via electrical currents (Duchenne, 1862/1990) and to let observers classify these.

A special type of learning occurs when remnant instinctive behaviors, which are no longer useful in the sense that they no longer impact on the physical world, are co-opted in a later stage of evolution in goal-directed processes that are at the service of the goal to communicate. In this way, the instinctive behaviors become again useful in the sense that they can now impact on the social world. Although the baring of the teeth can no longer impact on the physical world, it now serves to signal that the agent is angry. Such goal-directed processes may again grow into habits, and (across many generations) into instincts. As proof of the existence of such secondary instincts, Darwin (1872, p. 46) described the stereotypic postures that cats and dogs adopt to communicate acceptance and hostility. But even if an agent lacks the goal to communicate her mental state or if she has an active goal to suppress her behavior, some leakage will be inevitable due to the automatic nature of instincts. As a consequence, observers may still use this behavior as a cue to infer the agent's mental state. In a further stage in evolution, repeated goal-directed attempts of agents to keep their expressions in check may again grow into habits first and instincts later. As an example, Darwin (1872, p. 354) suggested that the oblique position of the eyebrows and the drawing down of the corners of the mouth in sadness originated from the goal to prevent screaming.

3.1.2 Scientific Definitions

As Darwin's (1872) theory does not present constitutive and causal-mechanistic explanations for emotions, and therefore cannot technically be considered an emotion theory, there are also no scientific definitions to be distilled. Nevertheless, the theory can still be evaluated in terms of its success in explaining some of the desiderata outlined in the working definition for emotions in Chapter 2. Darwin's (1872) habits and instincts can account for the automatic and irrational character that much of our emotional responses have. Their automatic nature is tied to the fact that they rely on associative operations. In addition, once initially goal-directed processes have passed the stage of being goal-directed and grown into habits and later instincts, they have lost their sensitivity to the current outcomes of behavior. Thus, habits and instincts qualify as practically irrational in the process-sense. They could, however, still yield adaptive behavior and hence qualify as practically rational in the output-sense. This aligns with Darwin's (1872) portrayal of habits and instincts as generating behavior that may not be "of the least use" (p. 48) but that may still be useful on occasion (p. 42, p. 50). If goal-directed processes tailored to the reward structure of one environment, later grow into habits and instincts, activation of these habits and instincts may continue to be adaptive in environments with the same reward structure but cease to be so in environments with a different reward structure. Suppose an animal developed the habit to bite a rival because that behavior repeatedly led to access to the best food. The habit may continue to be useful in the presence of a submissive rival, but cease to be so in the presence of a dominant rival, as the animal gets bitten back and robbed of the food. Likewise, instincts that grew out of the goal to communicate a mental state may be adaptive or maladaptive depending on the circumstances. For instance, baring one's teeth when angry may discourage some opponents from engaging in a physical fight while encouraging others.

3.1.3 Validation

I discuss three objections that have been raised about the internal consistency of Darwin's (1872) theory. Arnold (1960) called the teleological principle into question, asking how our animal ancestors, who are supposed to be less mentally developed would be capable of goal-directed behavior selection in the first place. Contra Arnold (1960), however, the capacity of animals to engage in goal-directed behavior is no longer called into question (e.g., Dickinson & Balleine, 1994). Still, it is legitimate to ask why our animal ancestors would have more access to goal-directed processes in the selection of emotional behavior than we would today.

Arnold (1960) also noted that the transition of habits learned in an individual's life into instincts (and their weakening through the loss of usefulness) fits more with Lamarckian than with Darwinian evolution theory. Lamarck's (1809) theory was based on the principle that

https://doi.org/10.1017/9781107588882.007 Published online by Cambridge University Press

characteristics acquired during the lifetime of an individual can be transmitted to the next generations. In Darwin's (1872, p. 41) defense, however, it may be noted that he did mention at some point that instincts not only grow out of habits that get inherited but that they can also develop from other mechanisms, such as prior, more primitive instincts. The variation that occurs in these prior instincts creates a basis for natural selection.

Dewey (1894) raised the objection that if the mental state at the start of Darwin's (1872) evolutionary story is a feeling, the behavior that results from it cannot be called "serviceable," strictly speaking. There is simply no behavior that is at the service of feelings. Only if the mental state is a goal, does it make sense to call a response serviceable (in reaching that goal), an idea that Darwin (1872, pp. 353–354) explicitly acknowledged when he wrote that "all those [movements] included under our first [teleological] principle were at first voluntarily performed for a definite object – namely, to escape some danger, to relieve some distress, or to gratify some desire." If Darwin's (1872) mental states were indeed restricted to goals, there would still be two different ways in which his instincts could be understood. Some instincts may be composed of a non-behavioral goal (e.g., goal to have food) connected to an abstract action tendency (e.g., goal to fight): [O^v–R]. Other instincts may be composed of an abstract action tendency (e.g., goal to fight) connected to a concrete action tendency, that is, a motor program or pattern of to-be-executed responses (e.g., goal to bite): [R–r]. In many of Darwin's (1872, p. 354) examples, the second option seems to be in operation. For instance, in the examples cited earlier, the dog was angry and *wanted to attack* ([R]), or the animal was in fear and *wanted to escape from danger* ([R]), and these abstract action tendencies were connected to more concrete action tendencies such as the tendency to bite ([r]) (later reduced to the tendency to bare the teeth) and the tendency to scan the source of threat ([r]) (later reduced to the tendency to widen the eyes). Although feelings are indeed not enough, Darwin (1872) did not carefully distinguish between various types of mental states (emotions, feelings, goals). As a result, several interpretations of Darwin's (1872) instincts have lived on, the two most important ones being: [emotion–R] and [R–r] (see Figure 3.1).

A further objection, also raised by Dewey (1894) is that Darwin (1872) did not explain the origin of the idiopathic principle, that is, where the excess neural discharge comes from. Dewey (1894) proposed that this occurs when an organism overestimates the feasibility of the selected action option. If an animal estimates that fleeing is possible, but next discovers that it is trapped, the energy that was mobilized to flee has to seek its way out via other channels. Thus, he concluded, the idiopathic principle marks the breakdown of the teleological principle. It could be

https://doi.org/10.1017/9781107588882.007 Published online by Cambridge University Press

argued, however, that excess neural discharge may also be produced when an organism overestimates the amount of effort needed to execute the selected action, resulting in overshooting. If an animal overestimates the effort required to flee, not all energy may be used up once fleeing is successful. Research shows that overshooting is more likely when the goal at stake is important, as when a football player kicks too hard for the decisive penalty (see Oudiette et al., 2019; Pessiglione et al., 2007). The latter type of error does not mark a complete breakdown but a miscalculation.

3.2 James

Unlike Darwin (1872), who took the existence of emotions for granted and tried to explain emotional behavior, James's (1884, 1890b, 1894) ambition was to explain emotions themselves.

3.2.1 Constitutive and Mechanistic Explanations

James (1884, 1890b, 1894, similar to C. G. Lange, 1885/1922) famously argued that perception of an exciting fact by the sensory cortex directly activates the motor cortex, which produces bodily responses, such as peripheral physiological responses and behavior. Feedback of these bodily responses is again perceived by the sensory cortex, and the resulting feeling or conscious experience *is* the emotion proper (i.e., constitutive explanation). For example, the perception of a crouching tiger produces a rise in heart beat and sweating palms, a facial expression, and fleeing behavior. These bodily changes are fed back to the sensory cortex, where they are registered and felt, and this feeling is the emotion we typically call fear. In James's (1884, 1890b) theory, bodily changes precede the emotion, as expressed in the aphorism "we feel afraid because we tremble and run." This view, foreshadowed by Descartes (1649/1989) and Spinoza (1677/1982) deviated from the view endorsed by Darwin (1872) and laypeople that emotions precede bodily changes, as expressed in the opposite aphorism "we tremble and run because we feel afraid."

James's (1890b) sequence can be split into four steps: a first step in which a raw external stimulus (eS) is linked to an afferent representation ([eS]) (i.e., cognitive component), a second step in which the afferent representation ([eS]) is linked to an efferent representation ([R]) (i.e., transition from cognitive to motivational component), a third step in which the efferent representation is manifested in overt responses (R) (i.e., transition from motivational to somatic and motor components), and a fourth step in which a raw internal stimulus (iS) is linked to an

https://doi.org/10.1017/9781107588882.007 Published online by Cambridge University Press

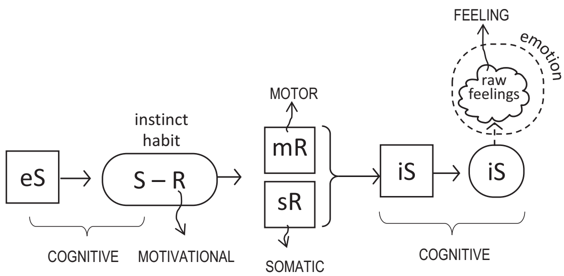


Figure 3.2 James's theory
Note: Feelings are in clouds; traditional components (motor, cognitive, motor, somatic, feeling) are outside the boxes; novel components are in squared (observable entities) or rounded boxes (mental entities); the constitutive explanation of emotion is surrounded by a dotted bubble.

afferent representation ([iS]) (i.e., cognitive component). Once the latter representation becomes conscious it is also felt (i.e., feeling component) (see Figure 3.2). I provide a detailed discussion of each of these four steps below.

Step 1

In the first step, an exciting stimulus is perceived (i.e., cognitive component). The standard case is a raw external stimulus (eS) that is linked to an afferent representation of this stimulus with concrete features ([ecS]). Most of the examples that James (1890b) cited fit this bill. But James (1890b) also included cases in which a stimulus is imagined or remembered. These are cases in which an afferent representation of a stimulus with concrete features ([ecS]) is also activated, but in which the source of the activation is unspecified. Some of James's (1890a) examples of emotions suggest that he was open to the possibility that more complex processes could figure in the first step (Ellsworth, 1994; Ratcliffe, 2005). A case in point is his analysis of self-related emotions like self-complacency (pride, vanity, arrogance) and self-dissatisfaction (modesty, humility, shame). As causes of these emotions, James (1890a, p. 306) mentioned success and failure. The process at work in these cases can thus be characterized as an evaluation process that takes external stimuli (eS) or representations thereof ([ecS]) as its input and produces representations of the abstract features success and failure ([eaS]) as its output.[38]

[38] James (1890a, p. 310) further specified the formula "self-feelings = success divided by pretensions." Self-feelings can be increased either by increasing the numerator or decreasing the denominator.

https://doi.org/10.1017/9781107588882.007 Published online by Cambridge University Press

Steps 2 and 3

In the second step, the afferent representation ([ecS] or [eaS]) from the first step is directly linked to an efferent representation ([R]) (i.e., transition from cognitive to motivational component). This matches the format of a stimulus-driven or [S–R] process. James (1890b) put forward innate [S–R] links, called instincts (i.e., primary processes), as well as learned [S–R] links, called habits (i.e., secondary processes).[39] Note that these are different from Darwin's (1872) instincts and habits, which were innate and learned [emotion–R] links (or broader: [mental-state–R] links). In fact, James's (1890b) sequence "S→[S–R]→ R" can be considered as a further reduction of Darwin's (1872) sequence "S→[S]→[emotion–R]→R" by simply erasing the [emotion] term. The third step, the translation from efferent representation ([R]) to overt responses (sR, mR) is straightforward. It occurs unless it is hindered by a competing efferent representation or a physical obstacle (in line with the ideomotor hypothesis, see Chapter 1 and Box 2.1).

Each innate [S–R] link is activated by a specific US (or a restricted class of USs) and gives rise to a specific UR. James (1890b) listed a number of instincts ranging from simple ones that produce behaviors like sucking, spitting out, clasping objects, holding the head erect, sitting up, and standing, to more complex ones such as the curiosity instinct (producing approach in response to novel stimuli), the fear instinct (producing flight in response to loud noises, strangers, solitude, and high places), and instincts of rivalry, pugnacity, and/or hunting (producing fighting in response to an even wider range of stimuli). James (1890b) stated that all instincts can produce emotional responses. However, this is easier to see for the complex instincts (e.g., fear instinct) than for the simple ones (e.g., sucking instinct). One way out for James (1890b) could be to argue that all instincts can give rise to emotions provided that the bodily responses they evoke are intense enough to be consciously felt, and this may be more likely for complex than simple instincts. McDougall (1908) avoided possible confusion in this respect by creating bins of instincts: complex ones that lead to emotions and simple ones that do not (see Chapter 4).

James (1890b) accepted Darwin's (1872) idea that instincts are installed by evolution and that they are vestiges of once instrumental actions that have now become innate. He stated that instincts "*produce certain ends, without foresight of the ends, and without previous education in the performance*" (James, 1890b, p. 383; italics in original). The ends or outcomes that

[39] James (1890b, p. 554) used the term habit to refer to all kinds of associations, not only [S–R] associations (as in contemporary usage) but also [S–S] associations or [R–R] associations, and even complex chains of [S–R–S–R] associations.

https://doi.org/10.1017/9781107588882.007 Published online by Cambridge University Press

are produced in this way comprise physical and social well-being that ultimately serve survival.

Instincts can produce "instinctual behavior" – outward deeds that take effect on the external world (e.g., fleeing from a wild animal) and that can serve physical well-being. But instincts can also produce "emotional responses" – physiological responses and facial expressions that do not go beyond the body. Facial expressions can still take effect on the social world, and in this way serve social well-being (James, 1890b, p. 422). James's (1890b) "emotional responses" piggyback on his "instinctual behavior": Every instinctual behavior is accompanied by emotional responses. But emotional responses can also occur in the absence of instinctual behavior, for instance, when activation of the instinct was too weak to produce the instinctual behavior or when the instinctual behavior was inhibited.

To explain the particular form of emotional responses confined to the body, James (1890b, pp. 481–482) adopted the teleological and idiopathic principles put forward by Darwin (1872) and added the principle of analogy. According to the latter principle, an innate [S–R] link can become activated by a stimulus that is similar to the representation of the stimulus that is part of this [S–R] link. For instance, the sucking movements of the lips in response to a sweet stimulus can generalize to any stimulus that is "sweet" if only in symbolic sense. Similarly, turning up one's nose at a bad odor can generalize to turning up one's nose in case of moral or social disdain. This takes us seamlessly to the next point: the role of learning.

Like Darwin (1872), James (1890b) pointed out that instincts are rapidly modified by learning. A first type of learning is when the set of stimuli that can trigger an instinct is extended via (a) generalization based on similarity (and that also forms the basis of the principle of analogy above) or (b) contiguity, that is, the co-occurrence in time and space of the initial stimulus with a novel stimulus (i.e., classical conditioning). The first route explains why fish bite at fake worms, the second route explains why Pavlov's dog salivated at the bell after it repeatedly co-occurred with the delivery of food. But James (1890b, p. 390) focused most on operant conditioning principles, which install goal-directed processes on top of the stimulus-driven process involved in instincts: "It is obvious *that every instinctive act, in an animal with memory, must cease to be 'blind' after being once repeated*, and must be accompanied with foresight of its 'end' just so far as that end may have fallen under the animal's cognizance." Once an instinct has led to overt behavior and an actual outcome, the outcome is stored so that from then on, the organism will be guided by an expectation of this outcome. A positive outcome reinforces the behavior and a negative outcome inhibits it.

Outcomes are positive or negative insofar as they appeal to other instincts. If the outcome of one instinct is the stimulus of another opposite instinct, the stimulus that initiated the first instinct may create a conflict between the two instincts. Thus, "a rat's memory [. . .] of a former escape from a trap would neutralize his impulse to take bait from anything that reminded him of that trap" (James, 1890b, p. 390). While "nature" has endowed lower animals with only a few instincts that make "them act *always* in a manner which would be *oftenest* right" (James, 1890b, p. 392), she reduces the risk of fatal errors in higher animals by endowing them with many conflicting instincts, so that hesitation occurs and slight alterations in the stimuli end up determining the behavior that carries the day. James (1890b, p. 293) thereby resisted the classic opposition between instinct and reason, arguing that reason in itself is powerless and can only defeat one instinct by creating an image of a stimulus that excites an opposing instinct.

Like many of his contemporaries (e.g., Darwin, 1872), James (1890b) sided with the view that extensive repetition of a goal-directed mechanism ([S:R–O^v]) transforms it into a habit (learned [S–R] link) (but see Box 2.1). This is the reason why James (1890b) used the terms learning and habit interchangeably.

James (1890b, p. 454) held that the shaping of instincts by learning processes creates an infinite variety in the stimuli that trigger them and in the emotional responses that they give rise to, as well as large individual differences in both stimuli and responses. "Jokes at which one explodes with laughter nauseate another, and seem blasphemous to a third" (James, 1890b, p. 448). Further, in contrast with Darwin (1872), who insisted that only innate processes can generate "true" – or should we say "truly emotional" – expressions, James (1890b) accepted that both innate and learned processes can produce emotional responses. Thus, whereas Darwin (1872) took the variety resulting from learning as a source of error, James (1890b) took this variety to be part of the very fabric of emotions. James's (1890b) idea that the total number of innate and learned [S–R] links is infinite, combined with his idea that [S–R] links not only lead to emotional responses (somatic and facial responses) but also to instinctual behavior (outward deeds), contributed to his view that emotional responses are produced by general-purpose mechanisms (i.e., not dedicated to emotional responses). Add to this his view that these [S–R] links are not localized in dedicated brain centers (James, 1884, p. 188, 1890b, p. 453), and it becomes clear why he resisted reifying emotions as "eternal and sacred psychic entities" (James, 1890b, p. 449) as so many of his contemporaries did.

Step 4

In the fourth step, the bodily responses produced in Step 3 return to the sensory cortex where they are perceived via interoception and

https://doi.org/10.1017/9781107588882.007 Published online by Cambridge University Press

proprioception (i.e., cognitive component). The process at work here is one that takes a raw internal stimulus (iS; the immediate outcomes of somatic and motor responses) as its input and produces an afferent representation of concrete stimulus features as its output ([icS]). To the extent that the content of this representation is conscious, it is also felt, and it is this feeling that James (1890b) called the emotion. This feeling qualifies as raw feeling and not as labeled feeling.

James (1890b) was convinced that the feelings he called emotions are exhausted by (changes in) bodily feelings: The emotion begins and ends with them. To make this insight persuasive, he presented his "subtraction" argument, which reads: "If we fancy some strong emotion, and then try to abstract from our consciousness of it all the feelings of its bodily symptoms, [. . .] a cold and neutral state of intellectual perception is all that remains" (p. 451). Two nuances are worth noting, however. First, in addition to the standard emotions discussed so far, James (1884, 1890b) also left room for non-standard emotions such as moral, intellectual, and aesthetic emotions. He split these emotions in a non-emotional stage involving purely moral or intellectual cognitions or purely aesthetic sensations, and an emotional stage in which these cognitions and sensations excite the body and without which they would not count as emotions. Second, James (1890b, footnote p. 459) relaxed his demand for the presence of bodily responses by also accepting "hallucinated" bodily responses as a sufficient basis for emotions, a possibility that was later relabeled as the "mental simulation" of bodily responses or the "as-if body loop" in the neo-Jamesian theories of Damasio (1994) and J. J. Prinz (2004a).

An implication of the claim that emotions are exhausted by bodily feelings is that the variety (intensity and quality) in feelings stems from the variety (intensity and nature) in the bodily responses. James's (1890b) hypothesis that each token emotion has its own specific pattern of bodily changes has been misinterpreted by his critics as the hypothesis that each emotion type *that has a label in natural language* has its own specific pattern of bodily changes (Gendron & Barrett, 2009). Yet James (1890b, p. 448) did not seem convinced that the variety within emotion types was smaller than that between emotion types. In addition, James (1890b, p. 450, pp. 472–474) argued that there are no special brain centers for Step 4 – the perception of the feedback that produces the feelings. All that is needed is a cortex that can perform an ordinary process of perception by operating as a sounding board or projection surface for every change in the body. Just like in Steps 2 and 3 then, the only mechanism at play here is general-purpose rather than emotion-specific.

3.2.2 Scientific Definitions

In this section, I examine what sort of scientific definitions would be suggested by the constitutive and mechanistic explanations proposed in James's (1890b) theory. To this end, I consider the criteria that these explanations suggest for (a) the demarcation of the set of emotions (scientific intensional definition) and (b) the partitioning of this set (scientific divisio definition). I also evaluate the adequacy of the resulting definitions in terms of the meta-criteria of apparent-similarity and fruitfulness.

3.2.2.1 Intensional Definition

CRITERIA FOR DEMARCATION

James's (1890b) constitutive explanation tells us that emotions are exhausted by bodily feelings. This is the first necessary criterion, which I call the "bodily feeling" criterion. It specifies the presence of the feeling component with the content of the feelings being bodily responses. James's (1890b) mechanistic explanation tells us that these bodily feelings must be caused in a specific way: by stimuli that, after being perceived, activate instincts or habits, which in turn lead to bodily responses that are again perceived. This is the second necessary criterion, which I call the "Jamesian mechanism" criterion. To determine whether these two criteria buy the set of emotions a scientific status, they must be evaluated in the light of the meta-criteria of apparent-similarity and fruitfulness.

ADEQUACY

Apparent-Similarity To evaluate apparent-similarity, we must ask whether the criteria of bodily feeling and Jamesian mechanism can account for the properties of emotions specified in the working definition of Chapter 2. Even if James (1890b) does not accept that all of these properties are real, he should ideally still be able to explain why they are apparent.

The bodily feeling criterion, which stipulates that emotions equal bodily feelings, accounts for the phenomenal aspect of emotions. The content of these bodily feelings – bodily responses – also accounts for the bodily aspect of emotions. James (1890b) further linked this bodily aspect to the heat of emotions, which allows demarcating them from cold, purely intellectual phenomena. Apart from these selling points, critics have argued that James's (1890b) bodily feeling criterion may not be necessary for emotion by pointing at counterexamples that lack a bodily glow such as hope and regret (Deonna & Teroni, 2012, p. 65). An easy way out for James (1890b), however, would be to reply that if hope and regret lack a bodily glow, they are cognitions rather than emotions. This is the line of thought that he followed for non-standard emotions (like moral, intellectual, and aesthetic emotions), for which he argued that if they do not excite bodily feelings, they remain in the cognitive stage and do not count as emotions at all.

Apart from necessity, critics have also argued that bodily feelings are not sufficient for emotions. If bodily feelings were understood as purely phenomenal sensations, they would lack Intentionality altogether (Deonna & Scherer, 2010). As James (1890b) argued that bodily feelings are the result of perception (in Step 4), however, bodily responses figure in the content of representations, and this ensures Intentionality. However, such a solution would not provide us with the right kind of Intentionality as emotions are typically not about bodily responses but about an object in the external world. Anger is not about an adrenaline rush, but about an offense, and disgust is not about a squeamish feeling but about a disgusting object (Deonna & Teroni, 2012; Salmela, 2011, p. 2). As Dewey (1895, p. 17) put it, "the child who ceases to be angry *at* something – were it only the floor at last – but who keeps up his kicking and screaming, has passed over into sheer spasm." Given that bodily feelings lack the typical world-directed Intentionality of emotions, they do not permit demarcation from non-emotional bodily feelings such as feeling tired or cold. As Worcester (1893, as cited in James, 1894, p. 521) put it, "shivering from cold . . . is the same sort of a movement as may occur in violent fright, but it does not make us feel frightened."

The bodily feeling criterion indeed does not seem sufficient to demarcate emotions from non-emotional bodily feelings. This is exactly why James (1890b) equips us with an additional necessary criterion: the Jamesian mechanism criterion. This criterion, which stipulates that bodily feelings must be caused by instincts and habits, does allow demarcating emotions from non-emotional bodily feelings such as feeling tired or cold because the latter are caused by physical exercise or a temperature drop respectively.[40] James's (1890b) advocates might even argue that this criterion accounts for the world-directed Intentionality of emotions. After all, instincts and habits are triggered by perceived stimuli. These stimuli can therefore act as the objects of emotions. If bodily feelings are produced by the encounter with a bear, chances are high that the bear is also on one's mind and constitutes the object of one's fear. As it turns out, James (1890b, 1894) gave more attention to the world-directedness of emotions than is often thought (Ellsworth, 1994; Ratcliffe, 2005).

[40] The further narrowing of the set of emotions (which James, 1890b, takes to be constituted by bodily feelings) by invoking its mechanistic origin (Jamesian mechanism) is reminiscent of the way in which the set of hangovers (constituted by a headache, nausea, and a dry mouth) was distinguished from viral infections by invoking the cause of drinking too much and the ensuing physical processes.

According to him, the stimulus and the bodily feelings are both perceived and combine in consciousness so that "an object-simply-apprehended" transforms into an "object-emotionally-felt" (James, 1890b, p. 474).[41]

The fact that James (1890b, pp. 458–461) allowed for the existence of objectless emotions such as morbid terror, groundless anger, melancholy, and apathy might suggest that he allowed for the existence of emotions that are not caused by stimuli; and this would suggest that he considered Intentionality to be optional or typical at best. Yet, a closer look at his writings shows that he framed these emotions as only "outwardly causeless" (James, 1890b, p. 461) because they result from instincts or habits for which the thresholds were set so low that any stimulus, however weak, could trigger them. So rather than being "objectless" emotions, these are better called emotions with "overgeneral objects."

Critics may object, however, that the concrete stimulus representation that serves as the basis for instincts and habits still does not deliver the right kind of Intentional object for emotions. Emotions are not about the concrete features of stimuli, but rather about their significance for us. My fear of the anaconda is not about its coiling form or its slippery surface, but about the possibility that it will eat me alive. Ward (cited in McDougall, 1908, p. 54) sketched the problem as follows: "Let Professor James be confronted first by a chained bear and next by a bear at large: to the one object he presents a bun, and to the other a clean pair of heels." This suggests the need for a process that does more than just register the concrete features of stimuli, such as an evaluation process that produces representations with abstract content like valence, danger, loss, and so on. This is precisely the path taken by stimulus evaluation theories (see Chapter 6).

Similar problems arise for the heat of emotions. James (1890b) stipulated that a state does not become emotional unless and until bodily responses become intense enough to break through into consciousness, but he did not explain in a satisfactory manner where this intensity comes from. James (1890b) might argue that a more intense stimulus input leads to a stronger activation of the instincts and habits producing the bodily responses, but this cannot be the whole story. If you whisper in my ear that you hate me, I will be as shocked as (or even more so than) when you shout it to me during an escalating fight. In addition, we must observe that James (1890b) does not provide a satisfactory account of valence either. The concrete features of stimuli are entirely value free (e.g., the

[41] Note, however, that a similar line of defense is adopted by theorists who equate emotions with Intention-less feelings, such as Whiting (2011), I. Goldstein (2002), and Reisenzein (2012). They would say, for instance, that fear of the ice is the sum of the objectless emotion of fear and a representation of the ice.

coiling form of a snake). This has made the call for a stimulus evaluation process even louder (see Chapter 6).

Some of James's (1890b, 1894) advocates were confident that he did have an evaluation process in mind, but that he found "[t]he role of interpretation (or appraisal) [...] so obvious that it need[ed] no special emphasis" (Ellsworth, 1994, p. 223). James (1894, p. 518) did indeed seem to assume a role for representations with abstract features when he wrote "The same bear may truly enough excite us to either fight or flight; according as he suggests an overpowering 'idea' of his killing us, or one of our killing him." It may further be recalled that James (1890b) discussed self-complacency and self-dissatisfaction as two emotions that are caused by success and failure. These few examples notwithstanding, the bottom line is that James (1890b) did not give special emphasis to an evaluation process and that many of his other examples seemed to do just fine with concrete representations requiring no further evaluation.

In conclusion, by adding the Jamesian mechanism criterion, the extension of James's (1890b) scientific intensional definition of emotion can be somewhat narrowed down but remains still broad. There is another sense, however, in which adding the mechanism criterion keeps James's (1890b) extension broad: His list of instincts is very wide-ranging. It not only includes instincts that resemble emotions such as the instincts of fear, pugnacity, rivalry, shame, and curiosity, but also instincts that seem neutral, such as the instincts of cleanliness, secretiveness, imitation, appropriation, and construction. And then we have not even counted the many modifications that these instincts can undergo through learning.

James (1890b) has an easier time accounting for the self-directed Intentionality of emotions, due to the presence of the action tendencies that are part of instincts and habits. Instincts can also explain the continuity in emotions between human adults, infants, and animals, whereas habits (i.e., the modification of instincts by learning) can explain why emotions of adults are more complex or rich than those of infants and animals. Learning can also account for the wide range of stimuli that can elicit emotions, including actual stimuli, representations of stimuli (i.e., remembered, imagined, or anticipated stimuli), and aesthetic stimuli, as well as for the variety across individuals within and across cultures.

The automatic nature of instincts and habits should account for the automatic nature of many emotions and perhaps also their control precedence. One dissenting voice, however, came from Cannon (1927), who objected that bodily responses are generated too slowly to determine emotions (see below).

The stimulus-driven nature of instincts and habits explains why emotions are practically irrational in the process-sense. Instincts and habits do not take into account the current outcomes of behavior, even if they evolved from goal-directed processes that did. The behavior generated by instincts and habits can still be adaptive if the reward structure of the current environment matches that of the reward structure of the environment in which the original goal-directed process was installed. Thus, emotions can still be practically rational in the output-sense. This is entirely analogous to my discussion of Darwin's (1872) instincts and habits.

Given that in James's (1890b) theory, emotional responses cause emotions and not the other way around, critics accused him of neglecting the influence of emotions on behavior (Damasio, 1994, p. 130) or to consider emotions as "a kind of froth on top of the real business of behavior" (Oatley, 1992, p. 133). In James's (1890b) defense, Barbalet (1999) replied that the fact that bodily responses precede emotions does not preclude them from having an influence on behavior afterwards. He illustrated this by citing James's (1897, pp. 96–97) discussion of an alpine-climber who is better able to execute a dangerous leap when feeling confident than when feeling fearful.

Fruitfulness To evaluate fruitfulness, we must ask whether James's (1890b) criteria of bodily feeling and Jamesian mechanism provide homogeneity among the members that he included in the set of emotions and whether they allow separating these members from those of other sets. This should ensure that meaningful generalizations can be made among these members. In James's (1890b) view, instincts, and especially habits, are ubiquitous causes of behavior. The only thing that separates emotional from non-emotional habits is that the bodily responses produced by the former are felt whereas those produced by the latter are not. Whether this will turn out to be a fruitful way of carving up the realm of psychological entities depends on whether it makes any difference for the organism whether bodily responses are felt or not. But perhaps we should not try to look for something that is not there. There are many indications that James (1890b) was a skeptic, who was not in the business of vindicating the set of emotions, but rather tried to give an account of the things that people call emotions. James (1890b) repeatedly argued that the mechanisms involved in emotions are general-purpose.

3.2.2.2 *Divisio Definition*

Criteria for Partitioning

The next question to ask is whether James's (1890b) constitutive and mechanistic explanations provide us with a meaningful way to carve up the set of emotions internally. His constitutive explanation that emotions

https://doi.org/10.1017/9781107588882.007 Published online by Cambridge University Press

are feelings of bodily responses, together with the assumption that bodily responses can vary in infinite ways, suggests that emotions can also vary in infinite ways. His mechanistic explanation that the bodily responses are caused by instincts may reduce the variety to some extent, as the bodily responses produced by the fear instinct probably differ from those produced by the pugnacity instinct. To illustrate, James (1884, pp. 193–194) described fear as a feeling of quickened heart beats, shallow breathing, trembling lips, weakened limbs, goose-flesh, and visceral stirrings. And he described rage as a feeling of ebullition in the chest, flushing of the face, dilatation of the nostrils, and clenching of the teeth. Yet James's (1890b) idea that habits can modify these instincts in infinitely arbitrary ways increases this variety again without offering any useful principles for partitioning in return. Thus, according to James (1890b), the infinite variety in bodily responses did not follow the fault lines of vernacular emotion types, which is why there can be no doubt that he was skeptical about the existence of discrete emotion subsets. This fits with his aversion to existing lists of descriptions of bodily responses for different emotions. He said such descriptions were tedious and that he "should as lief read verbal descriptions of the shapes of the rocks on a New Hampshire farm as toil through them again" (James, 1890b, p. 448). Like most other emotion theorists, however, James (1890b) kept using discrete emotion labels for the purpose of communication and he did occasionally provide descriptions of bodily responses for different emotions (like those mentioned above) for the purpose of illustration. This is probably why he has been misunderstood by his critics as making the claim that every vernacular emotion has a fixed bodily response pattern (e.g., Cannon, 1927; Hufendiek, 2016). James's (1890b, p. 454) rejection of the discrete view, however, should not be confused with a full-blooded commitment to the dimensional view in that he did not identify any meaningful dimensions to organize the set of emotions. He argued instead that "*any classification of the emotions is [. . .] as true and as 'natural' as any other*, if it only serves some purpose."

ADEQUACY

Apparent-Similarity and Fruitfulness Even if a theory denies that there is sufficient ground for partitioning the set of emotions into vernacular emotion subsets, it would be an asset if the theory can still make sense of them, and in this way pass the apparent-similarity test. James's (1890b) theory does not seem to possess this asset, however. As critics have argued, James's (1890b) bodily feelings are not sufficient to account for the variety of emotions in terms of vernacular emotion types. Cannon (1927) pointed at a failure of empirical research to identify specific visceral response patterns for specific emotions, which led to the conviction that bodily feelings are too undifferentiated to account for the variety in emotions (see Reisenzein, 1983). Although James (1890b) did not think bodily feelings are undifferentiated, he did agree that their variety could not be neatly organized in vernacular emotion types. In this sense, Cannon's (1927) data did not falsify James's (1890b) theory. It remains true, however, that James (1890b) could not make sense of the discrete variety that the critics (trying to vindicate these emotion types) were after. In other words, James (1890b) could not deliver apparent similarity.

It seems that adding James's (1890b) mechanistic explanation does not bring us any closer in this respect. Even if the bodily responses produced by different instincts (e.g., fear vs. pugnacity) are likely to be different, James's (1890b) idea that learning can modify both the stimulus and response representations of instincts in infinitely arbitrary ways entails that every ground for grouping emotions is again lost soon after birth. As long as the stimuli are only processed in terms of their concrete features, the variety remains innumerable. This is another reason why many critics did not settle for James's (1890b) perceptual process but expressed the need for a cognitive process that goes beyond mere perception.

What about fruitfulness? James (1890b) did not identify any concrete categories or dimensions for organizing the set of emotions. Hence, there is nothing that can be evaluated in terms of fruitfulness here.

3.2.3 Validation

Few emotion theories have been turned inside out in the way that James's (1890b) theory has. The current section concentrates on the empirical status of James's (1890b) theory. But first a note on internal consistency. The most often heard objection against James's (1890b) theory is that by placing responses before emotions, he placed the cart before the horse (Pineda, 2015a). According to common sense, it is simply unthinkable that emotions come after their outward expressions. Nevertheless, several researchers took up the glove and examined the empirical implications of James's (1890b) theory.

Cannon (1927) argued that it is implausible that bodily responses – which he restricted to visceral responses – would come before feelings because these responses and their feedback are too slow to determine feelings (see above). The solution offered by Cannon (1927) and McDougall (1908) was to argue that feelings already arise at the stage of the activation of the [S–R] links (i.e., centrally) and do not have to be postponed until the feedback (from the periphery) reaches the cortex again. Others have argued, on the other hand, that bodily responses are not restricted to visceral responses but also include musculo-skeletal responses, which arise and are fed back much quicker (e.g., Damasio, 1999).

Rather than focusing on *order*, one line of research investigated James's (1890b) assumption that feedback from the peripheral effects of bodily responses to the sensory cortex is *necessary* for emotions. If this is indeed the case, the separation between the peripheral organs and the central nervous system should hinder the occurrence of emotions. Research conducted with paralyzed patients and animals shows mixed results. Patient studies show diminished emotions. Animal studies show intact emotions. In a study by Hohmann (1966), paralyzed patients (due to spinal cord injuries) reported reduced intensity of fear and anger (but increased sentimentality after loss) compared to the time before they got paralyzed, and the degree of paralysis correlated with the intensity of these emotions. Chwalisz et al. (1988), however, found support for the alternative interpretation that changes in the pattern of these patients' emotions stemmed from changes in their daily lives. They found that people in a wheelchair reported a stronger intensity for some emotions but not others regardless of whether they were paralyzed or whether there was another reason why they were in a wheelchair. The patient studies need to be interpreted with caution, however. To conclude that peripheral feedback is necessary for emotions, it must be demonstrated that when peripheral feedback is absent, no emotions are experienced. All patients still reported emotions. On the other hand, none of the patients were completely paralyzed and feedback from their facial responses was still possible (Damasio, 1999). In addition, they may have merely "hallucinated" (James, 1890b, footnote, p. 459) or "mentally simulated" bodily responses (Damasio, 1994; J. J. Prinz, 2004a, pp. 58–59). A similar concern can be raised against the finding that people whose facial muscles were temporally paralyzed due to Botox injections did not show the expected reduction in feelings while watching positive or negative videoclips (J. I. Davis et al., 2010). These people could still have received somatosensory feedback from other body parts or they could have mentally simulated facial responses (Hufendiek, 2016; Reisenzein & Stephan, 2014).

Animal studies showed that animals that no longer received peripheral feedback due to surgical intervention (in which the spinal cord and vagus nerve were cut) still showed intact emotional behavior (dogs: Sherrington, 1900; cats: Cannon et al., 1927). This evidence was dismissed as irrelevant, however, because James (1890b) had claimed that *feelings*, and not emotional *behavior*, were dependent on feedback from the periphery (Barbalet, 1999; Ellsworth, 1994; but note that Cannon, 1927, did in fact admit that these data speak to behavior rather than to feelings).

Another piece of evidence that has been put forward as critical for validating James's (1890b) theory is evidence for the direct physical induction of emotions, either by taking drugs or adopting artificial facial expressions or coarse behavior (see also Chapter 5). Based on the

https://doi.org/10.1017/9781107588882.007 Published online by Cambridge University Press

Jamesian mechanism criterion, however, the mere activation of the bodily components independent of their instinctual/habitual path would not make the resulting bodily feelings genuinely emotional. Thus, it can be argued that such evidence is not a test of the theory and falls outside of its scope. Validating the theory would require demonstrating that direct activation of the instinct/habit would produce emotions (instead of emotions being produced indirectly via stimuli). But as James (1890b) did not think there are dedicated brain areas for instincts/habits, the electrical stimulation of certain brain sites does not seem to be an option here.

A final line of research, already hinted at, examined the claim falsely attributed to James (1890b) that each prototypical emotion is characterized by a specific pattern of bodily responses (e.g., Cannon, 1927; Hufendiek, 2016). This claim could not be confirmed by empirical evidence in James's (1890b) days, but the debate was reopened later under the impulse of evolutionary theories (see Chapter 4).

To wrap up, the constitutive and causal-mechanistic explanations that James (1890b) proposed for emotions allowed us to distill two criteria for a scientific intensional definition of emotion: Emotions are bodily feelings (i.e., bodily feeling criterion) and they are caused by instincts/habits that produce bodily responses that are fed back to the brain (i.e., Jamesian mechanism criterion). Although these criteria go a long way to fulfill the desiderata of embodiment, phenomenality, continuity, automaticity, control precedence, practical irrationality, and self-directed Intentionality, they are less satisfactory when it comes to the desiderata of world-directed Intentionality and even heat. In addition, James (1890b) bumped into the differentiation problem. He was unable to provide a clear ground for individuating the discrete emotion types from natural language in a scientific divisio definition. As James (1890b) was a skeptic, who explicitly resisted seeing emotions and emotion types as special kinds of entities, the issues raised did not pose a problem for him. Nevertheless, even skeptics should ideally be able to account for the descriptive properties of emotions, that is, the apparent properties that are typically ascribed to them. James (1890b) was only partially successful in this respect.

The "gaps" in James's (1890b) theory inspired the development of many later theories. One solution to the differentiation problem was offered by McDougall (1908). He proposed modifications to the [S–R] link in Step 2. In particular, he put forward the existence of specific instincts that are dedicated to specific vernacular emotion types. To make sure the specificity would not get lost after habit formation set in, he also postulated that the [S–R] links that grounded the vernacular emotions remained unchanged after learning. In his theory, habits can only extend the range of concrete stimuli that activate the [S–R] link as well as the range of concrete responses resulting from this activation, but they cannot touch the original [S–R] core. This [S–R] core is also supposed to generate feelings centrally (see above). As I will explain shortly, McDougall (1908) can be considered as the bridge between James (1890b) and evolutionary theories (see Chapter 4). The evolutionary theories gave further substance to McDougall's (1908) [S–R] links by anchoring them in the brain. Thus, they postulated dedicated brain mechanisms for a limited set of vernacular emotions, called basic emotions.

Another solution to the differentiation problem that James (1890b) faced is provided by a family of theories known as network theories (see Chapter 5). These theories embraced James's (1890b) proposal that learning takes on a major role in shaping emotions, but they did not – or at least not all of them – adopt his conclusion that learning erases all ground for having discrete emotion types.

A further set of solutions to solve the problems of differentiation as well as world-directed Intentionality is provided by theories that add an abstract cognitive process to James's (1890b) sequence. This process takes on different shapes in different theories. Stimulus evaluation theories (see Chapter 6) insert a stimulus evaluation process in Step 1, prior to the [S–R] connection in Step 2. Stimuli must be evaluated in order to give rise to action tendencies, behavior, and feelings (Arnold, 1960). These theories propose that the features of the stimulus representation that is part of the [S–R] connection are not concrete (e.g., coiling form of a snake) but abstract (e.g., valence, danger, loss). Psychological constructionist theories (see Chapter 8), on the other hand, insert a construction process (e.g., categorization) between the bodily responses in Step 3 and the feelings in Step 4 (Barrett, 2006b; Russell, 2003; Schachter, 1964). These theories propose that representations of diffuse bodily responses and representations of abstract stimulus features combine to yield labeled feelings.

A final adjustment to James' (1890b) theory is provided by response evaluation theories, in particular the goal-directed theory (Moors, 2017a; see Chapter 7). This theory replaces the [S–R] connection in Step 2 with a response evaluation process, more specifically, a goal-directed process, which involves [S:R–O^v] connections. This is another general-purpose mechanism grounding another skeptical approach. With this overview of adjustments in hand, I will move on to a discussion of the evolutionary theories, preceded by a discussion of the preparatory work of McDougall (1908).

https://doi.org/10.1017/9781107588882.007 Published online by Cambridge University Press

CHAPTER 4

Evolutionary Theories

4.1 Precursors

Darwin (1872) is usually cited as the founding father of evolutionary theories of emotion. Some of Darwin's (1872) ideas, however, only exerted their influence on evolutionary theories through the filters of the theories of James (1890b) and McDougall (1908). James (1890b) turned Darwin's (1872) instincts, understood as [mental-state–R] links into [S–R] links. Building further on these [S–R] links, McDougall (1908) in turn, made several fundamental changes to James's (1890b) theory in an attempt to fix what was lacking in that theory from a vindicator's point of view. These changes formed the perfect stepping stone to evolutionary theories.

James (1890b) argued that all instincts can elicit emotions, including those for which the connection with emotion did not seem intuitive (e.g., instincts to walk and construct). McDougall (1908) solved this problem by explicitly distinguishing between instincts that do (e.g., flight, repulsion, curiosity, pugnacity, self-abasement/assertion, parental) and do not elicit emotions (e.g., reproduction, food seeking, gregariousness, acquisition, crawling, and walking). Moreover, he postulated that each of the emotional instincts, alone or combined, accounted for the emotion types we know from natural language. For instance, the flight instinct is for fear, the pugnacity instinct for anger, and the repulsion instinct for disgust. Pugnacity combines with repulsion to create scorn and with flight to create loathing. This simple intervention allowed McDougall (1908) to demarcate the set of emotions from other sets and to account for the variety within this set. The fact that instincts could be combined in numerous ways still allowed McDougall (1908) to account for the infinite shadings that James (1890b) had emphasized, in the same way as mixing a few basic colors can produce an infinite palette of color shadings.

Another intervention that McDougall (1908) made to keep the separation between emotions and other sets tidy was to hold the influence of learning on emotional instincts at bay. James (1890b) argued that habits could completely overtake instincts, and he refused to give special status to instincts in generating emotions as compared to habits. McDougall

https://doi.org/10.1017/9781107588882.008 Published online by Cambridge University Press

(1908), by contrast, argued that learning could not modify the emotional instincts themselves and that these constituted the pure emotion (returning in this sense to Darwin's, 1872, idea that expressions stemming from instincts were more "true"). Learning increases the variety in the inputs and outputs of emotions without touching their innate core – the "hard liquor" so to speak (J. J. Prinz, 2004a).

On the input side, learning leads to an extension and refinement of the range of stimuli that can trigger the innate [S–R] links. Extension happens via associative learning based on similarity (i.e., generalization) and co-occurrence (i.e., classical conditioning). The co-occurrence of the initial stimulus and a neutral stimulus forges an association between the stimuli, after which the mere presentation of the neutral stimulus can now activate the [S–R] connection and elicit an emotion. This is how animals, for instance, learn to fear hunters: As hunters get paired with the loud noises from their guns and the chasing of their dogs, they come to activate the innate connection between loud noises and the flight tendency, felt as fear. Refinement happens via discrimination learning (i.e., a form of operant conditioning). This is illustrated by the undisturbed grazing of cattle on pastures next to railways. The fact that the loud noises from trains are never followed by negative outcomes leads to a refinement of the inlet of the flight instinct, resulting in selective responding to different types of loud noises. On the output side, variety is created by goal-directed attempts to control or regulate the manifestation of action tendencies in overt responses. Such attempts would be more successful for large movements of trunk and limbs than for facial expressions and futile for physiological responses. McDougall's (1908) view can be characterized as a dual-system view with a default-interventionist architecture in which the innate stimulus-driven process (instinct) is the default and the goal-directed process is the intervenor (see Box 2.1).

A final deviation that McDougall (1908) proposed from James's (1890b) theory has to do with the time point at which feelings are generated (**Axis 6a2***). James's (1890b) theory counts as a peripheralist theory in that feelings are postponed until feedback of the bodily responses is perceived. McDougall's (1908) theory, by contrast, is a centralist theory in that feelings are already produced by the instinct itself (see also Cannon, 1927). McDougall (1908) parsed instincts into three parts. The first part is the afferent part ([S]) of the [S–R] link, during which a perception of the object takes place (i.e., cognitive aspect). The second part is the central part or the bridge between the afferent ([S]) and the efferent part ([R]), during which a feeling (however faint) is already produced in regard of the object (i.e., affective aspect). This feeling constitutes the emotion and corresponds with the vernacular emotion types, such as fear and anger. The third part is the efferent part ([R]) or action tendency of the [S–R] link,

https://doi.org/10.1017/9781107588882.008 Published online by Cambridge University Press

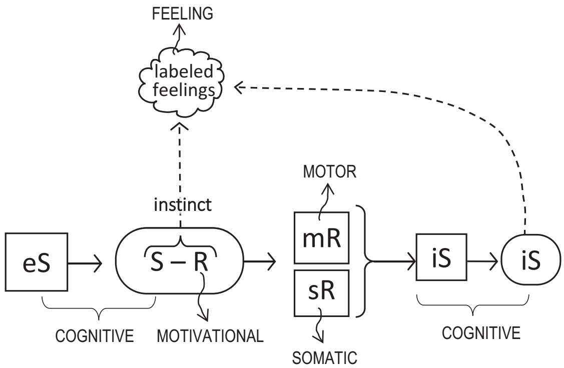


Figure 4.1 McDougall's theory

during which a striving toward/away from the object occurs. Action tendencies have their own central correlated feelings, such as the felt urges to flee and fight. Like in James's (1890b) theory, instincts get manifested in overt bodily responses, which can be fed back via perception and produce peripheral feelings. But while James (1890b) took these peripheral feelings to constitute the emotion, McDougall (1908) considered these only as accessory to the central feelings already produced earlier. McDougall's (1908) sequence is presented in Figure 4.1.

The changes that McDougall (1908) introduced to James's (1890b) theory paved the way for evolutionary theories (Buck, 1999; Damasio, 2004; Ekman, 1984, 1992a, 1999a; Ekman & Cordaro, 2011; Griffiths & Scarantino, 2005; Izard, 2011; Keltner, Tracy, et al., 2019; LeDoux, 1996; Levenson, 1994, 2011; Panksepp, 1982, 1998, 2005, 2012; Panksepp & Watt, 2011; Plutchik, 1980, 1984; Shariff & Tracy, 2011b; Tomkins, 1962, 1963, 1970, 1995; Tooby & Cosmides, 1990; Tracy, 2014). Members of the evolutionary family share the following mechanistic backbone: Emotions are (partly) constituted or caused by inherited innate [S–R] links in which the stimuli are recurrent challenges and the responses are the solutions to these challenges that were selected by evolution. Individual members of this family show differences in a number of corollary theoretical assumptions, which can be captured in the following axes (and values): (a) the width of the constitutive explanation (**Axis A**: from narrow to broad; **Axis 5c**), (b) the role of stimulus evaluation (**Axis B:** absent vs. present), (c) the time at which feelings are generated (**Axis C**: central vs. peripheral; **Axis 6a2***), and (d) details about the number and nature of the emotions that are included (**Axis D**). I start with the commonalities and indicate deviations from the consensus along the way.

https://doi.org/10.1017/9781107588882.008 Published online by Cambridge University Press

4.2 Constitutive and Mechanistic Explanations

The sequence of components proposed by evolutionary theories is similar to that in McDougall's (1908) theory. A stimulus, after being processed, activates the innate [S–R] link for a specific emotion type, which produces physiological responses (somatic component), motor responses (motor component), and feelings (feeling component) specific to that emotion type. The [S–R] links are no longer called instincts, however, but "affect programs," and they are implemented in brain mechanisms (Ekman & Cordaro, 2011; Tomkins, 1962; Levenson, 2011, called them "core systems"; Panksepp & Watt, 2011, called them "command networks"[42]). To illustrate this sequence, the processing of a crouching tiger activates the affect program for fear, which consists of a link between the representation of the tiger and the tendency to flee. This, in turn, generates an adrenaline rush and blood flow to the legs (mobilizing the body in general and preparing for flight), a fear expression (in face, voice, and/or gestures), actual flight behavior, and the feeling we call fear (see Figure 4.2).

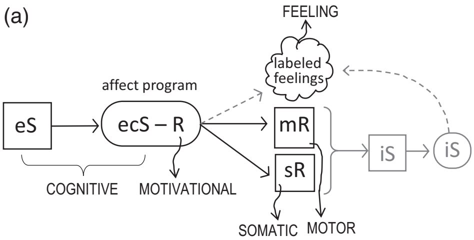
The constitutive explanations of evolutionary theories vary somewhat. Some evolutionary theories identify emotion with the affect program itself (e.g., Ekman & Cordaro, 2011;[43] Izard & Malatesta, 1987, p. 496; Tomkins, 1970). Still others lean more towards treating the affect program as the cause of emotion while equating the emotion itself with the sum of the resulting somatic, motor, and feeling components. A final group takes the entire emotional episode to be the emotion (Levenson, 2011; Scarantino, 2014; Tracy, 2014).

In many evolutionary theories, the emotional episode has the following sequence of steps: a first step in which a raw external stimulus (eS) is linked to an afferent representation ([eS]) (i.e., cognitive component), a second step in which the afferent representation ([eS]) is linked to an efferent representation ([R]) (i.e., transition from cognitive to motivational component), and a third step in which the efferent representation ([R]) is manifested in overt bodily responses (sR, mR) and feelings (i.e., transition

[42] Note that Tomkins (1970) used the term "affect" to indicate emotions, which he understood as responses and not as feelings. Panksepp and Watt (2011) rejected the term "affect program" because to them "affect" means "feelings" and they wanted to focus – just like Tomkins (1970) – on responses.

[43] Ekman and Cordaro (2011) considered the affect program as the sine qua non of emotion. When Ekman (1992a, 1992b, 1999a) wrote that nobody identifies emotions with a single component, I interpret this as meaning that nobody uses a single component to empirically diagnose the activation of an affect program and that we need to infer this from the total set of components.

https://doi.org/10.1017/9781107588882.008 Published online by Cambridge University Press



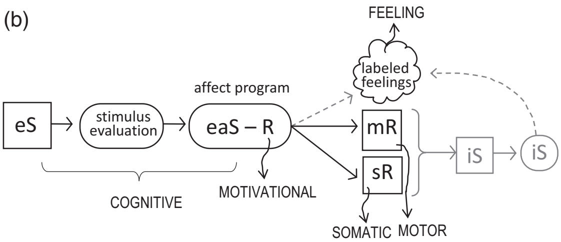


Figure 4.2 Evolutionary theories: (a) non-hybrid version; (b) evolution-evaluation hybrid version
Note: The gray arrow from the affect program to the feeling component is the route to feelings for the centralist version; the gray path from the motor and somatic components to the feeling component is the route to feelings for the peripheralist version.

from motivational component to somatic, motor, and feeling components). Let us consider these three steps in more detail.

Step 1

In some evolutionary theories, the first step (i.e., from stimulus to afferent representation) is nothing more than a simple perception process, which links a raw stimulus (eS) to a representation with concrete stimulus features ([ecS]) (see Figure 4.2(a)). In other evolutionary theories, the first step involves a stimulus evaluation process (after or instead of the perception process) that produces a representation with abstract features ([eaS]) such as (molar or molecular) appraisal values (Tracy, 2014). Theories that add stimulus evaluation can be considered as hybrids of evolutionary and stimulus evaluation theories (see Figure 4.2(b)). It must be noted, however, that the nature of this evaluation process varies across scholars and even across writings of the same scholar. Sometimes it is molded after the appraisal process in appraisal theories (Tracy, 2014); sometimes it is understood as a simple perceptual matching process (i.e., the matching of raw stimulus input with an internal perceptual template) (Matsumoto & Ekman, 2009); and sometimes as a process that must determine that stimuli pertain to emotion and to which one they pertain (Ekman, 1992a). Tooby and Cosmides (1990, pp. 407–408), moreover, suggested that the evaluation process itself may be shaped by evolution. It is prewired towards detecting and processing stimuli in terms of fundamental challenges and opportunities.

Steps 2 and 3

In the second step, the afferent representation ([ecS]/[eaS]) from Step 1 is linked to an efferent representation ([R]) in an affect program. Like the instincts in McDougall's (1908) theory, affect programs are installed by evolution. They contain preset instructions for how to deal with certain stimuli based on how our evolutionary ancestors dealt with them. The stimuli pose "fundamental life tasks," that is, prototypical survival-related challenges and opportunities, such as threats, offenses, losses, gains, and noxious substances (Ekman, 1992a; Levenson, 2011). The responses are those that evolution has selected as the best solution to deal with these stimuli (Ekman, 1992a; Johnson-Laird & Oatley, 1992). As Levenson (2011) put it, the ubiquity and importance of these challenges and opportunities created selection pressures to favor generalized solutions that have the highest survival value for the individual (and the group). As such, emotions "are the time-tested solutions to these timeless problems. Because of their importance, having each individual learn each solution de novo would be inefficient and uncertain" (Levenson, 2011, p. 382).

The affect program for each emotion type is not only innate (i.e., hardwired in the genes) but also brain-bound (i.e., hardwired in the brain), meaning that each emotion type has its own dedicated neural mechanism or "fingerprint."[44] This need not be a specific brain area but can also be a specific neural circuit comprised of interacting areas (including the neurotransmitters involved; see Panksepp, 1998) or even a specific brain network (Nummenmaa & Saarimäki, 2019). In sum, affect programs can be

[44] Hutto et al. (2018, p. 7) believe that evolutionary theories are not committed to the idea that affect programs are brain-bound, based on Ekman's (2003, p. 66; see also Ekman & Cordaro, 2011) claim that affect programs are a metaphor rather than a literal set of instructions localized in a specific part of the brain. I think this is a mistake. What Ekman (2003) tried to debunk was the idea that the instructions in affect programs are symbolically represented in a specific part of the brain, not the idea that they are brain-bound.

understood on the mental super-level as [S–R] links and on the brain super-level as neural mechanisms.

Returning to the mental super-level, in some theories, the afferent part of the [S–R] link is framed in terms of the concrete features of a small set of USs (e.g., loud noises, sudden loss of support, predators, dark places, strangers) (Izard, 2011; Levenson, 2011) or genetically prepared stimuli (i.e., which require a minimum of training; Öhman, 1986). In other theories, the afferent part of the [S–R] link is framed in more abstract terms such as threats, offenses, losses, gains, and so on. The format of the affect program that is chosen has implications for the type of cognitive process in Step 1 that must precede its activation: Affect programs of the type [*snake*–flee] must be preceded by a perceptual process yielding concrete stimulus representations (snake; [ecS]) whereas affect programs of the type [*danger*–flee] must be preceded by a stimulus evaluation process yielding abstract stimulus representations (danger; [eaS]).

The efferent part ([R]) of the [S–R] link holds instructions for (a) somatic responses, (a) subtle behavior in different modalities (face, voice, gestures), and (b) coarse behavior such as fight and flight. Once activated, the affect program automatically triggers somatic responses (sR) as well as subtle (smR) and coarse motor responses (cmR). The affect program belonging to a certain emotion also generates feelings characteristic of that emotion.

Some evolutionary theories follow McDougall's (1809) centralist view that feelings are the direct consequence of the affect program (e.g., Izard, 1977; Oatley & Johnson-Laird, 1987; Plutchik, 1980). These authors also see these feelings as having irreducible qualia (see Chapter 2). Other evolutionary theories return to James's (1890b) peripheralist view that feelings are generated by somatosensory feedback (e.g., Damasio, 1994, 1999, 2004;[45] Levenson, 2014; J. J. Prinz, 2004a). This has earned them the

[45] Damasio (1994) eventually proposed to divorce emotions, which he restricted to affect programs and the bodily responses directly triggered by them, from feelings, which are the depositions of these bodily responses in consciousness. This allowed him to make the further claim that emotions can remain unconscious until they are made conscious in feelings. Damasio (1994) further proposed that the function of these "gut" feelings is to guide decision making. Specifically, he proposed that past behavioral choices are stored together with the gut feelings that mark these choices as positive or negative in the light of their outcomes. When confronted with the same choice options on a new occasion, the positive or negative somatic markers of the options are re-activated in the absence of any conscious memory of the outcomes, steering agents towards positively marked options and away from negatively marked ones (for evidence, see Bechara et al., 1997, 1999; Damasio, 1994; for counterevidence and criticism, see Beer, 2017; Newell & Shanks, 2014). Finally note that Damasio (1994, p. 144) allowed mental images of the eliciting stimuli to be part of feelings, in juxtaposition to the experience of bodily changes. Since the cognitive component housing these mental images is also a

https://doi.org/10.1017/9781107588882.008 Published online by Cambridge University Press

name of neo-Jamesian theories, although they are perhaps more accurately named evolutionary-Jamesian hybrids.

Like the instincts in McDougall's (1908) theory, affect programs come in limited numbers that either alone or combined with other affect programs or other processes correspond to vernacular emotion types. Basic emotions are ones that have their own affect program. Although there is disagreement about their number and identity, many theorists include at least fear, anger, sadness, joy, disgust, and surprise. Non-basic emotions are ones that do not have their own affect program but that combine, or elaborate on, the affect programs of other basic emotions (see below). Some emotions are still in limbo land; their basic-emotion status is still under study, although the list is growing (see Section 4.4).

Like their predecessors, evolutionary theories assume that innate affect programs (i.e., primary processes) do not operate in isolation, but interact with learning (i.e., secondary processes) and computation (i.e., tertiary processes). The contribution of affect programs to emotions is phylogenetic, that of learning ontogenetic, and that of computation microgenetic. The question to ask is which part of the causal sequence can be changed by ontogenetic and microgenetic processes. The plasticity model endorsed by evolutionary theories is the avocado pear model (Goldie, 2000, p. 99; see Faucher & Tappolet, 2008, p. 111, for a review of six plasticity models): The core, which is the affect program (Step 2), is not malleable and forms the pure emotion; the soft flesh, which are the processes that happen prior to (Step 1) or after the affect program (Step 3), can be superficially modified by learning and computation (see also McDougall, 1908). On the input side (Step 1), learning processes can play a role, for instance, by making sure the genetically prepared stimuli get consolidated in representations (Öhman, 1986) or by extending the range of stimuli that can trigger the affect program (on the basis of similarity and co-occurrence) (Ekman, 1992a; Ekman & Cordaro, 2011; Izard, 2011; Levenson, 2011). Learning and computation are further involved in the stimulus evaluation process postulated by evolutionary-evaluation hybrid theories.

On the output side (Step 3), learning and computation are involved in emotion regulation and planning.[46] The default scenario is that the affect program automatically triggers all output components. Deviations from this scenario occur when emotion regulation steps in. Emotion regulation is thought to be subserved by goal-directed processes (Scarantino, 2014;

central component, Damasio (1994) cannot be said to be a pure peripheralist (see also Dub, 2022).

[46] Other processes that Ekman (1992b; Ekman & Cordaro, 2011) situated on the output side are the verbal labeling of emotions and the attitudes one holds towards them.

https://doi.org/10.1017/9781107588882.008 Published online by Cambridge University Press

see Box 7.1). If an initial, emotional action tendency does not fit with one's (long-term) goals, such as when an aggressive tendency threatens to blow up one's relationship, attempts will be made to suppress its manifestation. People are expected to have more control over their coarse behavior than over their facial and vocal expressions and no direct control over their physiological responses (Ekman, 1992a; see also McDougall, 1908). They can try to wipe off their facial expressions in an attempt to follow cultural *display rules* but occasional leakage may occur in the form of micro-expressions (Ekman, 2006; Ekman & Friesen, 1969). They may even try to regulate their feelings in an attempt to follow cultural *feeling rules* (Hochschild, 1979). In theories in which feelings are produced centrally, however, it should be virtually impossible to stop them once the affect program is activated.

Goal-directed processes are also granted a role in planning or behavioral concretization, that is, in the translation of the initial action tendency into a more concrete action tendency or motor program. The tendency to aggression may take the form of yelling, hitting, boycotting, turning one's back, or giving someone the silent treatment, and the choice between these depends on weighing up expected utilities given the specific situation or context one is in. Anger in traffic will manifest itself in yelling rather than hitting, for instance. The division of labor between a stimulus-driven (innate [S–R]) process that generates the initial emotional action tendency and the intervention of non-automatic goal-directed processes to regulate or refine this action tendency is consistent with a dual-system view with a default-interventionist architecture (see Box 2.1).[47]

At some point, Ekman and Cordaro (2011) started poking somewhat inside the pit of the avocado pear when they proposed that ontogenetic processes can also enter the affect program itself. They argued that an affect program is not a closed program in which "nothing can be inserted by experience" (p. 367), but rather an open program that "allows for additional input during the life-span of its owner" (p. 367; distinction after Mayr, 1974). They proposed, for instance, that learning can influence the preset instructions for physiological responses and expressions. That said, they added that learning these responses often develops in the same way for everyone, unless the person is confronted with unusual,

[47] It may be noted that some evolutionary theorists such as Scarantino (2014) proposed a parallel-competitive architecture in which stimulus-driven and goal-directed processes operate in parallel and compete with each other, but in which the stimulus-driven process often wins the competition because it is seen as automatic and the goal-directed process as non-automatic. As mentioned in Box 2.1, footnote 3, this brand of parallel-competitive architecture is almost indistinguishable from the default-interventionist architecture.

traumatic experiences. Learning can also influence the preset instructions for coarse behavior, but here more radical deviations from the innate instructions are possible (Ekman & Cordaro, 2011, p. 367). Once learned instructions have entered the affect program, they can be triggered and executed in the same automatic way as the innate instructions can.

It could be argued that by allowing affect programs to be open, Ekman and Cordaro (2011) landed on a slippery slope between evolutionary theories (which put phylogeny first) and network theories (which put ontogeny first; see Chapter 5). To avoid a complete slide-off, however, Ekman and Cordaro (2011) pointed out that learning cannot go so far as to delete or rewrite the innate instructions in affect programs. Novel instructions can be added, but the innate ones remain intact. This means that they can still rear their heads under the right circumstances, as can be seen, for instance, in the dramatic startle reaction that people have to unexpected loud noises. In Levenson's (2011, p. 382) words, "[t]he influence of basic emotions on behaviors and thoughts becomes most deterministic under those conditions in which antecedent conditions closely match prototypical elicitors ... When these conditions are not met, the plasticity and flexibility of the emotion system becomes more ascendant."

Ekman and Cordaro's (2011) proposal notwithstanding, some scholars (e.g., Griffiths, 1997; Scarantino, 2014, 2015, 2018) seem convinced that evolutionary theories leave affect programs open on the input side, but keep them closed on the output side. "We have to learn what to be afraid of, but not how to act afraid" (LeDoux, 2003, p. 213). In order to increase the power of evolutionary theories to account for flexible stimulus-response relations, Scarantino (2014, 2015) therefore proposed to adjust them by also opening up the output side. He dubbed his adjustment the "new basic emotion theory." I believe Ekman and Cordaro (2011) had already opened up the output side, albeit in a different way than Scarantino (2014) did. While the former authors allowed novel instructions to sit next to, and interact with, existing ones, Scarantino's (2014) solution consisted in the abstract reframing of affect programs, replacing more concrete affect programs of the type [snake–flee] by more abstract affect programs of the type [danger–defense]. Once danger has triggered the tendency to defend oneself or to regain safety, the behavioral options are not fixed, but are flexibly selected via a goal-directed process that is tailored to the affordances in the environment. Safety can be regained by fleeing but also by fighting or freezing.

Scarantino's (2014, 2015) solution was in fact not so new. Hybrids of evolutionary and stimulus evaluation theories had already proposed that the [S] part of the affect program can be abstract. For several theorists, this also counted for the [R] part. For instance, Oatley and Johnson-Laird

(2011; see also Roseman, 2011)[48] argued that emotions, which are installed by evolution, do not lead to fixed action patterns but rather constrain the action repertoire. Such a constraining would be useful when an organism is confronted with a novel, unexpected stimulus for which a rapid response is required, while at the same time leaving room for flexible responding depending on the specifics of the environment.

Leaving aside the question of novelty, however, I do not think the abstract (re)framing of affect programs makes them more open. The link between the abstract [S] and abstract [R] remains as fixed as before. It is true that making an affect program more abstract means that the concrete action tendency is less determinate. Saying that danger leads to the tendency to regain safety does not tell us yet whether the person will run away or hide. But the job must still be done. A concrete action tendency must still be determined. To this end, a goal-directed planning process is invoked. This process, however, is not considered to be part of the emotion itself. Thus, the flexibility is not delivered by the emotion but by an extraneous, non-emotional process.

Summing up, affect programs are innate [S–R] connections that are rigid. To increase flexibility between actual stimuli and actual responses, evolutionary theories add learning and computation to the mix. Generalization and classical conditioning increase the range of stimuli that can trigger the [S–R] connection on the input side; goal-directed regulation and planning increase the range of responses that can follow this connection on the output side. Authors have chosen to frame their [S–R] links on more concrete or more abstract levels, and this has the following implications. First, on the input side, concrete links can be triggered by a primitive perception process whereas abstract links must be preceded by a stimulus evaluation process. Second, to increase response flexibility on the output side, concrete links require a larger share of regulation processes (e.g., to change fight into flight) whereas abstract links require a larger share of planning (e.g., to translate defense into flight).

4.3 Scientific Definitions

4.3.1 Intensional Definition

4.3.1.1 Criteria for Demarcation

The intensional definition of evolutionary theories is straightforward: An episode is emotional if an affect program is activated. Let us call this the

[48] Note that Oatley and Johnson-Laird (2011) and Roseman (2011) fall more on the side of evaluation-evolutionary hybrids (i.e., a version of stimulus evaluation theories, see Chapter 6) than on the side of evolutionary-evaluation hybrids (i.e., a version of evolutionary theories).

https://doi.org/10.1017/9781107588882.008 Published online by Cambridge University Press

"affect program" criterion. This remains true regardless of whether the boundaries of the emotion proper are put around the affect program itself, only around the somatic, motor, and feeling components that form the output of the affect program, or around the entire emotional episode. Theories with a wide constitutive explanation can also invoke the presence of the output components as a criterion. Let us call this the "components" criterion. The status of this criterion is not so much one of necessity, however, but rather one of typicality (as in a cluster-type definition). This is because the output components, although supposed to flow automatically from the affect program, can be held back to a greater or lesser extent by emotion regulation.

4.3.1.2 Adequacy

APPARENT-SIMILARITY

How does this intensional definition fare in terms of the properties of emotions specified in the working definition from Chapter 2? The response and feeling components that are generated by the affect program provide the bodily and phenomenal aspects of emotion, respectively.

Affect programs are brain circuits on the brain super-level, but representations of [S–R] links on the mental super-level. Because they are representations, they carry Intentionality. The stimuli represented in the [S–R] links are the objects, which ensure world-directedness. Non-hybrid theories, which restrict the first step to a simple perception process, however, are amenable to the same criticism as James (1890b) and McDougall (1908) that representations with concrete features do not provide the right kind of object (Deonna & Scherer, 2010). Fear is not about the coiling form of a snake, but about danger. Hybrid theories, which allow a (sufficiently abstract) evaluation process to precede the affect program, on the other hand, effectively solve this problem, as they provide stimulus representations with abstract features such as danger, offense, loss, and so on. The response representations or action tendencies in the [S–R] links display self-directedness.

Like the instincts in James's (1890b) and McDougall's (1908) theories, the innateness of the affect programs in evolutionary theories can easily account for the continuity in emotional life between adults, infants, and animals. The increasing importance of learning and computation with age and across the phylogenetic ladder can explain the increasing variety in the stimuli that can trigger emotions and the increasing complexity of their manifestations. Affect programs, moreover, are supposed to be automatic and therefore to take control precedence.

The fact that affect programs are stimulus-driven processes that do not take into account the current outcomes of behavior should lead us to conclude that emotions are practically irrational, at least in the

https://doi.org/10.1017/9781107588882.008 Published online by Cambridge University Press

process-sense (Greenspan, 2004; Griffiths, 1997; Moors, 2017b; Tooby & Cosmides, 1990, pp. 418–419). Yet several evolutionary theories turn out to emphasize the adaptiveness of the behavioral solutions of our evolutionary ancestors (Ekman, 1992a; Shariff & Tracy, 2011b; Tomkins, 1962, but see Griffiths, 1997; Tooby & Cosmides, 1990). This seems to be at variance with their predecessors (Darwin, 1872; James, 1890b; McDougall, 1908), who emphasized that ancestral responses are often vestigial and hence no longer useful (Barrett, 2013; Fridlund, 1994). Jumping away from a snake in the zoo is not adaptive.

Three replies are possible. The first reply is the idea discussed earlier that stimulus-driven processes can occasionally lead to adaptive behavior provided that the reward structure of the current context matches the reward structure of the ancestral context. Thus, jumping away from a snake in the zoo is maladaptive whereas jumping away from a snake in the wild continues to be adaptive. The second reply, based on a dual-system view, is that innate [S–R] links are considered dirty but quick. They are dirty (biased) in the sense that they are triggered by minimal evidence, but this comes with the benefit that they are also quick (automatic). A process that is biased in this liberal way is more likely to produce false alarms but less likely to avoid misses. In cases in which it is better to be safe than sorry, innate [S–R] links may be practically rational after all. Jumping away from the slightest snake-like stimulus leads to false alarms in the zoo but makes sure that not a single snake is missed in the wild (Goffin, 2021). The third reply follows Darwin (1872) in arguing for the co-option of vestigial expressions as a means of communication so that they can still be adaptive today, if not to influence the physical environment then at least to influence the social environment. There is no doubt that information about a sender's emotion may be useful for the receiver. If the sender displays fear, this may help the receiver to prepare for danger. From the viewpoint of the sender, on the other hand, giving away one's emotions can be either beneficial or detrimental depending on whether the receiver is an ally, who wants to help, or a competitor, who is able to exploit this information (Fridlund, 1994, p. 109, 1997; Hinde, 1985b). Signaling distress is beneficial for an infant as this may recruit caretakers to alleviate the source of distress. In a poker game, on the other hand, it is more adaptive to conceal one's emotions. Note, finally, that evolutionary theories do not deny that emotions can also be maladaptive under some circumstances. The fact that they reserve an important role for emotion regulation underscores this.

FRUITFULNESS

The affect program criterion might be seen as bringing some sort of abstract unity to the set of emotions, but since all affect programs are different and therefore do not share an essence, the total set of emotions does not count as fruitful. Early proposals that all (or most) emotions are located in the visceral brain or limbic system whereas emotion regulation originates from the neocortex (MacLean, 1949) do not seem to cut it anymore (Calder et al., 2001; LeDoux, 1991, 2012a; Lindquist & Barrett, 2012; Mobbs et al., 2019).

4.3.2 Divisio Definition

4.3.2.1 Criteria for Partitioning

Evolutionary theories endorse a discrete approach to organizing the variety within the set of emotions. The set is populated by discrete emotions that each carry a name in natural language. These can be grouped into a subset of *basic* emotions, each of which has its own affect program, and a subset of *non-basic* emotions, which ride on the affect programs of basic emotions (see Table 4.1).[49] The precise lists of basic and non-basic emotions vary across authors. The subset of basic emotions further splits into the *primitive* form of these emotions, which include nothing but the innate affect program, and a *developed* form, which are complemented by learning and computation (Izard, 2011; Keltner & Haidt, 2001; Panksepp, 2012).[50] Primitive basic emotions occur in early infancy and are ubiquitous in non-human animals. They are triggered by USs such as loud noise and sudden loss of support. Developed basic emotions progressively occur in later stages of phylogeny and ontogeny. They are triggered by CSs such as cars, guns, and impending school exams.

The subset of non-basic emotions further splits into *mixed* emotions and *elaborated* emotions. Examples of mixed emotions are nostalgia (mixing sadness and joy), awe (mixing fear and surprise), contempt (mixing anger and disgust), scorn (mixing disgust and joy), and smugness (mixing contempt and joy) (Ekman, 1992a, 1999a; Plutchik, 1980, 2001). Mixed emotions can be (a) *blends*, in which

[49] In evolutionary theories, the basis for inclusion of an emotion in the set of basic emotions is evolutionary-biological. There are also theories that propose instead a conceptual, psychological, or social basis for inclusion in this set (see Ortony & Turner, 1990; Solomon, 2002).

[50] Keltner and Haidt (2001) call primitive basic emotions primordial emotions; Izard (2011) calls them first-order emotions. Note that Izard (2011) allows some influence of learning already on his first-order emotions in that these are not only elicited by USs but also by CSs.

Table 4.1. *Divisio definition of emotion by evolutionary theories*

<table>
<tr><th colspan="2"></th><th>only affect program</th><th>affect program + learning + computation</th><th>only learning + computation</th></tr>
<tr><td rowspan="2">PRIMARY EMOTIONS:
have an affect program</td><td>BASIC EMOTIONS:
have their own affect program: e.g., fear, anger, sadness, joy, surprise, disgust</td><td>PRIMITIVE EMOTIONS
e.g., primitive anger</td><td>DEVELOPED EMOTIONS
e.g., developed anger</td><td></td></tr>
<tr><td>NON-BASIC EMOTIONS:
ride on the affect program of basic emotions</td><td>MIXED EMOTIONS
e.g., nostalgia, awe, scorn, smugness
- blends
- rapid alternations</td><td>ELABORATED EMOTIONS
e.g., jealousy</td><td></td></tr>
<tr><td colspan="2">SECONDARY EMOTIONS:
have no affect program (exist only in split version of evolutionary theory)</td><td></td><td></td><td>PENSIVE EMOTIONS
e.g., admiration, adoration</td></tr>
</table>

two or more affect programs are activated simultaneously or (b) *rapid alternations*, in which affect programs occur sequentially (i.e., a possibility that Ekman, 1992a, reported having observed more frequently). Elaborated emotions, from their side, combine the affect program of one or more basic emotions with a specific type of stimulus. An example is jealousy, which combines the affect programs of fear, anger, and/or sadness with a particular cognitive plot, namely that another person has or threatens to have what you feel entitled to (Ekman, 1999a; Frijda, 1986). The question can be raised of whether there are any grounds for distinguishing between elaborated emotions (e.g., jealousy) – which are counted among the non-basic emotions – and the previously mentioned developed basic emotions (e.g., developed anger) – which are counted among the basic emotions. One difference that can be noted is that the former take on a different name than the basic emotion they derive from (e.g., jealousy is a different word than fear/anger/sadness) whereas the latter keep the same name (e.g., developed anger is still called anger).

By including developed basic emotions and non-basic emotions, evolutionary theories are able to cover a wide range of emotions that can occur in complex interpersonal situations. However, emotions without an affect program such as certain pensive emotions still fall outside the scope of these theories. Griffiths (1997) therefore proposed a split version of the evolutionary theory with (a) a subset of *primary* emotions, in which an affect program is activated (basic and non-basic emotions) and (b) a subset of *secondary* emotions, in which no affect program is activated and that are only governed by learning and computation. This proposal is also known as the "disunity thesis" (e.g., J. J. Prinz, 2004a). Deonna and Scherer (2010, p. 49; see also J. J. Prinz, 2004b) objected that the separation between primary and secondary emotions does not match with the obvious continuity between the two types of emotions: The fear of a viciously barking dog does have similarities with the fear "when I realize that I left my Stradivarius in the taxi 15 minutes before the concert." But perhaps this is not a suitable counterexample as it can be argued that the fear caused by the forgotten Stradivarius does not match the description of a secondary emotion, but rather the description of a developed basic emotion, which is a subset of primary emotions. A better example of a secondary emotion may be the fear that it will rain during the picnic tomorrow or the regret that you did not take your sunglasses when it turned out to be sunny.

4.3.2.2 Adequacy

APPARENT-SIMILARITY AND FRUITFULNESS

Evolutionary theories try to vindicate vernacular emotions. Their mechanisms are invented with the specific aim of meeting the similarity meta-criterion. In addition, if each basic emotion has its own unique affect program, fruitfulness is also guaranteed. Of course, everything will stand or fall with the existence of these affect programs. Merely inventing them is not enough, their existence should

https://doi.org/10.1017/9781107588882.008 Published online by Cambridge University Press

also be demonstrated empirically. This is the topic of the next section.

4.4 Validation

I first discuss approaches, methods, and evidence in favor of evolutionary theories (Section 4.4.1). After that, I discuss criticism and evidence contra evolutionary theories, replies, and counter-replies (Section 4.4.2).

4.4.1 Evolutionary Research Program and Evidence Pro

A first aim of the evolutionary research program is to demonstrate that some emotions are indeed hardwired. This entails showing that affect programs exist and that they cause the output components in emotional episodes. In addition to this "proof of principle" aim, another more exploratory aim is to discover which emotions are hardwired. The second aim is an extension of the first aim. Once a few emotions are already found to be hardwired (and preferably more than just positive and negative emotions), evidence for the first aim is achieved. To accomplish the second aim, more comprehensive and systematic research efforts are required. The present section focuses on the first part of the first aim – to demonstrate that affect programs exist for some emotions. Evidence for the existence of emotion-specific affect programs can be direct or indirect (Ekman, 1992a; Ortony & Turner, 1990). I discuss both types of evidence in turn.

4.4.1.1 Direct Evidence

The search for direct evidence has followed two approaches (T. J. Turner & Ortony, 1992). A first approach is to start from emotions that have a label in language, the vernacular emotions, and to examine whether these are hardwired in humans. Studies use brain imaging methods (F. C. Murphy et al., 2003; Phan et al., 2002; Vytal & Hamann, 2010) and the stimulation or blocking of particular brain circuits by pharmacological means or repetitive TMS (George et al., 1996; Panksepp & Watt, 2011). This research has led to the discovery of circuits involving the amygdala for fear (M. Davis, 1992; Johansen et al., 2011), the insula and globus pallidum for disgust (Wicker et al., 2003; P. Wright et al., 2004), the lateral orbitofrontal cortex for anger (F. C. Murphy et al., 2003), the subcallosal cingulate cortex (SCC) for sadness (Phan et al., 2002; Vytal & Hamann, 2010; but see F. C. Murphy et al., 2003), and the basal ganglia for joy (Phan et al., 2002).

A second approach (Panksepp, 2012) is to start from the brain circuits for fundamental functions that have been discovered in animal research and to

https://doi.org/10.1017/9781107588882.008 Published online by Cambridge University Press

use these as the basis to determine the list of basic emotions – also for humans – without striving for perfect correspondence with existing lists of vernacular emotions. Common methods used in these studies are the direct stimulation or blocking of particular brain circuits by means of dissection (Panksepp, 1982, 1998, 2000; reviews by Izard, 2007; Panksepp, 2007). Using this approach, Panksepp (2012) is confident about the existence of subcortical brain systems for the primitive basic emotions of FEAR, RAGE, PANIC/GRIEF, LUST, CARE, PLAY, and SEEKING (which he capitalized to mark the difference from the vernacular, developed forms of basic emotions).

4.4.1.2 Indirect Evidence

There are three kinds of indirect evidence based on the following inferences (Ekman, 1992a, 1992b; Ortony & Turner, 1990). First, *if* an emotion has a dedicated affect program with a unique function, *then* it is likely to have distinct output components: physiological responses, subtle and coarse behavior, and feelings. Second, *if* an emotion has a dedicated affect program, which is considered to be the common cause of the distinct output components for that emotion, *then* these output components should correlate. Third, *if* an emotion has a dedicated affect program, which is innate, *then* it should be universal. As a consequence, the output components caused by these affect programs should be universal as well, which means that they are consistent across cultures, developmental (ontogenetic) stages, and species that are phylogenetically close (e.g., humans and primates). The indirect research program of evolutionary theories considers for each candidate basic emotion whether there is evidence for distinct outputs (in short, distinctness), evidence for correlations among these distinct outputs (in short, concordance), and evidence for universality of these distinct outputs (in short, universality). Each piece of indirect evidence is taken to strengthen the hypothesis that the emotion under consideration is indeed a basic emotion.

(1) Research on *distinctness* has focused mostly on somatic responses and facial expressions. In research on emotion-specific ANS activity, emotions are manipulated indirectly by instructing participants to take on artificial facial movements and/or to relive certain feelings while ANS activity is measured with direct objective methods (e.g., Ekman et al., 1983; see also Averill, 1969; Ax, 1953; Christie & Friedman, 2004; Funkenstein et al., 1954; Levenson, 1992; Levenson et al., 1990; reviews in Ekman, 1992a; Kreibig, 2010; Levenson, 2014) or direct subjective methods (e.g., Nummenmaa et al., 2014). Using the former method, Ekman et al. (1983) found different patterns of ANS activity (across heart rate, skin temperature, and skin resistance) for fear, anger, sadness, happiness, disgust, and surprise.

Research on emotion-specific facial expressions (reviews in Ekman, 1992c, 1999b; Keltner et al., 2016; Matsumoto, Keltner, et al., 2008) was dominated in its early days by (a) perception studies in which participants were asked to judge which emotion words matched pictures of posed facial expressions, the so-called judgment method, and (b) to a lesser extent, production studies in which participants were instructed to produce facial expressions corresponding to emotion words and in which the resulting expressions were measured by independent observers using the judgment method (Ekman, 1972). This research confirmed the existence of stereotypic facial expressions for fear, anger, sadness, joy, disgust, and surprise.

(2) If the component values of an emotion indeed have an affect program as a common cause, they are likely to co-occur. The tendency to fight, for instance, should correlate with fighting behavior, blood flowing to the hands, a scowling face, and angry feelings. Research on *concordance* induces emotions and measures whether there is a correlation among two or more component values that are hypothesized to be specific for that emotion (reviews in Levenson, 2014; Matsumoto, Keltner, et al., 2008).

(3) To examine the *universality* of emotion-specific physiological responses and facial expressions, the methods used in Western samples have also been used to compare the physiological responses and facial expressions across cultures, including non-Western and isolated cultures. This research yielded evidence for cross-cultural consistency in somatic responses (e.g., Levenson et al., 1992) and facial expressions (see first reports in Darwin, 1872, and controlled studies by Ekman, 1972; Ekman et al., 1969; Elfenbein & Ambady, 2002; Izard, 1969; Tracy et al., 2013). Some studies report evidence for the universality of facial expressions based on consistency between the expressions of congenitally blind and sighted persons. For instance, Matsumoto and Willingham (2009) took the consistency between the facial expressions of blind and seeing athletes after having won or lost an important competition as a particularly strong argument for the existence of innate affect programs. A final piece of evidence for universality comes from ethological studies showing consistency between the facial expressions of humans and closely related species (Parr et al., 2007; Tomonaga et al., 2004; Tracy & Matsumoto, 2008; Vick et al., 2007).

4.4.2 Criticism and Evidence Contra

Although the evolutionary research program has yielded a rich body of empirical work, this work has met with severe criticism. A first type of criticism questions the empirical status of the findings. A second type of criticism questions the interpretation of the findings as evidence for the evolutionary theory. I discuss both types of criticism in turn, as well as the replies that have been offered by evolutionary theorists.

4.4.2.1 Criticism 1: Empirical Status of Findings

CRITICISM

The first criticism targets the robustness of the empirical evidence. While evolutionary theorists argue that much of the evidence is convincing and other evidence is encouraging, critics argue that it is disappointingly weak, not consistent across studies, and/or fraught with methodological problems (see reviews by Barrett, 2006a, 2011, 2012, 2017a, 2017b; Barrett & Satpute, 2019; Lindquist & Barrett, 2012; Lindquist et al., 2012; Ortony & Turner, 1990; Russell, 1994). A few illustrations of problems:

(1) Evidence for *distinctness* in central and peripheral nervous system activity has been underwhelming. Direct evidence for affect programs reported in certain meta-analyses (e.g., F. C. Murphy et al., 2003; Phan et al., 2002) turned out to be only partial and not consistent across meta-analyses (Barrett, Lindquist, Bliss-Moreau, et al., 2007; Barrett & Wager, 2006). Other meta-analyses found no direct evidence (Kober et al., 2008; Lindquist et al., 2012). The same conclusion has been drawn for indirect evidence in the form of physiological responses specific to basic emotions (e.g., Cacioppo et al., 2000; Kreibig, 2010; J. T. Larsen et al., 2008; Quigley & Barrett, 2014; Siegel et al., 2018). Kreibig (2010) suggested that emotion-specific ANS activity only obtained when emotions were divided into subtypes according to specific stimuli. For instance, different patterns of ANS activity emerged for disgust for contamination-related stimuli (e.g., rotten food, dirt) and disgust for mutilation-related stimuli (e.g., injured bodies, blood). Siegel et al.'s (2018) meta-analysis went further in suggesting that ANS patterning is only stimulus-specific and not emotion-specific, which led them to conclude that there are no emotional "fingerprints" to be found in ANS activity.

(2) Evidence for *concordance,* that is, correlations among the distinct outputs of affect programs is mixed. Some studies show concordance (Bonanno & Keltner, 2004; Mauss et al., 2005; see reviews by Evers et al., 2014; Lench et al., 2011) whereas others show dissociations instead (Bradley & Lang, 2000; Darrow & Follette; 2014; Lang, 1968; Mandler et al., 1961; Reisenzein, 2000; Reisenzein et al., 2006; see reviews by Durán et al., 2017; Evers et al., 2014; Fernández-Dols & Crivelli, 2013; Hollenstein & Lanteigne, 2014; Kanter et al., 2014; Mauss & Robinson, 2009; Reisenzein et al., 2013).

(3) Research on *distinctness* in facial expressions and the *universality* of these expressions has received extensive methodological criticism, and

replications with more suitable methods have often failed to yield the predicted results (Gendron et al., 2014a; Nelson & Russell, 2013; Russell, 1994; Russell et al., 2003). For instance, both the judgment method and the instructed production method have been found wanting. It was argued that if participants have to judge emotions from posed facial expressions or have to produce artificial facial expressions based on emotion words, they may tap into their learned stereotypical scripts to solve the task (Barrett, 2011; Fridlund, 1994; Lindquist & Gendron, 2013; Parkinson, 2013).

Recent studies using a spontaneous production method, however, have apparently failed to support the presumed link between facial expressions and basic emotions, both in the laboratory (Reisenzein et al., 2013) and in the field (Fernández-Dols & Crivelli, 2013; see Durán et al., 2017). Put simply, people do not always and do not only smile when they are happy, but also when they are proud, embarrassed, or in pain (Russell, 2017). Further, production studies in which facial expressions have been measured using EMG rather than with the judgment method yielded mixed evidence for emotion-specific facial expressions. Some researchers have reported no confirmatory evidence (Cacioppo et al., 2000; Russell et al., 2003). Others have reported only a limited number of the predicted facial expressions in sighted participants (Galati et al., 1997), infants (Bennett et al., 2002; Camras & Fatani, 2008), and congenitally blind adults or children (Galati et al., 1997; Roch-Levecq, 2006). Furthermore, recent cross-cultural studies have shown a lack of consistency between Western and remote cultures when methodological flaws have been corrected (e.g., Crivelli & Gendron, 2017; Crivelli, Jarillo, & Fridlund, 2016; Crivelli, Jarillo, Russell, & Fernández-Dols, 2016; Crivelli, Russell, et al., 2016, Crivelli et al., 2017; Gendron et al., 2014a, 2014b, 2018, 2020; Nelson & Russell, 2013). Finally, it has been argued that ethological findings rely on subjective judgments by observers and that their replicability is unknown (Barrett, 2013).

Taken together, failed attempts to produce direct and indirect evidence have yielded evidence for variety instead of consistency in the neurological correlates and output components of candidate basic emotions, and for dissociations instead of concordance among the output components. The variety holds across and within groups of participants (cultures, developmental groups, seeing/blind groups, species) but also within participants across contexts (Barrett, 2006b, 2017b; Barrett et al., 2009; Barrett, Lindquist, Bliss-Moreau, et al., 2007; Cacioppo et al., 2000; Hunt, 1941; Kagan, 2007; Lindquist et al., 2012; Mauss & Robinson, 2009; Ortony & Turner, 1990; Russell, 2003).

Replies and Counter-Replies

Evolutionary theorists have adopted three strategies to counter the first criticism that the data are not robust and show variety.

Strategy 1. The first strategy has been to emphasize that part of the variety that is observed in studies is already predicted by evolutionary theories, either because the variety is inherent in emotions themselves (see first and second sources of variety below) or because the variety stems from non-emotional processes that the theory acknowledges (see third and fourth sources of variety below). Another part of the variety is not predicted by the theory, but can be attributed to methodological constraints (see fifth source of variety below). I discuss each of these sources of variety as well as ways in which research can deal with them.

(1) A first source of variety is the combination of affect programs in mixed emotions. This can in principle be controlled for by striving for emotion induction procedures that target a single emotion instead of mixed emotions, although this will be easier said than done.

(2) A second source of variety is captured by the notion of emotion families in which the members have an innate theme and variation stems from learning and computation. One illustration is the variety in the disgust family between a primitive form of disgust and a developed form of disgust for moral transgressions. Primitive disgust may be manifested in the tendency to eject matter whereas the evolved form may be characterized by symbolic rejection. This intra-emotion variety is likely to work against evidence for distinctness. A lack of distinctness, in turn, is likely to produce a lack of concordance. In addition, given that learning is the vehicle of socio-cultural influences, it is likely to produce differences between groups. In this spirit, Elfenbein et al. (2007) proposed that differences in facial expressions between cultures should be understood as "regional dialects" that formed on top of a universally shared expressive "language." Learning further produces differences between individuals in the same group and computation accounts for differences within individuals over occasions. To obviate this source of variety, evolutionary theorists have resorted to studies with infants and lab-reared animals. The influence of learning is assumed to be low in infants and controlled in lab-reared animals and the influence of computation is assumed to be low in both (Watson, 1919).

Another illustration is the variety listed in Plutchik's (2001) emotion families. His fear family, for instance, is composed of apprehension, fear, and terror, each of which is associated with different action tendencies: apprehension with the tendency to be vigilant, fear with the tendency to avoid or flee, and terror with the tendency to fight or freeze. The variety in the action tendencies within this family seems to be a matter of the intensity of the input, corresponding to the imminence of the threat posed by the stimulus, rather than of more or less developed forms of fear.

https://doi.org/10.1017/9781107588882.008 Published online by Cambridge University Press

Moreover, the variety at stake here is not moderate (as in dialects) but threatens the very boundaries of the emotion subsets. If the fear family includes not only instances characterized by the tendencies to flee, but also instances characterized by the tendency to fight, a clear separation from the anger family is jeopardized (see Barrett et al., 2009; Blanchard & Blanchard, 2003; Lang et al., 1990; Russell, 2009).

To solve this problem, evolutionary theorists have resorted to three solutions. A first solution is to argue that the hypotheses of evolutionary theories about relations between basic emotions and their components are probabilistic rather than deterministic (see Frank, 1988; Oatley & Johnson-Laird, 2011; Scarantino, 2014, 2015, 2017a). Both anger and fear may be associated with the tendency to fight, for instance, but anger more so than fear. The kind of comparative research needed to test these probabilistic hypotheses, however, has not been systematically carried out.

A second solution already discussed in Section 4.2 is that of the abstract reframing of [S–R] links (e.g., Eickers et al., 2017; Lang et al., 1993; Lewin, 1935, cited in Neumann et al., 2003; Scarantino, 2014; Shiota, 2022; Sznycer et al., 2017; see Moors, 2017b; Parkinson, 2017b). Instead of framing the [S–R] links in fear at a concrete level as "predator–flee," for instance, they recast it at an abstract level as "danger–defense," and they consider the concrete tendencies to flee, fight, and freeze as possible manifestations of the abstract tendency to defend oneself (Bolles & Fanselow, 1980; Fanselow, 1994). This strategy does not impose any probabilistic restrictions as the previous solution does. It does not require that anger correlates more strongly with the tendency to fight than does fear, for instance. As discussed in Section 4.2, once the abstract tendency to defend oneself is activated, a goal-directed planning process takes over to determine the concrete action tendency. This may involve learning as well as computation.

Abstract reframing not only allows evolutionary theories to explain variety in behavior but also in somatic responses. If one emotion can be associated with different behaviors, somatic responses – which are supposed to subserve concrete behaviors – should also differ. Indeed, the somatic activity required to freeze should be quite different from that required to flee.

However, the problem with the strategy of abstract reframing is that if the [S–R] link is cast at too abstract a level, it risks becoming trivial. The hypothesis that danger (i.e., an impending lack of safety) leads to the tendency to defend oneself (i.e., the tendency to regain safety) comes down to the hypothesis that the discrepancy between a stimulus and some goal (here, the goal for safety) leads to the tendency to undo this discrepancy. This no longer looks like a special-purpose mechanism but rather a general-purpose one. It is not cut out for a specific type of evolutionary challenge, but for any challenge: Any discrepancy with a goal should be reduced. In addition to being trivial, such a hypothesis is also fairly empty (Moors, 2017b; Russell, 2009; see also Meehl, 1990). If an impending lack of safety indeed leads to the tendency to regain safety, the question remains how this abstract action tendency gets translated into the concrete tendency to flee, fight, or freeze.

A third solution, which is in a sense opposite to the formation of emotion families, is the strategy of breaking these families down into smaller families based on specific stimulus features (e.g., Kreibig, 2010; Shiota, 2022). Take the earlier mentioned finding that patterns of ANS activity differed between contamination-related disgust and mutilation-related disgust (Kreibig, 2010). If the common basis of different types of disgust starts to crumble, there is always the solution of dividing the disgust family into two new families that are no longer required to have anything in common. Shiota et al. (2011) applied a similar strategy to split the joy family into joy from winning a lottery game, joy from watching cute baby animals, and joy from watching grand nature scenes (i.e., awe).

(3) A third source of variety stems from emotion regulation. Emotion regulation may actively hinder the default automatic translation of the affect program to the output components. If the affect program for anger is activated, for instance, the scowling face and antagonistic behavior that are supposed to flow from it may be suppressed. This may produce a lack of distinctness and large intercultural differences. Intercultural differences are expected to be especially pronounced here because different cultures have different display and feeling rules. Moreover, based on the idea that emotion regulation will be most effective in the coarse action channel, less in expressive channels, and even less (and only in indirect ways) in the somatic and feeling channels (Ekman, 1992a), this may produce strong dissociations among the components of a single emotion (see Dan-Glauser & Gross, 2013, for an empirical test of the idea that emotion regulation reduces concordance).

Evolutionary theories assume that emotion regulation can be added to emotions but do not wipe out or rewrite the emotion itself (Ekman & Cordaro, 2011). This creates openings for empirical research. If the emotion is still intact below the surface, measures could be taken to bring it out. Evolutionary theories tend to see emotion regulation as a non-automatic process, meaning that it requires ample operating conditions such as abundant attention, abundant time, and the intention to engage in it. Reducing these conditions should therefore reduce the influence of emotion regulation. Participants could be put under mental load, their micro-expressions could be registered, and their need for emotion

https://doi.org/10.1017/9781107588882.008 Published online by Cambridge University Press

regulation could be minimized, for instance, by making sure there is no audience (Ekman, 1972; Ekman & Cordaro, 2011; see also Chapter 9). Under such conditions, evolutionary theories should predict the degree of variety to drop.

It could be argued, however, that removing the influence of emotion regulation is impossible especially for emotions that are strongly condemned by society because even if no real audience is present, an internalized audience (Fridlund, 1994) or an internalized moral norm (James, 1890b; McDougall, 1908) may cause regulation attempts. A related complication is that emotion regulation can become more automatic with practice (see Box 7.1), a possibility that evolutionary theorists do not deny (e.g., Ekman, 1972). Yet even if the influence of regulation cannot be totally removed, evolutionary theories should still predict that lower levels of regulation are associated with lower levels of variety.

(4) Emotions not only suffer competition from emotion regulation but also from other non-emotional processes. This constitutes a fourth source of variety. All response systems that subserve emotions also have non-emotional functions. The face is also used for non-emotional communication, for instance, and ANS activity serves homeostasis and non-emotional behavior (Levenson, 2014). Research can go some way in controlling for the influence of these other processes (e.g., asking participants not to talk and controlling their attention), but for some components, such as the somatic component, it may turn out to be quite a challenge to detect the emotional "signal" from the non-emotional "noise" (Levenson, 2014).

(5) A final source of variety is not inherent in the theory but stems from methodological limitations, such as the use of (a) weak methods for emotion induction and (b) insensitive methods for the measurement of components. Stimuli used in experimental research are often weak due to ethical constraints (Levenson, 2014). They are often also impoverished because they only pass through one sensory channel (visual, auditory) and they are static (e.g., pictures of facial expressions). Weak and impoverished stimuli are more likely to produce weak emotions with less pronounced components. A weak emotion is also less likely to show concordance, presumably because some of its components can be more easily regulated and/or are more susceptible to the influence of other non-emotional processes (Bradley & Lang, 2000; Roseman, 2011, 2013; Tassinary & Cacioppo, 1992; but see Durán et al., 2017).

If weak emotion induction methods are combined with insensitive methods for measurement, things may get worse. Insensitivity can be situated at the level of the parameters measured (e.g., narrow sampling in the spatial and/or temporal sense) but also at the level of the research design (e.g., between-subjects designs instead of within-subjects designs for establishing concordance) (Levenson, 2014; Shiota, 2022).

The problem of weak and impoverished stimuli can be addressed by conducting observation studies in natural environments, even if they come at the cost of reduced experimental control. The problem of impoverished stimuli can also be addressed in the lab by presenting richer stimulus material. For instance, while older expression research has mainly relied on static pictures of facial expressions, more recent research has shifted to the use of dynamic film clips (Keltner et al., 2016; Krumhuber et al., 2013).

In addition, evolutionary researchers are hopeful that more sensitive methods for measurement and data analysis will bring solace for establishing distinctness and concordance (Tracy & Randles, 2011). Three examples may illustrate this. First, future research may broaden the range of parameters used to measure components. Levenson (2014), for instance, listed a number of underexplored parameters for somatic responses, some of which are visible (e.g., blushing, piloerection, crying) and others invisible (e.g., gastro-intestinal parameters). Second, when it became clear that there were no dedicated brain areas or even circuits for specific emotions (e.g., amygdala or amygdala-mediated circuit for fear) (Lindquist et al., 2012; Wager et al., 2015), evolutionary researchers started to search for and claimed to have found distinct patterns of brain activity (as measured with fMRI) using multivariate pattern recognition techniques (e.g., Kassam et al., 2013; Kragel & LaBar, 2014, 2015, 2016; Nummenmaa & Saarimäki, 2019; Saarimäki et al., 2016, 2018, 2020). This development parallels a general paradigm shift in brain research, moving from a modular view of the brain in which functions are localizable in brain regions to a distributed view in which functions map onto distributed patterns of whole-brain activity and in which deficits in some areas can be compensated for by other areas (i.e., degeneracy). However, as recently argued by Clark-Polner et al. (2017; see also Barrett & Satpute, 2019; Kragel et al., 2018, p. 268), the distinct patterns obtained with these statistical methods do not reflect localizable patterns of activation since they represent averages. Even if localizability were abandoned as the hallmark of an affect program (Scarantino, 2012a; Shiota, 2022), the challenge would still be to show that the patterns replicate across studies (Kragel et al., 2018), which is not currently the case (Barrett & Satpute, 2019). A similar trend to employ pattern-matching techniques can be noted in the search for distinct patterns of somatic responses (e.g., Christie & Friedman, 2004; Kolodyazhniy et al., 2011; Kragel & LaBar, 2015; Kreibig et al., 2007) and facial expressions (e.g., Li & Deng, 2020; Shan et al., 2009). Here again, consistency across studies does not seem within reach. Third, some evolutionary researchers (Levenson, 2014;

Shiota, in press) have argued that concordance should not be studied using between-subjects designs, as is now often done, but rather using within-subjects designs (as in Mauss et al., 2005; Sze et al., 2010). I think this may be a start, but in view of universality, evolutionary researchers should ultimately have the ambition to find similar patterns in all people.

Strategy 2. The second strategy of evolutionary theorists has been to refine the if-then premises set out at the start. The argument runs as follows: The premise "if an emotion is hardwired, then it is likely to have distinct outputs" is not the same as the premise "if an emotion is hardwired, then it has to have a distinct physiological response pattern, a distinct facial expression, and a distinct coarse behavior." Some basic emotions may not have a distinct coarse behavior, and therefore lack a distinct somatic response pattern to subserve it (Ekman & Cordaro, 2011; Shiota, in press). Some basic emotions may lack a distinct facial expression but have a distinct expression in another channel such as the voice, gestures, or even touch. There could also be emotions that have a distinct pattern of values scattered across expressive channels (e.g., part of the face, part of the voice) or even across components (e.g., part of the face, and part of the somatic responses). This is the path taken in recent research (see reviews by Keltner, Sauter, et al., 2019; Keltner et al., 2016; Keltner, Tracy, et al., 2019). Thus, while older research has strongly focused on evidence for unique physiological response patterns and unique facial expressions, recent research expands its scope to (a) expressions in other modalities such as the voice (Cordaro et al., 2016; Sauter et al., 2010), gestures (de Gelder, 2006; Lhommet & Marsella, 2014), and touch (Bonnie & de Waal, 2004; Hertenstein et al., 2009), (b) patterns of expressions across modalities (e.g., emotions with the same facial expression but a different vocal expression count as distinct), and (c) patterns across components (e.g., emotions with the same expression but a different physiological response pattern count as distinct).

Equipped with these multimodal, dynamic expression signatures for basic emotions, evolutionary theorists have recently been able to expand their lists of basic emotions beyond the original six to up to twenty basic emotions including awe, gratitude, pride, shame, guilt, embarrassment, love, sympathy, and so forth (Keltner, Sauter, et al., 2019). Whereas awe, gratitude, pride, and love were initially not considered to be basic emotions because they lack a unique facial expression, awe is now considered basic because it has a characteristic vocal mark, gratitude is reliably expressed in touch, pride combines an expanded chest and tilted head, and love combines smiling, head tilting, open hand gestures, and oxytocin release (Cordaro et al., 2016; Gonzaga et al., 2006; Tracy et al., 2010).

Strategy 3. A third strategy taken by some contemporary evolutionary theorists (e.g., Shiota, 2022) has been to water down the assumption that

https://doi.org/10.1017/9781107588882.008 Published online by Cambridge University Press

basic emotions require the existence of localizable or other signatures in the brain and to shift focus to the assumption that basic emotions are evolved solutions to recurrent challenges. This boils down to the suggestion that basic emotions are still characterized by innate [S–R] links (e.g., [danger–defense]) on the mental super-level without making any strong commitments about how these links are implemented on the brain super-level.

The emphasis on emotions as evolved functions, however, brings to the surface another pressing question relevant for evolutionary theories of all stripes: What makes an evolved function (or innate [S–R] link) qualify as an emotion? It may be recalled that the same question was raised in my discussion of the instincts listed by James (1890b) and McDougall (1908). While James (1890b) did not require an answer since he was a skeptic, McDougall (1908) solved the problem via stipulation. He a priori categorized some instincts as emotional (e.g., flight, repulsion, curiosity) and others as non-emotional (e.g., reproduction, food seeking, crawling).

To sum up, evolutionary theories take the *presence* of direct and indirect evidence as strong indications *in favor of* the existence of affect programs and they make three moves that allow them to avoid taking the *absence* of direct and indirect evidence as evidence *against* the existence of affect programs: They point to (a) various sources of variety that they either predict or acknowledge, (b) overly strict if-then premises guiding indirect research programs, and (c) an undue fixation on demonstrating neural encoding of basic emotions in direct research programs.

4.4.2.2 Criticism 2: Interpretation of Data

Critics have not only taken issue with the robustness or consistency of the findings brought forward by evolutionary researchers, but also with the interpretation of these findings (should they become more robust in the future) as evidence for dedicated affect programs. Even if findings were to replicate across experiments, it may still be questioned whether the effects are caused by affect programs, that is, processes that are specific to emotions and that are innate. To conclude that the effects are caused by affect programs, two types of alternative explanations must therefore be ruled out: that the effects are (a) not specific to emotions but to other things (i.e., Alternative explanation 1) and (b) not driven by innate processes (i.e., primary processes) but by learning or computation (i.e., non-primary processes) (i.e., Alternative explanation 2). I discuss both types of alternative explanations in turn, followed by a reply from evolutionary theorists.

https://doi.org/10.1017/9781107588882.008 Published online by Cambridge University Press

ALTERNATIVE EXPLANATION 1: EFFECTS NOT SPECIFIC TO EMOTIONS
Both direct and indirect effects may not be specific to entire emotions, but rather to (a) molar or molecular components of emotions and (b) non-emotional entities (Fridlund, 2017; Moors, 2017a; Ortony & Turner, 1990). I discuss this issue separately for indirect and direct evidence.

Indirect evidence. Let us first consider research on *distinctness*. To establish whether an emotion is characterized by a distinct component, researchers should manipulate or measure the emotion, they should measure the component, and they should examine whether there is a unique relation between the emotion and the component. The problem is that there is no other way to manipulate or measure an emotion than via one or more of its components (Moors, 2017a). To illustrate, Ekman et al. (1983; Levenson, 1992) investigated whether there are distinct physiological response patterns for different emotions by manipulating emotions via facial expressions (i.e., motor component) and by measuring physiological responses (i.e., somatic component). Rosenberg and Ekman (1994) investigated whether there are distinct facial expressions for different emotions by manipulating stimuli that they expected to be processed in a certain way (i.e., cognitive component) and by measuring facial expressions (i.e., motor component).[51] Clearly, proof of the relation between two components is not proof of the relation between a basic emotion and a component. In other words, presumed evidence of distinctness is in fact evidence of *concordance*. It could be argued, however, that evidence of concordance between two components is a first step towards finding evidence of concordance among all the components of a basic emotion. If research has no choice but to operate at the level of components, this may not be fatal because the components can eventually be pieced together to form the emotion. Perhaps, if all the presumed components of a candidate basic emotion show strong concordance, evolutionary theories have what they need, and the emotion is just a label to summarize the package.

But the problem runs deeper still. The complaint is not just that research is doomed to operate at the component level, but rather that the most robust relations may not obtain among full components (i.e., molar values of components) but rather among subcomponents (i.e., molecular values of components; Ellsworth, 1991; Ortony & Turner, 1990). For instance, the correlations between the molar values of an appraisal of offense, a scowling face, fighting behavior, and angry

[51] Many studies do not even examine the relation between two components but rather the relation between a component and an emotion label. In facial perception studies, participants match facial expressions with emotion labels; in instructed production studies, they produce facial expressions based on emotion labels.

https://doi.org/10.1017/9781107588882.008 Published online by Cambridge University Press

feelings may be less robust than the correlations between the molecular values of an appraisal of goal incongruence, a furrowed brow, and frustrated feelings (C. A. Smith, 1989). A number of findings support this alternative interpretation. For instance, there is evidence that certain somatic responses are not specific to fear and anger, but rather to appraisals of threat and challenge (Blascovich & Mendes, 2000; Quigley et al., 2002; Tomaka et al., 1993, 1997), and to positive and negative affect (Cacioppo et al., 2000; Lang et al., 1993). What is important to realize here is that most subcomponents are not unique to one particular emotion, and that they are not even unique to emotions in distinction to non-emotional phenomena. For instance, an appraisal of goal incongruence is part of all negative emotions (fear, anger, sadness) and even of mundane goal pursuit (e.g., scratching an itch or wiping hair away from one's face). Likewise, a furrowed brow may indicate anger, but also non-emotional effort (e.g., when lifting a heavy weight; Barrett et al., 2009; Ortony & Turner, 1990).

Direct evidence. The problem raised for indirect evidence – that there is no independent, agreed-upon index of emotion – also presents itself for research trying to establish direct evidence for the existence of affect programs. As a consequence, similar issues about emotion-specificity can be raised. In early neuroimaging research, for instance, several studies showed that amygdala activity is not specific to fear (see Barrett et al., 2009; Barrett, Lindquist, Bliss-Moreau, et al., 2007; Cunningham & Brosch, 2012), but rather to appraisals of goal relevance (Mohanty et al., 2008; Ousdal et al., 2008; Sander et al., 2003), novelty (Blackford et al., 2010; C. E. Schwartz et al., 2003; C. I. Wright et al., 2008), or uncertainty (Herry et al., 2007; Rosen & Donley, 2006; Whalen, 2007). These appraisals count as subcomponents of emotions, and they need not even be specific to emotions (Ousdal et al., 2008). In addition, it has been argued that evidence of neural circuits for emotions can better be interpreted as evidence of neural circuits for action tendencies or behavior (e.g., Barrett, 2011; Barrett, Lindquist, Bliss-Moreau, et al., 2007; Barrett & Satpute, 2019) or for goals that do not necessarily align with emotions (LeDoux, 2012b; LeDoux & Daw, 2018). Examples are Adams's (1982) circuits for defense, offense, and submission, Fanselow's (1994) circuits for freezing, fleeing, and fighting, LeDoux's (2012b) circuits for defense, energy maintenance, fluid balance, thermoregulation, and reproduction, and even Panksepp's (2012) circuits for FEAR, RAGE, PANIC/GRIEF, LUST, CARE, PLAY, and SEEKING, which are ultimately based on animal behavior. Why insist that play, lust, care, and seeking are emotions rather than behaviors?

The same issue is at play in more recent neuroimaging research aimed at discovering affect programs in the form of patterns of whole-brain activity (e.g., Nummenmaa & Saarimäki, 2019). More controlled designs are needed to show that the patterns are specific to entire emotions, and not to emotion components or non-emotional entities (see Barrett & Satpute, 2019; Kragel et al., 2018). In many brain studies, participants are asked to watch pictures with emotional content or to relive past emotional episodes. There are no guarantees that these emotion-induction procedures elicit emotions rather than cold perceptions or thoughts, for instance.

Alternative Explanation 2: Effects Driven by Non-Primary Processes

Even if future research were to show consistent and emotion-specific evidence for distinctness, concordance, and universality, there would still be an additional caveat. There would be no reason to accept that the emotion-specific effects are caused by affect programs, understood as innate entities. Instead of primary processes, the effects could also be due to non-primary processes of emotion causation, as other theories of emotion causation would have it (e.g., non-biological versions of network theories and stimulus evaluation theories; see Chapters 5 and 6). Another option is that the effects are due to non-primary processes that are not emotional. Again, I discuss this issue for indirect and direct evidence separately.

Indirect evidence. It has been pointed out that when evolutionary researchers accept indirect evidence as valid, they fall prey to the fallacy known as "affirming the consequent." This fallacy entails that based on the premises "if p then q" (e.g., if the lamp is broken then it is dark) and "q" (e.g., it is dark), one accepts the conclusion "p" (e.g., the lamp is broken). The conclusion is false because there may be other explanations for "q" (e.g., the lamp was not switched on). Applied to our case: Based on the premises "if an emotion has an affect program, then it is likely that there will be distinctness/concordance/universality" and "there is distinctness/concordance/universality," evolutionary theorists accept the conclusion "the emotion has an affect program." Here as well, the conclusion is false because there may be alternative explanations for distinctness/concordance/universality.

Consider research on *distinctness*. Suppose this research were to confirm the intuition that fleeing is more likely in the case of fear/danger, fighting is more likely in the case of anger/offense, and ejection is more likely in the case of disgust/disgustingness (Scarantino, 2017a). This would not yet show that the links between danger–flight, offense–fight, and disgustingness–ejection are innate [S–R] links. Distinct action tendencies could be selected because they are instrumental. Individuals could compute, for instance, that fleeing is the best response to danger, fighting

https://doi.org/10.1017/9781107588882.008 Published online by Cambridge University Press

the best response to offense, and ejecting matter the best response to a risk for contamination. This could result in learned [S:R–O^v] links (i.e., goal-directed knowledge) or learned [S–R] links (i.e., habits). Once the right action tendency ([R]) is selected, there are again two possibilities. A first possibility is that the translation of the selected action tendency into a concrete motor program rests on an innate [R–r] link (corresponding to the second interpretation of Darwin, 1872, see Chapter 3). For instance, the individual may need to learn that ejection is the best response to remove potentially contaminated matter from the oral cavity, while the translation from the tendency to eject into a concrete motor program with specific muscle contractions is built in (Allport, 1924). A second possibility is that the translation into a concrete motor program is also computed or learned during the individual's life, either in the form of an [S:r–R] relation (i.e., goal-directed knowledge) or a learned [R–r] link (i.e., habit). For instance, an individual may not only need to learn that in the case of danger it is best to increase information intake, but also that raising one's eyebrows is the best way to do this as it allows more light to fall on the retina (see discussion in Ekman, 1992a).

To the extent that learning of the best responses takes the same form in all members of the species across cultures – a phenomenon called "species-constant learning" (Allport, 1924) or "convergent cultural evolution" (Fridlund, 1994) – it can also serve as an alternative explanation for *universality*. Learning processes also provide an alternative explanation for consistency in facial expressions between congenitally blind and sighted people as it is likely that blind people are similarly rewarded for producing culturally appropriate facial expressions (Barrett, 2011). Note that while evolutionary theorists invoke learning to explain a potential *lack* of consistency across groups, critics invoke learning to explain a potential *presence* of consistency across groups. Also note that some authors (e.g., Crivelli, Jarillo, & Fridlund, 2016; Fridlund, 2017) have called into question the premise "if the emotion has an affect program, then universality is likely," pointing out that natural selection may also produce diversity instead of uniformity among cultures.

Finally, *concordance* among the components of a single emotion can alternatively be explained as the result of a common cause that is not an innate affect program, but that is based on learning or computation. For instance, network theories propose that concordance stems from the activation of a learned emotion network (see Chapter 5) and appraisal theories propose that it stems from an appraisal process that produces a specific appraisal profile (see Chapter 6).

In addition to these alternative processes for emotion causation, concordance can also stem from a non-emotional common cause. To borrow an example from Russell (2009), suppose focused attention would cause both a frown and cardiac acceleration. This would result in a correlation between frowning and cardiac acceleration, but it would not count as evidence for concordance among the components of anger.

What is more, concordance can even be explained without invoking a common cause. First, correlations among components may stem from correlations among features of the stimuli that cause components. To again borrow an example from Russell (2009, p. 1273), suppose a correlation exists between novel and goal-incongruent stimuli, and that novelty produces a frown whereas goal-incongruence produces cardiac acceleration. This would result in a correlation between frowning and cardiac acceleration in the absence of a common cause. Second, correlations among components may also stem from direct influences among the components themselves (Colombetti, 2009; J. Lange et al., 2020; J. Lange & Zickfeld, 2021; Russell, 2009). To illustrate, suppose producing a threat face increases breathing and muscle tension, which in turn leads to cardiac acceleration. Again, this would result in a correlation between the threat face and cardiac acceleration in the absence of a common cause. The components in an emotion may even interact in such a way that they form a stable pattern. As will be explained in more detail in Chapter 5, the distinction between theories that explain concordance by invoking a common cause and theories that explain concordance by the dynamic interactions among components aligns with the distinction between psychometric models known as latent variable models and psychometric network models (Coan, 2010; J. Lange et al., 2020; J. Lange & Zickfeld, 2021).

Direct evidence. Suppose that future neuroimaging research were to deliver robust and distinct patterns of brain activity for basic emotions. This would not yet show that the patterns are innate (Kragel et al., 2018). Consistency across samples from different remote cultures, suggesting universality, would certainly help, but as discussed above, it would still not escape the alternative explanation of species-constant learning.

REPLIES

In reply to the suggested alternative explanations for direct and indirect evidence, evolutionary theorists have justified their conclusions as "inferences to the best explanation" – accepting the explanation that fits best with or is most plausible given the available evidence (Douven, 2017; Thagard, 1978). For instance, even if the consistency in facial expressions between congenitally blind and sighted individuals can be explained via social learning, Shariff and Tracy (2011a) believe that innateness rings truer. Ekman (1992b) argued at the time that the existing evidence for the alternative explanations (e.g., in terms of subcomponents; Ortony & Turner, 1990) was weak and that it had to be held to the same standards that evolutionary research has been held to. Meanwhile, such research has proliferated and makes it harder to stick to the powerful intuition that there must be emotions in the brain. Taken together, given the ongoing debate about what "the available evidence" is today and the fact that this evidence may continue to evolve in the future, judgments about the plausibility of evolutionary explanations are necessarily colored by whether one is a vindicator or a skeptic, and this ultimately boils down to intuition.

To summarize, evolutionary theories insist that a limited but still open-ended set of basic emotions is neurogenetically encoded in so-called affect programs. Each basic emotion has its own unique affect program, which is triggered by a unique set of stimuli and results in a unique set of components (distinctness) that correlate amongst each other (concordance) and generalize across groups (universality). Although evidence for the existence of affect programs is not convincing to date, proponents of evolutionary theories do not seem willing to bow out, even if some of them seem willing to water down the theory so much that there is hardly any liquor left. Nevertheless, alternative explanations loom large. These alternative explanations point either at non-emotional entities or at non-primary processes such as learning and computation. The role of learning is put center stage in network theories, to be discussed in the next chapter (Chapter 5). The role of computation is more prominent in evaluation theories, which are discussed after that (Chapters 6 and 7).

CHAPTER 5

Network Theories

5.1 Precursors

James's (1890b) theory distinguished between innate [S–R] links (i.e., primary processes) and learned [S–R] links or habits (i.e., secondary processes). This set the stage for two theoretical traditions: evolutionary theories, which focus more on the innate stuff, and network theories, which focus more on the learned stuff. McDougall's (1908) theory formed the ideal bridge between James's (1890b) theory and evolutionary theories because he protected the innate process by depicting it as the unchangeable, purely emotional core, which cannot be fundamentally altered by learning. This was echoed by evolutionary theories (e.g., Ekman & Cordaro, 2011). Network theories, on the other hand, mark a return to James (1890b) in the sense that they open up again to a more substantial influence of learning processes. Unlike James (1890b), however, they do not insist that feelings can only be produced peripherally (i.e., only via feedback from the periphery) but they allow them to be also produced centrally (e.g., Leventhal, 1980, p. 168) similar to several evolutionary theories. Network theories come in many shades depending on how they construe the relative contributions of innate and learned processes and hence how much overlap they still have with evolutionary theories (Philippot et al., 2002). Rather than merely accepting the influence of learning, however, network theories have actively worked out the principles involved in emotional learning and they have set up empirical research programs to study these principles.

For this, they have drawn on established learning theories of behavior developed in the behaviorist tradition. Radical behaviorists took as their explanandum "behavior," for which they proposed purely causal explanations in which the explanantia are regularities in the environment (i.e., relations between stimuli). Although these theories were initially developed to explain behavior, several of them expanded their set of explananda to emotions. Because radical behaviorists such as Watson (1919, 1929) and Skinner (1945) eschewed mental constructs (e.g., representations, consciousness, feelings), which they considered to be non-observable entities that do not have a place in science, they reconstituted

https://doi.org/10.1017/9781107588882.009 Published online by Cambridge University Press

emotions as purely observable phenomena and aligned them with behavior (Dashiell, 1928; Skinner, 1945) or patterned responses (i.e., clusters of somatic responses and subtle and coarse motor responses) (Watson, 1919). The approach taken was not much different from that in evolutionary theories: Researchers started from a limited number of vernacular discrete emotions and tried to align them with relations between stimuli and behaviors or patterned responses. For instance, Watson (1919, 1929) postulated the existence of three sets of US–UR links present at birth (either inherited or established by prenatal conditioning): loud noises leading to a fear response, restraint of movement leading to an anger response, and stroking or caressing leading to a love response. Like evolutionary theories, he held that more complex emotions developed by combining multiple US–UR links or by learning principles such as substitution (which allows the range of stimuli that elicit emotional responses to grow) and response integration (which allows the emotional responses to become more complicated; Watson, 1929, p. 24).

In an attempt to capture the component that others called "feelings," radical behaviorists proposed substituting "feelings" by "verbal emotional behavior" (Skinner, 1974, pp. 61–62). Verbal behavior can be either overt (public speech) or covert (private or inner speech). In Watson's (1929, p. 18) words, "[s]peaking overtly or silently is just as objective a type of behavior as baseball." This is because both types of behavior can be observed, even if the silent variant can only be observed by the person engaging in it. The message is that instead of trying to explain feelings (e.g., the feeling of anger), researchers should try to figure out the learning principles governing verbal self-ascriptions of feelings (e.g., "I feel angry"). An example of such a principle is arbitrarily relational responding (Friman et al., 1998). Objects and their labels may end up in the same equivalence class because they share similar functions. This explains why anxious people show anxious responses to so many stimuli that it seems as if they are "afraid of their own shadow" (Friman et al., 1998, p. 145).

In contrast with radical behaviorist theories, neobehaviorist theories went beyond purely causal explanations by also proposing mental mechanistic explanations in which representations and operations mediate between the regularities in the environment and behavior (e.g., Hull, 1943). The content of these representations consisted largely of innate or learned [S–R] links and the operations involved were typically associative.

A final precursor of network theories of emotion is to be found in semantic network models of memory (e.g., Collins & Loftus, 1975; Holland, 1975). The explanandum here is not "behavior" but "cognition" and the aim is to figure out how semantic knowledge is encoded in the mind. This concerns the properties of stored representations (i.e., memories) and the way in which these representations interact with one another.

https://doi.org/10.1017/9781107588882.009 Published online by Cambridge University Press

Early semantic network models are *localist* in that each concept is localized in a separate node. These models received competition from *feature* models in which each concept is made up of features that are distributed across semantic space. A final type of models are *distributional* models. These are more concerned with the way in which concepts are learned and they propose that this happens via the extraction of statistical regularities from the environment. There is debate about whether this extraction follows error-free, purely associative principles (i.e., Hebbian learning) or rather error-driven, predictive processing principles (see Kumar, 2020, for a review).

5.2 Constitutive and Mechanistic Explanations

Network theories of emotion (Bower, 1981, 1991; M. S. Clark & Isen, 1982; Colombetti, 2009; Forgas, 1999, 2017; Isen, 1984; Isen et al., 1978; Lang, 1984, 1985, 1993, 1994; Leventhal, 1980, 1984; M. D. Lewis, 2005; Teasdale, 1999) postulate that emotions are represented in memory in the form of a network made up of nodes that stand for (a) representations of eliciting stimuli, (b) representations of components of emotions such as appraisals, somatic and motor responses, and feelings, and (c) representations of related information (e.g., emotion words, social norms) (see Figure 5.1, which depicts only "a" and "b"). The nodes are linked to one another in bidirectional associations. An emotion is caused when the network gets activated. Once activation in a node passes a certain threshold, activation spreads to all other nodes in the network. In the case of the cognitive and motivational components, which are themselves representations, the nodes merely need to be activated in order for the components to occur. In the case of the somatic and motor components, which are not representations but overt responses, the nodes representing these responses

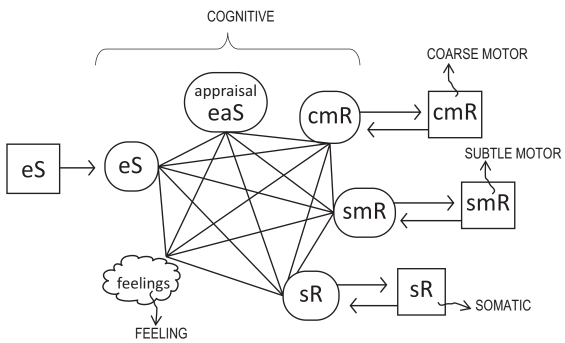


Figure 5.1 Network theories

https://doi.org/10.1017/9781107588882.009 Published online by Cambridge University Press

must be activated strongly enough to spark the occurrence of the components (Bower, 1981).

Emotion networks may be activated via the stimulus side, by raw external stimuli (eS) that show similarity with concrete stimulus representations that are already part of the network ([eS]). The stimulus representation then activates all the other nodes in parallel: an appraisal of the stimulus ([eaS]), a somatic response representation ([sR]), a subtle motor response representation ([smR]), a coarse motor response representation ([cmR], and feelings. For example, the fear network may be activated by coming across a mad dog, which activates the stimulus node for mad dogs and associated nodes representing danger, an adrenaline rush and sweating palms, a fearful expression, fleeing behavior, and feelings of fear.

Emotion networks may also be activated via the response side. This can happen via artificial induction of the motor or somatic responses represented in the network (Lang, 1994; see also Darwin, 1872; Ekman et al., 1983; McDougall, 1908).[52] For instance, the fear network may be activated by putting on a fearful face or by taking amphetamines. Artificially induced responses may activate their corresponding nodes via somatosensory feedback. In the case of artificially induced motor responses, moreover, the intention to engage in the artificial response may simply coincide with the corresponding response node.

The only operation required for the activation of an existing network and for the spreading of activation throughout the network is an associative operation. Associative operations are assumed to proceed in an automatic fashion (Cacioppo et al., 1992; Lang, 1994; Leventhal, 1984; Teasdale, 1999). A first threshold is required for the activation to pass from one node to another. A second threshold is required for a node to become conscious (Moors, 2016). Networks of opposing quality inhibit

[52] It may be noted that activation of emotions via the response side was already present in the writings of several forerunners. This is obvious for James (1890b), who only allowed for peripheral causation of feelings, but the idea was also present in evolutionary theories and their precursors. Darwin (1872, p. 31) cited Bain (1864) who wrote that "actions, sensations, and states of feeling, occurring together or in close succession, tend to grow together, or cohere, in such a way that when any one of them is afterwards presented to the mind, the others are apt to be brought up in idea". McDougall (1908, p. 98) proposed that sympathy is not a separate instinct but "founded upon a special adaptation of the receptive side of each of the principal instinctive dispositions, an adaptation that renders each instinct capable of being excited on the perception of the bodily expressions of the excitement of the same instinct in other persons." Ekman et al. (1983) must also assume the existence of such a mechanism given that in order to examine distinct somatic response patterns for basic emotions, they used a method in which emotions were manipulated via the artificial induction of the facial expressions supposedly belonging to these emotions.

each other (e.g., fear and joy) whereas networks of similar quality may blend (e.g., sadness and surprise may blend to disappointment) (Bower, 1981, p. 135).

Most network theories adopt a broad constitutive explanation of emotion by equating the emotion with the entire emotional episode. But the constitutive explanation could also be narrower, limited either to the memory network alone or, alternatively, to overt responses and feelings (**Axis A**; **Axis 5c**). Network theories come in two main versions: a biological version and a non-biological (or no longer biological) version (**Axis B; Axis 6d**). I discuss these in turn.

5.2.1 Biological Version

Biological network theories (Bower, 1981; Forgas, 1999, 2017; Johnson & Multhaup, 1992; Lang, 1984, 1985, 1993, 1994; Leventhal, 1980, 1984; Leventhal & Scherer, 1987; Power & Dalgleish, 1997; Teasdale, 1999; Teasdale & Barnard, 1993) remain close to evolutionary theories. Organisms come into the world with a handful of innate emotion networks, built around innate [S–R] links, with nodes for stimuli, responses, and feelings. Activation of an emotion network produces the preset responses and feelings. The range of stimuli that can trigger an emotion network and the range of responses that can follow from them are progressively expanded via learning principles (Forgas, 1999; Leventhal & Scherer, 1987; Martin & Levey, 1978; Teasdale, 1999, p. 594). In this version, each emotion has a separate network. The innate networks can be expanded by learning, but they naturally remain clustered around their innate cores. Biological network theories differ from evolutionary theories in the weight they assign to learned vs. innate processes and the research programs they breed. Evolutionary theorists emphasize innate processes and search for affect programs whereas network theorists emphasize learning and study how emotions can be learned (e.g., fear learning). But there is more. Evolutionary theorists tend to treat the innate affect programs as the sine qua non of emotion (i.e., the pure emotion or the part of the emotional episode that makes it truly emotional), whereas network theorists take the entire network including all the learned information as the sine qua non of emotion. Some network theorists even go so far as to claim that *only* the learned stuff gives rise to full-blown emotions, arguing that the innate stuff consists of meaningless reflexes (e.g., loud noise leading to a startle response) (Leventhal, 1980).

If the retrieval of emotion knowledge is indeed the cause of emotions, as network theories claim (Bower, 1981), the question can be raised how these theories can account for dissociations between the cold retrieval of information without emotion (e.g., thinking about fear) and the hot retrieval with emotion (e.g., fear) (Teasdale, 1999). In other words, it seems that the mere retrieval of emotional information is not sufficient for emotion. A related dissociation, often observed in therapy, are recalcitrant emotions. These are emotions that persist despite convincing evidence that there are no grounds for them, such as the case of a patient who continues to feel guilty after the therapist has convinced her that she is not to blame (Teasdale, 1999, p. 666). To accommodate these dissociations, network theories propose that emotion networks include different layers of information processing and that only some of these layers can directly trigger emotional components (Lang, 1994; Leventhal, 1980; Leventhal & Scherer, 1987; Power & Dalgleish, 1997; Teasdale & Barnard, 1993).

Different network theories have cut the cake in different ways, however. Lang (1994) proposed three layers based on the content or type of information involved in the nodes: stimulus nodes, response nodes, and nodes with semantic knowledge. He asserted that a full-blown emotion will only occur when response nodes are activated. Semantic knowledge nodes can only serve as an entry point for emotion elicitation if the activation is spread further to the response nodes. Leventhal (1984; Leventhal & Scherer, 1987) postulated three layers: the sensorimotor layer, the schematic layer, and the conceptual layer. Different layers involve different types of operations that operate on different formats of representations or codes. The sensorimotor layer refers to innate expressive-motor programs that connect specific sensory codes to reflex-like responses (i.e., innate [S–R] links; primary processes). The schematic layer refers to associative operations that take care of encoding and retrieval (i.e., secondary processes) and operate on perceptual (image-like) codes. The conceptual layer refers to computations (i.e., tertiary processes) involved in reasoning and decision making that operate on conceptual (verbal-like) codes (abstracted from emotional episodes). Processing at the sensorimotor and schematic layers is assumed to take place in an automatic sense; processing at the conceptual layer is typically marked as non-automatic. In this theory, it is the schematic layer that holds the strongest emotion-eliciting power. The sensorimotor layer is supposed to produce reflexes, which may not qualify as full-blown emotions, and the conceptual layer is only allowed to elicit emotions indirectly by "calling up" the schematic layer. In a similar vein, Teasdale (1999) framed perceptual codes as memories *of* emotions (i.e., "gut" memories) and conceptual codes as memories *about* emotions (i.e., intellectual memories). It remains obscure, however, why the schematic layer would be more likely to elicit emotions than the conceptual layer (Frijda & Zeelenberg, 2001, p. 146). Leventhal (1984) suggested that perceptual codes are concrete and therefore more vivid whereas conceptual codes

https://doi.org/10.1017/9781107588882.009 Published online by Cambridge University Press

are abstractions that leave out familiar cues and context information, which are crucial for the actual provocation of emotions.[53]

The multi-layered nature of network theories is not only capable of explaining what is required to turn a cold memory into a hot one, but also why a hot memory cannot easily be changed by providing cold information as is the case in recalcitrant emotions. Here, authors hint at the automaticity and relative cognitive impenetrability (i.e., modularity) of processing at the schematic layer (e.g., Leventhal, 1980; Teasdale, 1999). If a person's fear of spiders is rapidly elicited and relatively encapsulated, any cognitive work that remains strictly at the conceptual layer will be too slow and powerless.[54]

5.2.2 Non-Biological Version

Non-biological network theories are not committed to the evolutionary origin of emotion networks (Camras, 1992; Colombetti, 2009; J. Lange et al., 2020; M. D. Lewis, 2005; M. D. Lewis & Liu, 2011). Some theories that fit here nevertheless allow for the existence of packages of components that correspond to those of basic emotions in evolutionary theories. The concordance between the components of each of these emotions is not the result of an affect program, however, but the result of the mutual interactions among the components themselves.

The distinction between evolutionary theories, on the one hand, and non-biological network theories, on the other hand, has been aligned with the distinction between two psychometric models known as latent variable models and psychometric network models (J. Lange et al., 2020). In latent variable models, the output components of an emotion are understood as symptoms of a common cause: the affect program. In psychometric network models, on the other hand, emotions are understood as

[53] Note that Leventhal's (1984) theory is more complex than the picture drawn so far. He discussed several additional antecedents of emotional experience (i.e., the feeling component). Thus, he mentioned unlearned stimuli that violate expectations raised at the schematic layer, such as the sight of a decapitated head. He also mentioned complex interactions between the schematic and conceptual layers. In the "laughter in church" phenomenon, the discrepancy between the automatic tendency to laugh at a silly joke (generated at the schematic layer) and the voluntary tendency to remain serious (generated by the conceptual layer) ironically boosts the tendency to laugh. Conversely, the phenomenon that you cannot tickle yourself is explained by the lack of a discrepancy between the action effects predicted at the conceptual layer and the action effects occurring at the schematic layer. Note that several of these ideas come back in more recent predictive processing theories of affect (see Van de Cruys, 2017).

[54] In Leventhal's (1984) theory, moreover, recalcitrance may even add to the intensity of emotional experience, as is suggested by the laughter in church phenomenon.

syndromes that emerge from the mutual interactions among the components instead of by an underlying common cause. Latent variable models can easily explain concordance (i.e., packages), but in order to explain variety, they have to invoke extra factors that moderate the relation between the common cause and the components. Psychometric network models only deal with components that can interact in various ways and are therefore naturally equipped to account for variety. At the same time, the interactions between the components can also settle on stable configurations, which can account for concordance.

Another source of inspiration for non-biological network theories is the dynamic systems (DS) framework (Colombetti, 2009; M. D. Lewis, 2005; M. D. Lewis & Liu, 2011). This is a general framework of information processing that developed in contrast to the classic computational framework. In the classic framework, information processing often takes the form of units of linear, sequential input-output relations. The theories discussed in the previous chapters took this approach: An incoming stimulus, after being processed, activates an action tendency, which in turn translates into overt responses and feelings. The DS framework describes the activity of a system as it unfolds over time (Kelso, 1995; Port & Van Gelder, 1995). The activity of the system as a whole emerges from the interactions among its parts, resulting in the self-organization of the system. The interactions are not linear but involve positive and negative feedback loops that can (a) self-amplify or (b) oscillate until they again stabilize the activity of the whole system (M. D. Lewis, 2005). Self-amplification, based on positive feedback loops, can be illustrated by the case in which an appraisal of offense leads to anger, and the anger further intensifies the appraisal of offense, and so on, leading to an escalation of anger over time. Oscillation between positive and negative feedback cycles can be illustrated by the case in which an appraisal of uncertainty leads to an increase in vigilance, which in turn leads to a decrease in uncertainty, after which the vigilance is reduced in turn, followed again by an increase in uncertainty, and so on.

The activity of the system as a whole can be described by placing it in a multidimensional space in which the dimensions reflect possible values of its parts. The activity of the system over time traces a trajectory, with phases of relative instability, and phases to which the system tends to gravitate, so-called attractor states. The trajectory of the system may jump from one attractor state to the next because of small changes in the activity of one or more of the parts. Some of these changes may be caused by external stimuli that impinge on the system and that temporarily destabilize it before the system self-stabilizes again by settling into one of the possible attractor states. In addition to the bottom-up influence or constraints from parts to the whole, there is also a top-down influence or constraint from the whole to the parts. These bidirectional inter-level influences are referred to as circular causation (but see Bakker, 2005). Applied to emotions, the idea is that an emotion is shaped by the reciprocal interactions over time among its components. This provides the glue among the components, so that there is no need to invoke a common cause (such as an affect program) to explain concordance. Packages produced in this way count as soft packages because variety in the context is allowed to produce variety in the shape of the packages (Colombetti, 2009).

A number of issues are worth noting. First, accepting the existence of packages is still a far cry from accepting the existence of the blue-ribbon packages put forward in evolutionary emotion theories. As most of the components that are said to figure in emotional episodes are ongoing processes (cognition, motivation, somatic and motor responses, and feelings), there will be an ongoing stream of patterns among these values (Russell, 2009). Some of these patterns may stand out because they mark an unusual quantitative change in all components within some brief time interval (a big change in cognition, motivation, somatic and motor responses, and feelings). This is called coherence. But coherence does not speak to the quality of these changes. There could be an infinite number of qualitatively different coherent packages. This is the point where the DS framework might invoke (the specific nature of) the emotion as a whole to provide top-down constraints, the idea that the activity of the system can fall in a limited number of attractor states. The problem with this proposal is that there is nothing in the DS framework per se that warrants a limited number of attractor states, nor that if these attractor states are indeed few, they will match with the blue-ribbon packages that evolutionary emotion theories put forward (see J. Lange et al., 2020). Colombetti (2009) suggested that the fact that our language has words for the blue-ribbon packages indicates that these packages must exist prior to our labeling. This argument is not convincing, however. It is true that words are sometimes invented to label previously existing entities such as animals and plants. But words have also been invented to label constellations of stars, and these did not form entities prior to our labeling (Russell, 2009, p. 1276).

It could be argued that if the biological core, the affect program, is left out (or massively overwritten by learning soon after birth), networks may not (or no longer) be organized according to the discrete emotions postulated by evolutionary theories, but rather in culturally salient ways or even in highly idiosyncratic ways (see James, 1890b; M. D. Lewis, 2005; M. D. Lewis & Liu, 2011). M. D. Lewis (2005) proposed a network theory embedded in the DS framework in which the role of biology is kept minimal based on the assumption that evolution does not endow us with affect programs that contain ready-made instructions for how to behave, but rather dispositions for behavior that require substantial maturation after birth. M. D. Lewis and Liu (2011) further argued that under the impulse of DS principles, maturation involves a shift from more distributed brain activity (destabilization) to more specialized brain activity (stabilization) with the help of pruning. Crucially, the precise shape of this specialization may differ across individuals depending on their learning history.

Embracing the dominant role of learning processes entails embracing variety, relying on the consensus that environments of individuals across and within cultures are not identical. But the question is how much variety there is between cultures and individuals. Next to species-variable learning (individuals within the species learn different things), there could also be species-constant learning (individuals within the species learn similar things) (Allport, 1924; Fridlund, 1994). Thus, even if emotions were entirely learned, it remains an empirical question just how much variety vs. consistency this would entail. If all cultures and all individuals within each culture are faced with the same challenges and opportunities, and if these are more or less the same as those faced by our ancestors, non-biological network theories might discover the same empirical patterns as evolutionary theories but simply explain them differently.

As for the DS framework itself, it is still up for debate whether this framework offers a radical alternative to theories bred in the classic tradition or whether it is instead complementary to them. The DS framework draws attention to reciprocal interactions among parts that become apparent only if the time window is stretched. Based on the assumption that these interactions "are so dense that it is not possible to track the way in which each individual part contributes to the system's overall behavior" (Colombetti, 2009, p. 414), theorists may be led to vote for the DS approach as a radical alternative. Yet several arguments speak in favor of complementarity (see also Delancey, 2005; J. J. Prinz & Barsalou, 2000). It could be argued that the macroscopic perspective taken in DS theories is not in itself incompatible with the microscopic perspective of classic theories. The idea that the interactions between components are dense need not deter researchers from parsing the stream of components into units and analyzing the mechanisms operative in a single unit (see also D. Evans, 2008; Moors & Boddez, 2017). In fact, proponents of several other theories have flirted with the tools of the DS framework. Examples can be found in evolutionary theories (Izard et al., 2000, p. 17), biological network theories (e.g., Teasdale, 1999; Teasdale & Barnard, 1993), appraisal theories (Scherer, 2000, 2009a), and psychological constructionist theories (Barrett, 2006b; Barrett, Ochnser, et al., 2007). In sum, a DS

https://doi.org/10.1017/9781107588882.009 Published online by Cambridge University Press

framework opens our eyes to evolutions over time that tend to be overlooked, but it need not be placed in opposition to, or proposed as a substitute for, the micro-mechanistic proposals in classic theories.[55]

5.3 Scientific Definitions

5.3.1 Intensional Definition

5.3.1.1 Criteria for Demarcation

The constitutive and mechanistic explanations proposed in network theories yield the following necessary criteria for the intensional definition of emotion. An episode counts as an emotional episode if its components are caused by the activation of an emotion network in memory. Let us call this the "network" criterion. In biological network theories at least, not all layers in the network are capable of eliciting emotions, however. In Lang's (1994) theory, emotions are caused by the response layer. In Leventhal's (1984) theory, they are caused by the schematic layer, where nodes with perceptual codes must be activated automatically. Let us call this the "layer" criterion. In these network theories, moreover, emotion networks are still built up around affect programs, which equips them with an additional "affect program" criterion. A final criterion, shared by all network theories, is the "component" criterion, which stipulates the presence of a number of components. This varies depending on the width of the constitutive explanation that is endorsed.

5.3.1.2 Adequacy

The questions to ask in this section are how well the above-listed criteria account for the apparent properties of emotions listed in the working definition (apparent-similarity meta-criterion) and whether they allow demarcating emotions from non-emotional phenomena (fruitfulness meta-criterion).

APPARENT-SIMILARITY

The components in the emotional episode that stand for responses and feelings account for the bodily and phenomenal aspects of emotion, respectively. As an emotion network is made up of representations, it exhibits Intentionality. The stimulus representations ([S]) in the network cater for the world-directedness of emotions. To alleviate the complaint

[55] Another option is to consider DS principles not only at the macro-level but also at the nano-level. The entities at the micro-level can themselves be further decomposed into entities at the nano-level and these may self-organize so that the entities at the micro-level emerge as Gestalts.

https://doi.org/10.1017/9781107588882.009 Published online by Cambridge University Press

that concrete stimulus representations ([ecS]) do not provide the right kind of object (e.g., coiling form of a snake), however, most network theories have added abstract stimulus representations – appraisal nodes – ([eaS]) to their networks (e.g., Bower, 1981; Leventhal & Scherer, 1987). This has turned them into network-evaluation hybrids. The action tendency ([R]) in the network supplies the self-directedness of emotions.

The notion of multiple processing layers was added to make sense of the dissociation between cold (e.g., thinking *about* fear) and hot retrieval (e.g., thinking *with* fear), and in this way to help demarcate emotions from purely intellectual states. The response nodes and/or vivid perceptual nodes that must be activated for the network to qualify as an emotional network can be said to deliver the intensity of emotions. While intensity covers the quantitative part of the heat of emotions, the qualitative part, namely valence, is still missing from the picture. If valence is missing, emotions generated at response or schematic layers cannot be distinguished from non-emotional phenomena such as mundane habits, which also involve response and perceptual nodes. Here again, appraisal can come to the rescue, as it is well-suited to deliver valence. This solution is not ironclad, however, because even in network theories that include an appraisal node, activation of the other nodes in the network can happen without prior activation of the appraisal node. This means that in episodes in which appraisal is in fact bypassed, earlier complaints about the wrong kind of object and missing valence continue to apply.

As explained above (Section 5.2), multiple processing layers also help to understand theoretically irrational emotions known as recalcitrant emotions, as in the case in which a person's emotion of guilt does not match up with her belief that she is not to blame. In Leventhal's (1984) theory, for instance, the schematic layer is responsible for the guilt whereas the conceptual layer is responsible for the belief that she is not to blame. As explained above, the guilt can survive due to the automaticity and relative encapsulation of the processes residing at the schematic layer. This entails that the layer criterion also takes care of the automaticity and control precedence of emotions.

But what about practical rationality? In biological network theories such as Leventhal's (1984), everything starts with affect programs, which are innate [S–R] links, situated on the sensorimotor layer. At the schematic layer, these links get elaborated via principles of generalization and classical conditioning. The resulting connections are learned elaborate [S–R] links, which when activated, cause emotions. Leventhal (1984) explicitly contrasted the stimulus-driven processes from the schematic layer with the goal-directed processes for behavior causation and emotion regulation located at the conceptual layer. Since stimulus-driven processes do not take into account the outcome of behavior, they guarantee that emotions are practically irrational in the process-sense. As always, the behavior resulting from this process can still be adaptive in certain contexts, allowing for instances of emotions that are practically rational in the output-sense. Networks in non-biological network theories resemble those in biological ones except that they do not (or no longer) have their innate core. If learning follows the same principles, however, these networks may also qualify as elaborate learned [S–R] links.

The affect programs at the heart of the networks in biological network theories further account for the continuity between adult emotions and those of infants and animals. Non-biological network theories are less well equipped in this regard. In both versions, however, the strong emphasis on learning processes is especially apt to explain differences in the complexity and sophistication of emotions between these groups.

FRUITFULNESS

To evaluate fruitfulness, the question to address is whether the listed criteria create a homogeneous set in which generalizations can be made from one member to the next and not to members outside of the set. The network and layer criteria specify that emotions are caused by memory networks dedicated to these emotions, and in particular by a specific layer within these networks. Thus, a *network* is emotional by virtue of the "emotional" content of its nodes. For instance, the fear network has nodes for stimuli, appraisals, responses, and feelings that are specific to fear. A *layer* has emotional power if it houses special types of nodes (response nodes according to Lang, 1994; perceptual nodes according to Leventhal, 1984) that are activated by an associative operation. Given that different emotions are caused by different emotion networks, however, generalizations cannot be made across these networks. Thus, the fruitfulness of the overarching set of emotions likely remains low.

5.3.2 Divisio Definition

5.3.2.1 Criteria for Partitioning

Biological network theories endorse a discrete view regarding the variety within the set of emotions. Each biologically basic emotion has a separate network (Bower, 1981; Leventhal, 1980).[56] If the biological sting is pulled out of networks, however, as is the case in non-biological network theories, there could in principle still be other forces that ensure the clustering of components, such as the sheer interactions among components and

[56] Note that Lang (1985, p. 164) does not see the boundaries between discrete emotions as strict.

https://doi.org/10.1017/9781107588882.009 Published online by Cambridge University Press

environmental influences (Colombetti, 2009). Although it remains possible that interactions among components result in the blue-ribbon packages identified by evolutionary theories, there is no guarantee that they do so. Given that non-biological network theories reserve an important role for environmental influences, it is more likely that cultural and idiosyncratic factors will sculpt the packages. The degree to which environments differ across cultures and across individuals will determine the degree of consistency in the packages that will be found across cultures and individuals. If the variety becomes too great, a dimensional view may become more fitting.

5.3.2.2 Adequacy

APPARENT-SIMILARITY AND FRUITFULNESS

Network theories with a discrete view are well-equipped to deliver the vernacular emotion types that laypeople identify. The fruitfulness of each of these networks could in principle be guaranteed. Network theories with a dimensional view reject the claim that discrete emotions have a special scientific status. The latter theories are sketches of theories that lack detailed hypotheses about dimensions associated with different emotions, and are therefore difficult to falsify (see Delancey, 2005; Kaup & Clarke, 2005).

5.4 Validation

Network theories have inspired various lines of empirical research. Network researchers with a background in learning psychology have studied ways in which emotion networks can be elaborated via learning principles (see research on fear conditioning; Beckers et al., 2013; Boddez et al., 2020; Vervliet & Boddez, 2020). Exploiting the idea that not all representations are equally effective in eliciting emotion, but only those that contain responses (Lang, 1994), intervention strategies to unlearn fear bet strongly on exposure therapy rather than on talk therapy. In exposure therapy for spider phobics, for instance, the aim is to let the patient experience that spiders do not produce physiological responses.

Network researchers with a background in cognitive psychology, on the other hand, have studied the biasing influence of emotions on cognitive functions such as attention and memory, under the supposition that these are key to the onset and persistence of anxiety disorders (MacLeod et al., 2019; McNally, 2019) and depression (LeMoult & Gotlib, 2019; Teachman et al., 2019; F. C. L. Wilson & Gregory, 2018). Network theories bred in the classic computational framework make the straightforward prediction that any source of activation of an emotion network cumulates

https://doi.org/10.1017/9781107588882.009 Published online by Cambridge University Press

with the already present activation. This is in line with all kinds of mood congruency effects in attention and memory (Bower, 1981; Lang, 1994; LeMoult & Gotlib, 2019). It also accords with the finding in emotion regulation research that venting in the case of anger does not curb but rather increases anger (Bushman, 2002; Lohr et al., 2007). Network theories inspired by the DS framework, on the other hand, widen the time window, thereby holding the promise of generating more nuanced hypotheses: After an initial phase of self-amplification, the network may self-organize again into a stable state. Thus, by activating the anger network, anger may increase at first, but perhaps settle down eventually (Khoo & Adkins, 2020; H. L. Smith, 2020). Neither version of network theories, however, is well-equipped to explain contrast effects that are sometimes observed, such as when positive stimuli produce more negative moods (e.g., Pillaud & Ric, 2020) or when negative moods increase attention towards positive stimuli (Schwager & Rothermund, 2013, 2014).

Another line of research examines the peripheral feedback hypothesis – that emotions can be induced from the response side instead of via the stimulus side. This can be achieved by the artificial induction of (a) physiological responses via chemical substances, (b) facial expressions, and (c) coarse behavior. Drugs are supposed to inject us with various feelings. Putting on a happy or sad face is supposed to make people feel happy or sad. And approaching or avoiding stimuli is supposed to breed liking or disliking. But let us consider the evidence.

Few scholars will deny that drugs can change the way we feel (James, 1890b; C. G. Lange, 1885/1922). A dramatic illustration of a negative effect comes from de Montigny (1989), who showed that the intravenous administration of cholecystokinin (a gastric peptide) produced panic attacks in healthy volunteers. Recreational drugs, on the other hand, often lead to positive feelings (Solowij et al., 1992). There is disagreement, however, about whether the influence of drugs on feelings is direct or indirect. Appraisal theories (e.g., Arnold, 1960, p. 111; see Chapter 6) raise the possibility that drugs change the way we perceive or evaluate the world, which in turn changes our feelings. Psychological constructionist theories (e.g., Barrett, 2006b; Schachter, 1964; see Chapter 8) suggest that the arousing (and raw positive/negative) effects of drugs combine with contextual information to shape our feelings.

The facial feedback hypothesis is the hypothesis that facial expressions can have a direct influence on feelings (Adelman & Zajonc, 1989; Allport, 1924; see review by Niedenthal et al., 2006). Two versions of this hypothesis have been formulated: (a) a modulation (or weak) version, according to which facial expressions can influence the intensity of feelings that were themselves caused by stimuli and (b) an initiation (or strong) version, according to which facial expressions can create qualitatively novel feelings. Testing both versions requires demonstrating that facial expressions have a direct influence on feelings rather than an indirect influence that is mediated by other processes such as appraisals and attributions or self-perceptions (D. J. Bem, 1967). To this end, researchers avoid giving participants explicit instructions to put on a happy or sad face, but ask them instead to move parts of their face, or instruct them to do other things which requires them unobtrusively to move parts of their face. For instance, Strack et al. (1988) reported that participants who held a pen between their teeth (producing a smile-like expression) rated cartoons as funnier than participants who held a pen between their lips (preventing a smile) (but see Wagenmakers et al., 2016). Finding evidence for the initiation version of the facial feedback hypothesis is even more daunting as it requires showing that the feelings were caused by the facial expression instead of by stimuli. This entails the use of neutral stimuli or showing that the facial expression pushes feelings in the opposite direction than the stimuli do by themselves.

Despite initial enthusiasm to test the modulation (Laird, 1974; Lanzetta et al., 1976; Strack et al., 1988) and initiation versions of the facial feedback hypothesis (Duclos et al., 1989; Zajonc et al., 1989), later replication studies and meta-analyses concluded that effects are weak and mixed (Coles et al., 2019; Wagenmakers et al., 2016). The cautious interpretation is that if a direct influence of facial expressions on feelings does exist, it must be fragile. And if this influence is already difficult to detect in the laboratory, it should have even more trouble surviving in the wild, especially if the emoter felt a different emotion to begin with. For in that case, activation of one network would have to compete with activation of another network. Indeed, even if putting on a happy face activated the joy network including happy feelings, it would still have to beat an already active sadness network.

Next to facial expressions, there is also evidence for the influence of vocal feedback (Hatfield et al., 1995), postural feedback (Riskind, 1984; Riskind & Gotay, 1982; Rossberg-Gempton & Poole, 1993), and feedback of coarse behavior (Cacioppo et al., 1993; Jones et al., 2013; Kawakami et al., 2007; Koch, 2014; Van Dessel et al., 2016; Wells & Petty, 1980; Woud et al., 2013; see Fuchs & Koch, 2014; Hatfield et al., 1994) on feelings. For instance, Cacioppo et al. (1993) showed that when people were trained to approach certain stimuli and avoid others, they developed a preference for the approached stimuli (but see Clore & Centerbar, 2004). Wells and Petty (1980) showed that nodding the head when hearing one type of message and shaking the head when hearing another type of message led to a more positive attitude towards the former type of message. These effects have been framed as evaluative conditioning effects in which the valence of the responses (USs) transfer to the stimuli (CSs). To the extent

https://doi.org/10.1017/9781107588882.009 Published online by Cambridge University Press

that this transfer is mediated by feelings generated by the responses, these effects support the hypothesis that peripheral feedback from the responses influences feelings. More direct evidence was delivered by Duclos et al. (1989), who showed that when participants adopted fearful, angry, or sad body postures, this influenced the intensity of the corresponding feelings, measured via self-report. This study also went beyond previous studies in that it showed an emotion-specific feedback effect rather than a mere valence feedback effect. Another piece of more direct evidence was provided by Koch (2014), who showed that movements with smooth transitions (e.g., rope skipping) led to more positive affect and receptivity than movements with sharp transitions (e.g., ball-kicking).

To summarize, while evolutionary theories capitalized on innate affect programs in emotion causation, network theories spotlighted learning processes. In the biological version of network theories, affect programs were still allowed to play a role in partitioning, making sure that each vernacular emotion had its proper emotion network. In the non-biological version, on the other hand, the affect programs were thrown overboard. Yet not all proponents of the latter version seemed prepared to accept the full implication of this move, namely that this could also dissolve the vernacular emotions. If learning remains the only active principle, vernacular emotions can only subsist if they are learned. I have argued, however, that there is no guarantee that they will indeed be learned or that they will be learned equally well in each culture and in each individual life.

The next two chapters are dedicated to evaluation theories. The majority of these theories are characterized by the assumption that a process of evaluation sits between the stimulus input and the rest of the emotional episode. The family of evaluation theories houses two subfamilies: stimulus evaluation theories (henceforth, SETs; Chapter 6) and response evaluation theories (henceforth, RETs; Chapter 7). SETs focus on the evaluation of the stimulus input but remain vague about the evaluation of response options. RETs accept some form of stimulus evaluation but shift focus to a response evaluation process. As discussed in the previous chapters, some evolutionary theories (Chapter 4) and network theories (Chapter 5) already made space for a stimulus evaluation process and therefore count as hybrids with SETs.

---

https://doi.org/10.1017/9781107588882.009 Published online by Cambridge University Press

CHAPTER 6

Stimulus Evaluation Theories

6.1 Precursors

SETs have a long history. Precursors from philosophy date back to Aristotle (*Rhetoric*), Hume (1739), Spinoza (1677/1982), Brentano, (1874), Meinong (1894), C. D. Broad (1954), Kenny (1963), Thalberg (1964), Pitcher (1965), and many others. Precursors from psychology are James's (1890b) theory and the non-hybrid forms of evolutionary and network theories that followed. The latter theories can be considered as precursors in the sense that SETs tried to remediate their presumed shortcomings.

Two objections against James (1890b) were that his mechanism did not provide the right kind of (world-directed) Intentionality and that it could not differentiate between the discrete emotions that vindicators were after. To solve the differentiation problem, evolutionary theories postulated innate affect programs (i.e., primary processes) that held preset instructions for packages of responses and feelings that corresponded to a limited set of discrete emotions called basic emotions. Network theories, from their side, emphasized the role of learning processes (i.e., secondary processes), even if these were not in themselves capable of grounding discrete emotions. As extensively discussed in the previous chapters, evolutionary theories took on board some learning processes and some network theories still granted a place to affect programs. Crucially, however, neither the innate nor the learning processes in both theories were able to provide the right kind of Intentionality. Innate processes are tailored to fundamental challenges (e.g., danger, offense, loss). They make sure that when organisms encounter USs (e.g., loud noise, predators, dark places) that fall in the category of a fundamental challenge (e.g., danger), these organisms respond in a way that was appropriate for our evolutionary ancestors (e.g., by fleeing). Learning processes allow for more flexibility in the stimuli triggering these innate responses so that they are better tailored to modern living environments. They ensure, for instance, that flight is not just triggered by USs such as thunder and predators, but also by CSs such as fast-approaching cars and fire alarms. In non-hybrid evolutionary and network theories, in which the stimulus representation that is part of the [S–R] link has a content with concrete features (e.g., the smell of a cat for a rat; [ecS]), the only process required to precede this [S–R] link is a perceptual process that detects these features. There is no need for the organism to understand that the stimulus (e.g., the smell) is important for a fundamental challenge (e.g., that it is dangerous; [eaS]).

SETs disagree that mere perception is enough for emotions. For them, it is not enough for stimuli to fall in the category of certain challenges, the organism must also process this. It is not enough that a stimulus qualifies as dangerous from a third-person perspective (i.e., objectively), the individual must also understand this (i.e., subjectively; even if not necessarily consciously). Only then do emotions have the right kind of Intentionality or meaning (Arnold, 1960). Now, what better way to make sure that an organism represents danger than by implanting an evaluation process that takes the stimulus as its input and produces a representation with the abstract content "danger" as its output. This is the solution that the majority of SETs propose for endowing emotions with the right kind of Intentionality (Arnold, 1960; Irons, 1879a, 1879b). These theories all belong to a brand of SETs that I will call "evaluation-first SETs."

An alternative solution to the Intentionality problem is provided by a second brand of SETs, called "embodied SETs." These theories propose that stimulus evaluation is embodied in the action tendencies and/or the physiological responses that follow on the (non-evaluative) perception of the stimulus (Deonna & Teroni, 2012, 2014, 2015; Dewey, 1894; J. J. Prinz, 2004a),[57] or that stimulus evaluation and response components are so entangled that they can no longer be separated (e.g., M. D. Lewis, 2005; Colombetti, 2003, 2009).

In sum, all SETs share the idea that evaluation of the stimulus is part of the broad emotional episode, but different brands of SETs differ with respect to the form of this evaluation process and/or the place it occupies in the emotional episode. This chapter's main focus will be on the first brand of SETs. It is for this brand that I will scroll through the phases of the demarcation-explanation cycle: constitutive and mechanistic explanations (Section 6.2), scientific definitions (Section 6.3), and validation (Section 6.4). The second brand will be discussed more concisely at the end of the chapter (Section 6.5).

6.2 Constitutive and Mechanistic Explanations

In evaluation-first SETs, the stimulus evaluation process (Step 1) precedes any of the other components of the emotional episode (Step 2). For

[57] Note that embodied SETs deviate from the earlier description of SETs as theories in which a process of stimulus evaluation sits between the stimulus and the rest of the emotional episode.

instance, a crouching tiger is evaluated as dangerous before leading to the action tendency to flee, a racing heart and sweaty palms, actual fleeing behavior, and feelings of fear. Another example that brings out the crucial role of stimulus evaluation even more is the case of hearing a noise in the hall at night. The sound has no emotion-eliciting power in itself. If I evaluate it as a threat to my safety because it might be a burglar, it leads to fear. If I evaluate it as a disturbance to my sleep caused by my teenage son coming home drunk, it may lead to anger. And if I evaluate it as irrelevant because I know it is the cat jumping off the table, it produces no emotion.

Theories that belong to this brand are known under the names of (a) appraisal theories (Arnold, 1960; Ellsworth, 1991, 2013; Frijda, 1986, 2007b; Lazarus, 1991; Roseman, 1984, 2013; Scherer, 1984, 2001a, 2009b; C. A. Smith & Ellsworth, 1985; see Moors et al., 2013), (b) judgmental theories (Castelfranchi & Miceli, 2009; Gordon, 1987; Green, 1992; W. Lyons, 1980; Marks, 1982; Neu, 2000; Nussbaum, 2001; Reisenzein, 2012; Searle, 1983; Solomon, 1993), (c) quasi-judgmental theories (Armon-Jones, 1991; Greenspan, 1988), and (d) perceptual theories (Deonna, 2006; de Sousa, 1987; Döring, 2007; Johnston, 2001; Roberts, 2013; Tappolet, 2000, 2016; Tye, 2008). Appraisal theories belong to psychology; judgmental theories, quasi-judgmental theories, and perceptual theories are prominent in philosophy.[58] Appraisal is just another word for evaluation, but I will reserve it for the stimulus evaluation process put forward by appraisal theories.

I will compare SETs from both disciplines in one go with the help of six axes. First of all, SETs vary in the role they confer to stimulus evaluation (**Axis A**). Some take it to be a mere constituent of emotion (Figure 6.1(a)), others take it to be the cause of emotion (Figure 6.1(b)), and still others take it to be a constituent of emotion as well as the cause of the remaining components of the emotion (Figure 6.1(c)) (see also Reisenzein, 2012; Scarantino, 2016; Scarantino & de Sousa, 2018).

Constituent-only SETs can be *pure*, in which case they see stimulus evaluation as a necessary and sufficient component of emotion. Judgmental theories that equate emotions with evaluative judgments can be situated here (e.g., Solomon, 1993). According to these theories, fear is the judgment that one is in danger and sadness the judgment that something is lost for good. But also perceptual theories that equate emotions with evaluative perceptions belong in this group (e.g., de Sousa, 1987).

[58] Reisenzein (2012) sits between chairs, as he is a psychologist whose theory leans most closely to judgmental theories in philosophy. The theory qualifies as a naturalized theory because it spells out the mechanistic details.

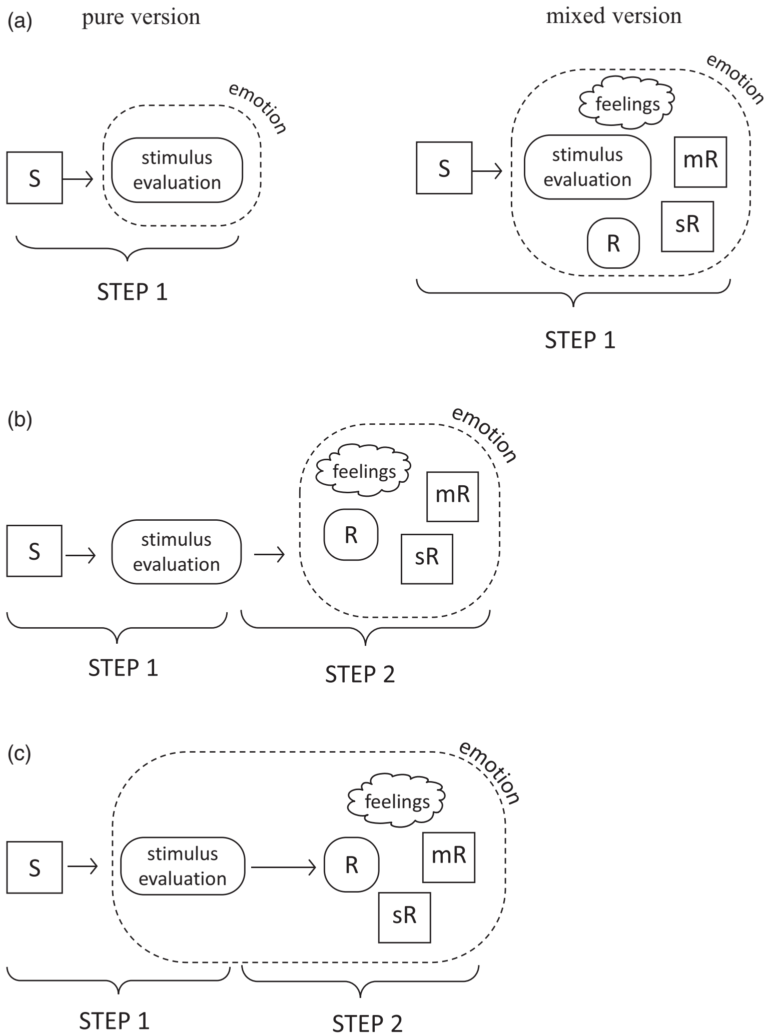


Figure 6.1 Versions of SETs according to Axis A: (a) constituent-only version; (b) causal-only version; (c) constituent-causal version

For them, fear is the perception that one is in danger and sadness the perception that something is lost for good.

Constituent-only SETs can also be *mixed*, in which case stimulus evaluation is a necessary component, but other components must be present as well. These theories take emotions to be aggregates of various components without specifying the relations among the components.

https://doi.org/10.1017/9781107588882.010 Published online by Cambridge University Press

A strand of judgmental theories called constituent belief-desire theories fits here. Trying to bring order in this strand is tricky because the "belief" is sometimes understood as an evaluative judgment and sometimes as a non-evaluative judgment, and the "desire" is sometimes understood as a non-behavioral goal and sometimes as an action tendency (see also Deonna & Teroni, 2012, chapter 3). Sampling from these possibilities, some theorists combine an evaluative judgment with an action tendency (Green, 1992; Marks, 1982). They would say, for instance, that fear is the belief that one is in danger combined with the desire to flee. This is sometimes enriched with feelings (e.g., W. Lyons, 1980). Other theorists combine a non-evaluative judgment with a non-behavioral goal (Castelfranchi & Miceli, 2009; Miceli & Castelfranchi, 2015).[59] They would say, for instance, that a negative emotion is the belief that some stimulus is present and the desire that the stimulus was absent. This can also be enriched with feelings. Mixed theories also differ in other respects. In *weak emergent* theories (e.g., Marks, 1982), the whole equals the sum of the parts. In *strong emergent* theories, the whole is more than the sum of the parts (Castelfranchi & Miceli, 2009; Green, 1992; see Reisenzein, 2012).

Causal-only SETs take stimulus evaluation to be the cause of emotion. Here, evaluation is separate from the emotion. Theories that fit here are a handful of appraisal theories (e.g., Frijda, 1986; Roseman, 2013) and a strand of judgmental theories called causal belief-desire theories (e.g., Reisenzein, 2009, 2012). In the latter theories, the belief is a non-evaluative judgment and the desire is a non-behavioral goal. The comparison of the belief and the desire is an evaluative judgment and this is the cause of the ensuing emotion. Authors of these theories would say, for instance, that a negative emotion is caused by the comparison between the belief that some stimulus is present and the desire that the stimulus was absent. The constitutive explanation of emotion can be narrow, including only feelings (Reisenzein, 2009) or broader, including also action tendencies, and/or somatic responses (Frijda, 1986; Roseman, 2013).

Finally, *constituent-causal* SETs take stimulus evaluation to be a necessary or at least typical[60] component of emotion that is also the unique driver

[59] Castelfranchi and Miceli (2009) stated that although they include beliefs and desires as necessary constituents of emotions, they make the further claim that an evaluative judgment (in which the beliefs and the desires are compared) is not required. So perhaps their theory cannot properly be called an SET.

[60] Most contemporary appraisal theories accept marginal cases in which emotions are not caused by appraisal with the aim of sidestepping endless discussions about necessity (see debate between Lazarus, 1984, and Zajonc, 1980; Section 6.4).

of all the other components in the emotion. Many contemporary appraisal theories fit in this group (e.g., Frijda & Parrott, 2011; Scherer, 2009b; see Moors, 2013b).

Let us now zero in on the different steps in the emotional episode. Constituent-only SETs have only one step in their causal chain during which evaluation of a stimulus (e.g., tiger) results in a representation with an abstract evaluative content (e.g., danger), and this content *is* the emotion, either alone (pure theories) or together with other components (mixed theories). The causal-only and constituent-causal SETs assume that after this first step is taken, the output representation gets translated into the other components in subsequent steps.

Step 1

The process of stimulus evaluation takes as its input either (a) a raw external stimulus (eS) or (b) a stimulus representation with concrete features ([ecS]), and produces as its output a representation with abstract features ([eaS]). A crouching tiger is either directly evaluated as dangerous or it is first perceptually preprocessed as a tiger or a creature with sharp teeth before being evaluated as dangerous.

SETs vary with respect to the way in which they characterize the evaluation process in terms of the granularity of the content (**Axis B; Axis 5b**), the precise content (**Axis C; Axis 5a**), and the format of the output representations (**Axis D; Axis 6c**), the operations producing these representations (**Axis E; Axis 6d**), and the conditions under which this all happens (**Axis F; Axis 6e**).

A first distinction worth making has to do with granularity (**Axis B**). *Molar* SETs propose molar values such as danger, offense, loss, and so forth. Advocates of this approach are rife in judgmental theories, quasi-judgmental theories, and perceptual theories. They portray the stimulus evaluation process as the extraction of evaluative properties from the stimuli in terms of how dangerous, offensive, admirable, and disgusting they are.[61]

Molecular SETs, such as appraisal theories and certain belief-desire theories, propose molecular values. They unpack the molar values by searching for the minimal number of variables needed to capture the

[61] Some philosophers (Tappolet, 2016) do not list properties such as danger and loss, but rather properties that remain literally closer to emotion words, such as fearsome and sad. As the property "dangerous" justifies the property "fearsome" and "fearsome" justifies "fear," de Sousa (2018) argued that the middle-term "fearsome" is redundant and could be left out.

differences between them. Appraisal theories, for instance, unpack danger and loss in terms of the variables goal incongruence or valence, certainty, and controllability. Each molar value can be described with a different pattern of values on these variables, that is, molecular values. Dangers and losses are both incongruent with some goal or desire and thus negative, but dangers involve an uncertain (future) goal incongruence whereas in losses the goal incongruence is certain (present). SETs can be purely molecular, but they can also combine with a molar approach. *Purely molecular* SETs sometimes still use the molar values as summary labels. *Molar-molecular* SETs, on the other hand, take the summarizing itself as part of the mechanism involved in the transition towards the other components (see below; Moors, 2014a; see Teroni, 2021).

A second distinction worth making concerns the precise content of the molar/molecular values (**Axis C**). Appraisal theorists have taken up the challenge to discover the molar appraisals and/or the patterns of molecular appraisals for each of the vernacular emotions they want to see explained.[62] When it comes to molar appraisals, most theorists follow Lazarus's (1991) proposal to link fear to danger, anger to a demeaning offense against me or mine, sadness to an irrevocable loss, guilt to having transgressed a moral norm, shame to failing to live up to an ego-ideal, and pride to the enhancement of one's ego-identity by taking credit for a valued object or achievement by oneself or one's group.

If we turn to molecular appraisals, the variation between theories rises significantly. Examples of appraisal variables about which there is still fair consensus are: goal relevance, goal congruence, certainty, expectedness, controllability, and agency. Goal relevance refers to the impact that a stimulus has on the satisfaction status of a goal.[63] Goal in/congruence indicates whether the stimulus mis/matches and hence frustrates/satisfies a goal. Goal relevance determines whether an emotion occurs and how intense it will be; goal in/congruence determines whether the emotion will be negative/positive. For example, meeting a tiger in the woods is incongruent with my goal to be safe and has more impact on this goal than meeting a medium-size dog. Further differentiation within the sets

[62] Note that hypotheses about links between appraisal variables and emotions belong to Step 1 for constituent-only SETs only. For causal SETs, they belong to the subsequent steps.

[63] Walentowska et al. (2016; see also Severo et al., 2017, 2018) distinguished three partially dissociable meanings of goal relevance: (a) *reliability*, which is the degree to which a stimulus informs about the satisfaction/frustration status of a goal, (b) *impact*, which is the magnitude of the satisfaction/frustration, and (c) *task relevance*, which is the degree to which the stimulus forms an opportunity for the implementation of a behavioral or task goal. In emotion theory, impact is most important, but this presupposes a sufficient degree of reliability.

https://doi.org/10.1017/9781107588882.010 Published online by Cambridge University Press

of positive and negative emotions can be obtained by adding the remaining appraisal variables. For instance, in fear and hope, the stimulus is pending and therefore uncertain, whereas in joy, anger, and sadness, the stimulus has already happened and is thus certain. Before a student receives her grades, she may be torn between fear and hope. Once received, good grades make her happy whereas bad grades make her angry or sad. Un/expectedness is considered as a booster/attenuator of both positive and negative emotions (Moors et al., 2021; Scherer, 1988). Unexpected goal in/congruence leads to more intense negative/positive emotions than expected goal in/congruence. The failing/passing of an exam leads to more un/happiness when it is unexpected than when it is expected. The ease/difficulty to control the stimulus (i.e., to turn a goal-incongruent stimulus into a goal-congruent one) differentiates anger and fear. Anger occurs when a goal-incongruent stimulus is easy to control; fear occurs when a goal-incongruent stimulus is difficult to control. As Ellsworth and Scherer (2003, p. 580) put it, "[i]n the case of an obstructive event brought about by a conspecific aggressor or predator, the comparison between the organism's estimate of its own power and the agent's perceived power is likely to decide between anger and fear and thus between fight and flight." Finally, agency refers to whether the stimulus was caused externally (by others)[64] or internally (by oneself). External agency contributes to anger about a goal-incongruent stimulus and gratitude for a goal-congruent stimulus. Sam will be angry at Sunny if she eats the last cookie but grateful if she offers it to him. Internal agency contributes to pride about a goal-congruent stimulus and guilt about a goal-incongruent stimulus. Sunny will be proud if she offered Sam the last cookie but feel guilty if she ate it. These effects should be boosted if the agent caused the goal-in/congruent stimulus on purpose and is therefore to blame/praise for it. Sam will be more angry at Sunny if she ate the last cookie on purpose than if she did it by accident.

Hypotheses in appraisal theories show substantial overlap with those in certain belief-desire theories (e.g., Castelfranchi & Miceli, 2009; Reisenzein, 2012). The latter theories put the variable of goal in/congruence, or in their terminology, the mis/match between beliefs (about the actual state) and desires (the desired state or goal) center stage. Other variables such as certainty and expectedness have been added to increase the scope of these theories from simple positive/negative emotions to specific emotions. For instance, un/happiness arises from the mis/match between a belief and a desire, fear/hope arises when this mis/match is

[64] Several theories split external agency into other-agency and circumstances-agency. For instance, fear can arise from threats caused by others (e.g., an aggressive conspecific) or impersonal circumstances (e.g., earthquake, illness).

https://doi.org/10.1017/9781107588882.010 Published online by Cambridge University Press

uncertain, and disappointment/relief arises when a mis/match is unexpected.

SETs not only differ in terms of the granularity and the precise content of the representations resulting from the stimulus evaluation process, but also in terms of the format of these representations and/or the Attitudes held towards them (**Axis D**). Judgmental theories (e.g., Reisenzein, 2009; Solomon, 1993) take the representations involved in the evaluation process to be doxastic representations whereas quasi-judgmental and perceptual theories believe them to be infradoxastic (see Charland, 1997). Doxastic representations are propositional representations to which the consumer has an Attitude of belief (i.e., accepting the proposition as true) and/or an Attitude of desire (i.e., wanting the proposition to come true). Infradoxastic representations are ones that are in some sense less than fully-fledged believed/desired propositional representations. Quasi-judgmentalists propose propositional representations that are not necessarily believed, but merely entertained or construed (e.g., Armon-Jones, 1991; Greenspan, 1988). For instance, to feel fear, one need not firmly believe that one is in danger, one may also construe or even imagine being in danger. Perceptualists (e.g., Deonna, 2006; de Sousa, 1987; Döring, 2007; Johnston, 2001; Roberts, 2013; Tappolet, 2000, 2016; Tye, 2008) propose perceptual representations. To feel fear is not to believe or entertain the thought that one is in danger but rather (a) to perceive that one is in danger – according to more liberal strands of perceptualism (e.g., Döring, 2007; Roberts, 2013)[65] – or (b) to perceive danger – according to more literal strands of perceptualism (e.g., Tappolet, 2016). The perception *that one is in danger* is a propositional representation to which the consumer takes an Attitude of "appearance of truth" (Döring, 2007, p. 378; Roberts, 1988). In fear, it appears to be true that one is in danger even if one does not (yet) assent to its truth. The perception *of danger*, on the other hand, is a simple or nominal representation. Simple representations cannot be evaluated as true or apparently true. One cannot judge "danger" to be true just as one cannot judge "a tree" to be true. Nevertheless, some perceptualists argue that the consumer can still take an Attitude of "presence" towards simple representations. In perceiving danger, one takes danger to be present (see Tucker, 2013).[66]

Doxastic and subdoxastic representations are further aligned with specific types of operations (**Axis E**) and specific sets of operating conditions

[65] The boundary between the quasi-judgmental theories and liberal strands of perceptualism is fluid, and so grouping authors into one or the other category is tricky and must therefore be taken with a grain of salt. Roberts (2013), for instance, uses the term "construal" but nevertheless cashes this out in terms of perception.

[66] Thanks to Fabrice Teroni for bringing this to my attention.

https://doi.org/10.1017/9781107588882.010 Published online by Cambridge University Press

(**Axis F**), in classic dual-system style. Doxastic representations belong to System 2. They are assumed to be generated by rule-based computation in a non-automatic way (i.e., under ample operating conditions). Infradoxastic representations belong to System 1. They are assumed to be generated by the activation of learned or innate associations between stimuli and evaluations (e.g., [tiger–danger]) in an automatic way (i.e., under poor operating conditions). Perceptual representations, moreover, are endowed with properties such as modularity, encapsulation, and cognitive impenetrability (Charland, 1995, 1997; Fodor, 1981; Tappolet, 2010, 2016). The upshot of these characteristics is that perceptions can be resistant to changes in beliefs. This is exemplified in perceptual illusions. The perception of a stick that is half-submerged under water as bent is not altered by the knowledge that the stick is straight. Recalcitrant emotions are modeled after perceptual illusions. Here too, the perception of oneself as being in danger is not altered by the knowledge that one is safe.

SETs in philosophy tend to put most of their eggs in one basket in that they capitalize either on System 2 or on System 1: Judgmental theories lean more on System 2 whereas quasi-judgmental and perceptual theories rely more on System 1 (see also Charland, 1997). Appraisal theories in psychology, on the other hand, embrace both systems. They believe emotions can be caused either by rule-based computation (System 2) or by the activation of preset (learned or innate) associations between stimuli and evaluations (System 1) (Clore et al., 2005; Frijda, 2007b; Kappas, 2006; Lazarus, 2001, pp. 51–52; Leventhal & Scherer, 1987;[67] C. A. Smith & Kirby, 2000, 2001; E. R. Smith & Neumann, 2005; van Reekum & Scherer, 1997). As appraisal theories have worked out the details of the operations involved in the two systems, it is worth taking a closer look (see Moors, 2010a, 2010b, 2013a, 2013b, 2014a, 2014b).

At first, stimuli are evaluated on each of the hypothesized appraisal criteria via rule-based computation, and this results in a pattern of appraisal values (System 2). Once such a pattern has been computed for a stimulus, an association between the stimulus representation and the representation of the pattern is stored in memory so that on subsequent occasions, the stimulus directly activates this association and reinstates the pattern, thereby circumventing the more complex process of computing every appraisal value anew (System 1). Here also, the dichotomy between operations is aligned with dichotomies between formats of

[67] Leventhal and Scherer (1987; van Reekum & Scherer, 1997) have a triple-system model with a sensorimotor (innate), a schematic (learned), and a conceptual (computation) layer. The sensorimotor system does not just provide a mechanism for evaluation but also for the connection between the afferent stimulus representation and the efferent motor representation ([S–R]). This already belongs to Step 2.

representations and operating conditions. Thus, rule-based appraisal operates on conceptual representations and is non-automatic (System 2) whereas associative appraisal operates on perceptual representations and is automatic (System 1). These alignments are not always strict, however, and small deviations can be noted. For instance, Clore and Ortony (2000, p. 39) admitted on one occasion that "rule-based reasoning is not necessarily conscious, explicit, or deliberative" but "can be utterly implicit, as evidenced by the fact that it can be demonstrated even in preverbal infants (e.g., Kotovsky & Baillargeon, 1994; Needham & Baillargeon, 1993)." C. A. Smith and Kirby (2000, 2001) noted that associative processes are not limited to perceptual representations but can also operate on conceptual ones. Frijda (2007b, p. 108) argued that the separate processing of each appraisal value requires online pick up or detection (instead of rule-based computation), but given the sequential nature of this process, it is time-consuming and therefore non-automatic. Regardless of whether the alignments are strict or more liberal, however, the fact remains that theorists make them on an a priori basis, sometimes propped up by one-sided evidence. It can be argued, however, that dichotomies are best treated as orthogonal until proven otherwise (Moors, 2014c; Moors & De Houwer, 2006b). In previous years, this led me to defend an appraisal view, which left open the possibility that vertically complex (i.e., multiple-input or constructive) operations (whether rule-based or associative) involved in appraisal can nevertheless be automatic (Moors, 2010a, 2010b).

Subsequent Steps

Constituent-only SETs equate emotion with the evaluation in Step 1, either pure or mixed with other components. In the perceptual variant, for instance, the process of perception is automatic, but the output of this process is often a conscious perception or perceptual experience. In these theories, the feeling component is included in, or directly attached to, the cognitive component, so to speak. It could be argued, however, that as long as feelings logically follow on the perception process, they count de facto as a subsequent step. It is further worth noting that constituent-only theories that do not include any response-related components inside the emotion sometimes still provide an account of how emotions influence these components in subsequent steps. The mechanisms they propose for this transition are similar to those proposed by causal SETs (e.g., Tappolet, 2010, 2016).[68]

[68] Tappolet (2010, 2016) suggested that the perceptual evaluation of the stimulus can, but does not have to, activate an action tendency. While rejecting the idea that evaluative properties activate concrete action tendencies (called behavioral dispositions), she

https://doi.org/10.1017/9781107588882.010 Published online by Cambridge University Press

In all causal SETs (i.e., causal-only and causal-constituent SETs), stimulus evaluation is openly considered to be the cause of the other components in the broad emotional episode. Appraisal theories have provided details about the mechanisms involved in this transition. Like network theories, appraisal theories come in a biological and a non-biological version (**Axis G**; see Figure 6.2) (Moors, 2014a, 2014b). The biological version is an evaluation-evolutionary hybrid, which forms the counterpart of the evolutionary-evaluation hybrid discussed in Chapter 4. The two types of hybrid theories mainly differ in the mechanistic step that they focus on: the former on the appraisal process, the latter on the affect program. The non-biological version of appraisal theory, on the other hand, manages without the mediation of affect programs. I discuss both versions in turn.

Biological version. Biological appraisal theories (e.g. , Arnold, 1960; Irons, 1879b; Lazarus, 1991; Oatley & Johnson-Laird, 2011; Roseman, 1984, 2011, 2013; a.k.a. Flavor 1 in Moors, 2014a; see Figure 6.2(a)) propose the following mechanism for the transition from Step 1 (appraisal) to the subsequent steps (other components). The pattern of molecular appraisal values generated in Step 1 is first summarized in a molar appraisal value. For instance, the pattern goal-relevant, goal-incongruent, and difficult to control is summarized as danger. This summary appraisal value subsequently activates an innate [S–R] link – an affect program – framed on an abstract level (e.g., [danger–avoidance]) (Oatley & Johnson-Laird, 2011; Roseman, 2011). Like in evolutionary theories, the affect program fixes the emotion's identity. The tendency to avoid caused by danger fixes fear, for instance, and the tendency to harm caused by offense fixes anger. The abstract nature of the [S–R] links grants flexibility on the input side as well as on the output side. Several concrete stimuli can be appraised as dangers (e.g., wild animals, climate change) and several concrete responses can implement avoidance (e.g., flee, hide, avert gaze). Responses and feelings generated by the [S–R] link are a foregone conclusion unless they are hindered by a goal-directed regulation process that targets these components (see Box 7.1). For instance, unfair treatment by one's boss may blindly generate the tendency to attack, but the long-term goal to keep one's job may prevent the tendency from manifesting itself in behavior. In addition to their role in emotion regulation, goal-directed processes are also involved in planning. They help translate

replaced them with abstract action tendencies (called desires), which can be concretized in a later stage to adapt to the specifics of the situation. This is entirely in line with appraisal theories and evolutionary-evaluation hybrids (see Chapter 4). The existence of so-called purely contemplative emotions, such as emotions caused by fictional events (e.g., in a movie), led her to argue that action tendencies are not a must.

https://doi.org/10.1017/9781107588882.010 Published online by Cambridge University Press

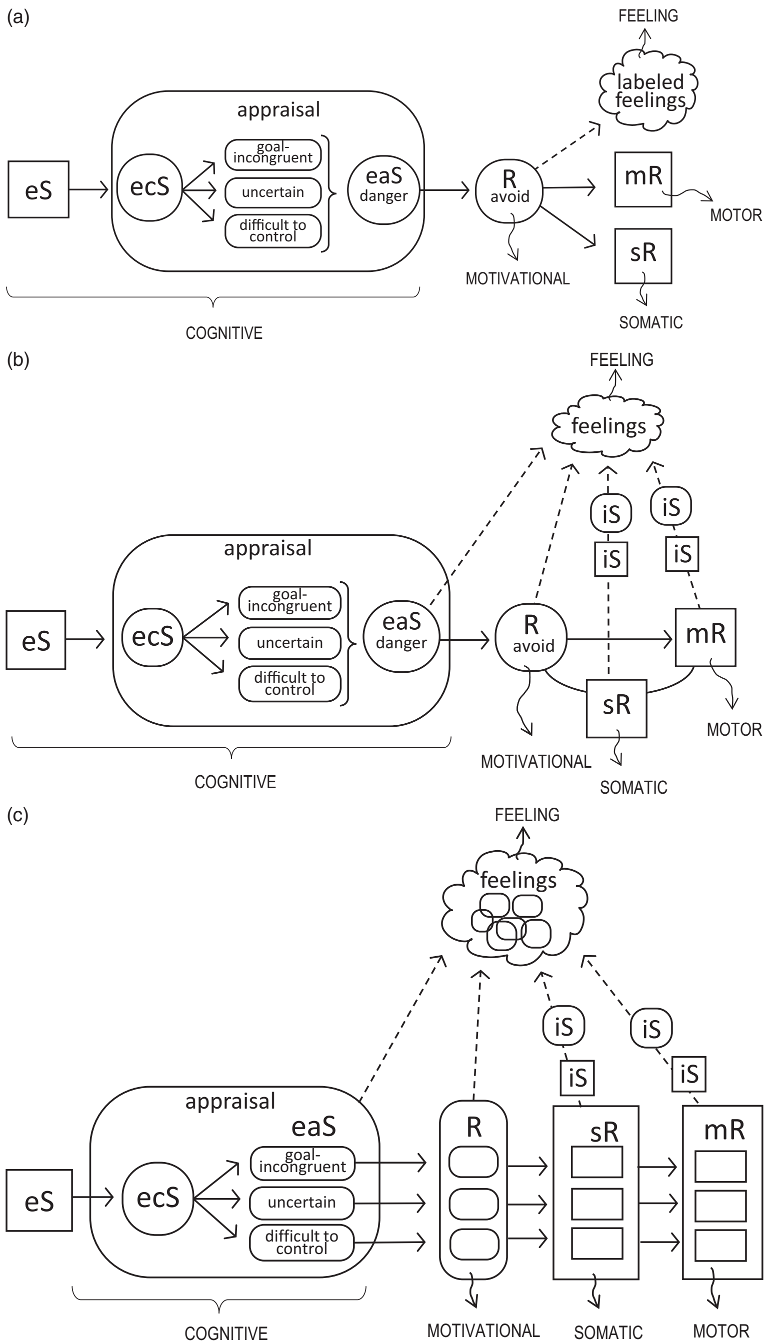


Figure 6.2 Versions of appraisal theories according to Axis G: (a) biological version = evaluation-evolutionary hybrid version; (b) non-biological summary version; (c) non-biological elemental version

https://doi.org/10.1017/9781107588882.010 Published online by Cambridge University Press

abstract action tendencies into concrete ones. If nothing is standing in the way of attack, the person still needs to decide whether to attack with the fists or with words. Like in evolutionary theories, the idea that stimulus-driven processes are responsible for the emotion and goal-directed processes are responsible for emotion regulation and planning matches a dual-system model with a default-interventionist architecture (see Box 2.1).

Non-biological version. Non-biological appraisal theories reject the idea that an innate affect program mediates between appraisal and the other components. This variant can itself be split into two sub-versions: a summary sub-version and an elemental sub-version. In the *summary* sub-version (see Figure 6.2(b)), the molecular appraisal pattern is again summarized into a molar appraisal value that determines the ensuing action tendency. Although in this non-biological version, the action tendency ([R]) is not part of an innate [S–R] link, the hypotheses follow closely on those put forward by the biological version. Danger is still hypothesized to lead to the tendency to avoid, for instance. This is presumably based on the assumption that organisms had the chance to learn that avoidance is still (even in the present day) the best action one can take when faced with danger, at least on average, or when there is no information available about the stimulus other than that it is dangerous. The action tendency, in turn, is manifested in somatic responses, which prepare and support overt behavior. Aspects of appraisals, action tendencies, and overt responses percolate into consciousness where they together form the content of the feeling component.

Although I am not aware of any appraisal theorist who explicitly endorses this sub-version, I have seen glimpses of it in the writings of Arnold (1960) and Lazarus (1991), as if they are not fully committed to the biological view. It might also be compatible with Frijda and Parrott (2011), who propose that a number of highly abstract action tendencies, called "ur-emotions," figure among other components in the multi-componential episodes called emotions. Examples of ur-emotions are the tendency to accept, dominate, be submissive, avoid, and reject. They are often, but not always, caused by appraisal, and they are manifested in concrete action tendencies, somatic responses, and behavior. Ur-emotions are supposed to be innate and universal and to be few in number because there are only so many ways in which organisms can engage with the world. The theory differs from the biological version, however, in that the hard liquor in this theory is not "affect programs," understood as innate [S–R] links, but only encompasses the [R] part. This means that, in principle, it allows for more flexibility between stimuli (as appraised) and action tendencies. This is reflected in the idea that there is not a one-to-one relation between ur-emotions and vernacular emotion types. A single ur-emotion (e.g., the tendency to be submissive) can figure in different vernacular emotions (e.g., shame, awe, and admiration) characterized by different eliciting stimuli or appraisals (e.g., failure to live up to an ego-ideal, grand nature, and an admirable person).

In the *elemental* sub-version (e.g., Scherer, 2000, 2001a, 2009a, 2009b; a.k.a. Flavor 2 in Moors, 2014a; see Figure 6.2(c)), finally, each of the molecular appraisal values directly activates a molecular action tendency value. To give just a few examples, goal relevance increases the overall intensity of the action tendency, goal in/congruence activates the direction of the action tendency (e.g., approach/avoidance), ease/difficulty to control determines the direction of adaptation that the action tendency tries to achieve (e.g., stimulus to person, person to stimulus), and internal/external agency determines the target of the action tendency (e.g., self/other). Again, hypotheses are chosen based on the action tendency values that seem to best fit the appraisal values on average.[69] In turn, these molecular action tendency values may activate molecular values of somatic responses, and further downstream, molecular values of motor responses. As the components take shape, aspects or fragments of them may ooze into consciousness, where they are integrated in an emotional experience (i.e., feeling component). This elemental sub-version differs from the other two versions in that the organism does not need to determine at any point which emotion is at stake. Instead, emotion is an emergent phenomenon. The nature of the emotion becomes apparent when all components have taken on a value. This emotion, moreover, may remain entirely unlabeled. It may also get labeled, and this labeling may color the person's experience, but the labeling process is not part of the emotion itself, narrowly understood.

It is important to note that proponents of the elemental sub-version apply two nuances to the strict order of components presented above (Ellsworth & Scherer, 2003; Frijda, 2007b; Moors & Scherer, 2013; Scherer, 2001a). A first nuance, known as "immediate efference," entails that early components (e.g., appraisal) need not be entirely completed before they can influence later ones (e.g., action tendencies, responses, feelings). As soon as a single molecular appraisal has been made, it can already have downstream influences on other components. A second nuance, known as "recurrence," entails that changes in later components are fed back to earlier components. Changes in responses and feelings are fed back to the appraisal component (either directly or via the stimulus), where reappraisal occurs. For instance, running away from a goal-incongruent stimulus that is difficult to control leads to a new stimulus, which is

[69] Note that although innate connections between appraisals and action tendencies are rejected on the molar level, they may not be excluded on the molecular level.

appraised anew as goal-congruent or easy to control. But the somatic responses that prepare for running may also lay bare a hitherto unknown source of energy leading to an appraisal of the stimulus as easier to control. Feelings may also strengthen appraisals. For instance, fear may strengthen the appraisal of uncontrollability (Arntz et al., 1995) and anger may strengthen the appraisal that someone is to blame, or in Frijda's (2007b) words, "[a]nger searches [for] a reason, a justification, and a culprit" (p. 98). Because of recurrence, several cycles may run in parallel, with appraisal as the cause of every new cycle and the effect of every previous cycle. The picture that emerges in this way is one of massive recurrence from the other components back to the appraisal component in iterative cycles (Cunningham et al., 2013; Cunningham & Zelazo, 2007; Scherer, 2000).

Immediate efference and recurrence are two features that fit well with network theories discussed in Chapter 5, and in particular with the DS framework adopted by Colombetti (2009) and M. D. Lewis (2005). This is reflected in the fact that some appraisal theorists such as Scherer (2000, 2009a) have also adopted a DS framework. Although this reduces the gap between these theoretical families substantially, two differences are still worth pointing out. First, network theories postulate equipotent bidirectional relations among all components whereas appraisal theories postulate recurrence in which appraisal is the motor of every new cycle. This means that in the latter theories, feelings can influence action tendencies and responses if they first travel again via appraisal. Thus, feelings of fear of a house spider can themselves be appraised as incongruent with the goal to appear brave but as difficult to control, and thereby amplify the tendency to avoid.

A second difference is that Scherer's (2000, 2009a) DS application is less radical than those of Colombetti (2009) and M. D. Lewis (2005). The latter authors advocate a complete merging of the traditional components thereby losing any anatomical and functional separation between them. Scherer (2000, 2009a), by contrast, sides with the view (described at the end of Chapter 5) that a DS approach that acknowledges massive recurrence is not incompatible with a classic approach that analyzes a single cycle (see also D. Evans, 2008).

Like network theories, elemental appraisal theories can be aligned with the psychometric network (PN) model also mentioned in Chapter 5. In that chapter, the PN model was contrasted with the latent variable (LV) model that undergirds evolutionary theories (see Coan, 2010; Coan & Gonzalez, 2015; J. Lange et al., 2020). In a recent paper, J. Lange et al. (2020) further specified that evolutionary theories fit with a *reflective* latent variable (RLV) model, which assumes that the emotion (latent variable, which should ultimately be replaced by a unique brain

https://doi.org/10.1017/9781107588882.010 Published online by Cambridge University Press

mechanism) is the common cause of (i.e., is reflected in) all other components. The RLV model predicts strong covariation among the components. The authors contrasted this model with a *formative* latent variable (FLV) model, which assumes that emotion is not the cause of all components but rather an emergent entity resulting from the components. In weak emergence, emotion is the sum of the components; in strong emergence, emotion is more than the sum of the components, that is, new qualities arise that were not present in the components as such. In the FLV model, the components need not covary. In fact, the FLV model fits best with mixed constituent-only SETs, which hold that emotions have components without specifying the relations among the components (e.g., Castelfranchi & Miceli, 2009). This is in sharp contrast with the massive interactions among components that are postulated in the PN model adopted in network theories and elemental appraisal theories. Nevertheless in the PN model, the emotion is still identified with the Gestalt emerging from these interactions.

6.3 Scientific Definitions

6.3.1 Intensional Definition

6.3.1.1 Criteria for Demarcation

The constitutive and mechanistic explanations proposed in SETs yield the following criteria for demarcating the set of emotions from other sets. The criteria are considered to be necessary by some (as per a classic intensional definition) but only typical by others (as per a cluster-type definition). SETs take stimulus evaluation to be a central component of the broad emotional episode. This is the first criterion, which I call the "stimulus evaluation" criterion. In constituent-only SETs, this evaluation is also a central component of the emotion narrowly understood, either alone (in the pure version) or together with other components (in the mixed version). In causal-only SETs, stimulus evaluation is separate from, but a cause of, the emotion. In constituent-causal SETs, evaluation is both a component and a cause of the other components in the emotion.

The stimulus evaluation criterion, however, is not in itself sufficient to demarcate emotional from non-emotional episodes. The operation of a stimulus evaluation process is not unique to emotion. Individuals are constantly scanning their environment for challenges and opportunities. Only when the output of this evaluation process has a particular content does an episode qualify as emotional. This is the second criterion, which I call the "content" criterion. In purely molar SETs, stimulus evaluation results in the representation of an evaluative property such as danger, offense, or loss. Here, the "content" criterion takes the form of an "evaluative property" criterion. In molecular SETs, on the other hand, an episode is emotional if it is evaluated as relevant to a goal with a sufficiently high value. This is expressed in Frijda's (2007b, p. 351) law of concern, which articulates the idea that emotions are goal relevance detectors (see also Frijda, 1986, 1994; Lazarus, 1991; Moors, 2007, 2009; Oatley & Johnson-Laird, 1987; Ortony et al., 1988; Reisenzein, 2009; Roberts, 1988; Roseman, 1984; Scherer, 1994a; C. A. Smith & Lazarus, 1993).[70] Here, the "content" criterion takes the shape of a "goal relevance" criterion. It could be argued that goal relevance is inseparable from goal congruence given that a stimulus cannot be relevant for a goal without being congruent or incongruent with it. Thus, the goal relevance criterion may be said to conceal a "goal congruence" criterion.[71]

Judgmental, quasi-judgmental, and perceptual theories are also bound to a third criterion, called the "system" criterion, covering formats and Attitudes, operations, and operating conditions. In judgmental theories, representations of evaluative properties are doxastic and generated by non-automatic rule-based operations (System 2), whereas in quasi-judgmental and perceptual theories, they are subdoxastic and generated by automatic associative operations (System 1). This restriction does not apply to appraisal theories because they embrace multiple systems.

A fourth criterion for demarcation is only present for theories that include subsequent steps beyond the first step of stimulus evaluation (i.e., all causal SETs). While the stimulus evaluation process (independent of its content) is considered to be a general-purpose mechanism, the [S–R] link that subserves the transition from the stimulus evaluation to the action tendency is instead a special-purpose mechanism. The connection

70 Although all appraisal theorists believe that goal relevance is a central appraisal, not all consider it to be strictly necessary for all emotions. In Scherer's (1984) theory, emotions can already result from appraisals of novelty (e.g., surprise) or intrinsic valence (e.g., distress) (but see Lazarus, 1991, pp. 146–147). Reisenzein (2012), a causal belief-desire theorist, also leaves room for epistemic emotions (e.g., surprise), which result from matches and mismatches with expectations instead of goals. This is also in line with predictive processing accounts of affect (e.g., Van de Cruys, 2017; see Moors et al., 2021). Berkowitz and Harmon-Jones (2004), who are self-declared critics of appraisal theory, propose that negative intrinsic valence (e.g., pain) is sufficient for anger and that neither goal incongruence, external agency, nor blame are necessary.

71 Other criteria for demarcation proposed by appraisal theorists are the control precedence of the action tendency (Frijda, 1986) and synchronization among the various components (Scherer, 2000). I listed both properties in the working definition of emotion. As such, I treat them as part of the explanandum rather than as part of the explanans. I believe the "goal relevance" criterion goes a long way in accounting for control precedence and synchronization. Stimuli appraised as more goal-relevant may elicit action tendencies with more control precedence and lead to a higher level of synchronization among all components (Moors, 2017a).

between danger and the tendency to avoid, for instance, is specific to the emotion of fear. This criterion can be called the "[S–R]" criterion.

Finally, all theories that include other components in the emotion – either in addition to or instead of stimulus evaluation – dispose of a fifth criterion, namely the "component" criterion. In the next section, I examine how these five criteria fare in accounting for the desiderata from the working definition.

6.3.1.2 Adequacy

APPARENT-SIMILARITY

The stimulus evaluation criterion was specifically devised to take care of the world-directed Intentionality of emotions. Stimulus evaluation processes generate the formal objects of emotions, whether these come in molar (e.g., danger for fear, offense for anger, loss for sadness) or molecular form (i.e., specific appraisal patterns for specific emotions). In this respect, SETs outperform any of the other (non-hybrid) theories discussed in the previous chapters.

We saw that in purely molar SETs, the content criterion takes the form of an evaluative property criterion. By treating evaluative properties as the formal objects of emotions, these theories are able to make sense of theoretically irrational emotions. An emotion is inaccurate if its particular object does not represent or instantiate the right formal object, as when a person fears a bug that is not dangerous, or is angry about a blue joke that is not offensive (Kenny, 1963). As mentioned in Chapter 2, theoretical irrationality can trickle down to practical irrationality. Fearing a harmless bug (i.e., inaccurate) may lead to avoidance that is unnecessary and can even be costly, say, if it prevents a person from going outside (i.e., maladaptive). As also mentioned in that chapter, however, not all cases of practical irrationality can be explained via theoretical irrationality. Examples mentioned were costly aggression (e.g., Sanfey et al., 2003) and arational actions (e.g., Hursthouse, 1991). I return to this issue below.

In molecular theories, the content criterion takes the shape of a goal-relevance criterion. This criterion can account for the intensity of emotions. Stimuli with a higher impact on a goal elicit stronger emotions. It can also account for control precedence. The higher the impact on one goal, the more likely this goal will be to take over the show. Note that Frijda (1986, 2007b) supplemented the goal-relevance criterion with an urgency criterion (based on an appraisal variable of urgency). This criterion can probably account for other aspects of the quantitative profile of emotions such as their quick onset and steep rise in intensity.

While goal relevance and urgency take care of the quantitative aspect of the heat of emotions, the qualitative aspect of heat, namely valence, is

https://doi.org/10.1017/9781107588882.010 Published online by Cambridge University Press

covered by the appraisal of goal congruence, which I presented as inseparable from goal relevance. Goal-in/congruent stimuli yield negative/positive valence (see e.g., Moors et al., 2021).[72]

Judgmental theories portray emotions as judgments, which are doxastic representations. This proposal has encountered a long list of objections. The first complaint is that judgments do a poor job in accounting for the continuity across developmental stages and species. Newborns and non-human animals, who lack the capacity to form propositions and to assent to their truth, would be exempted from having emotions. A second objection is that judgments are overintellectual. They lack bodily stirrings and intensity, which is why they cannot do justice to the phenomenology of emotions. One can judge a stimulus, say, climate change, to be dangerous, without experiencing a single bout of fear. A third, related objection is that judgments do not explain how emotions motivate to action. Again, the mere apprehension of the danger of climate change may not suffice to mobilize people into pro-environmental actions. A fourth objection is that judgmental theories have difficulty accounting for recalcitrant emotions, a particular type of theoretically irrational emotions that exist or persist despite our better knowledge. To make sense of these emotions, these theories would have to argue that contradictory judgments can coexist, say, that a house spider is both dangerous and not dangerous (Nussbaum, 2001). Although it is possible that people simply tolerate ambiguity (Tymula et al., 2012), they may also deceive themselves, which involves suppressing one belief at the expense of another (see Griffiths & Scarantino, 2005). Many philosophers consider these solutions to be unsatisfactory, however.

Quasi-judgmental theories circumvent this problem, for it does not seem anomalous to believe one thing (e.g., that the spider is not dangerous) and to construe another (e.g., that the spider is dangerous). But quasi-judgmental theories are still not able to explain ontogenetic and phylogenetic continuity and they fare even worse than judgmental theories regarding some of the other desiderata. If emotions are mere construals, they lose part of their connection with the world, and in the wake of this, their assessability in terms of accuracy (Tappolet, 2016). Emotions can be false but construals cannot. We can judge a person's fear of a house spider to be absurd, but not a person's casual construal of a spider as dangerous. As mere construals, emotions also lose their connection with action. "If you construe a cloud as a horse, you are not likely to be tempted to ride on it" (Tappolet, 2016, p. 15; but see Bayne & Hattiangadi, 2013; Ichino, 2019; Velleman, 2000).

[72] Note that for all theories that consider molar values to be summaries of molecular values (i.e., an identity relation), all desiderata fulfilled by molar values should also be fulfilled by the corresponding molecular values and vice versa.

Perceptual theories were designed to preserve the benefits of judgmental theories while solving their problems. They do so by replacing the conceptual representations of the previous theories with perceptual ones, and by adopting other properties of System 1, such as automaticity, associative operations, and modularity. This results in the following four selling points. First, perceptual representations guarantee Intentionality in the world-directed sense. Second, the rich, salient phenomenology of conscious perception goes unquestioned: There is something it is like to have a perception. Third, because these representations are not conceptual and hence do not require the possession of concepts, emotions are also within the reach of infants and non-human animals. Finally, perceptual theories can make sense of recalcitrant emotions without losing the connection with accuracy and motivation. There is no incompatibility between seeing a house spider as dangerous and not believing it for a fact. This is analogous to visual illusions such as seeing that a stick held in water is bent when knowing that it is not. Seeing a house spider as dangerous can be judged to be wrong in the same way that seeing a stick as being bent can (but see Dokic & Lemaire, 2013; Helm, 2015). In addition, the compelling nature of perceptions can urge us to act, although this is by no means obligatory. Seeing a spider as dangerous may be enough to avoid touching it. The idea that action is not a necessary consequence of perception is underscored by the existence of purely contemplative emotions, such as emotions caused by fictional events, which remain behaviorally inert. The perception of the shark in the film *Jaws* as dangerous does not make us run out of the movie theater (Tappolet, 2010, 2016).

Despite the rosy picture of perceptual theories sketched so far, several objections have been raised (see Brady, 2007; Deonna & Teroni, 2012; Helm, 2001, 2015; Dokic & Lemaire, 2013; Salmela, 2011). I will highlight just a few. First of all, a number of disanalogies between normal perceptions and emotions have been noted such as that (a) normal perception is equipped with specific sensory channels whereas there are no sense organs for emotions; (b) normal perception has direct access to the world whereas emotion requires the intermediary step of non-affective perception (to perceive a spider as dangerous one has to first perceive the spider as a spider); and (c) the phenomenology of normal perception may be salient and rich, but it lacks the characteristic heat of emotions related to valence and bodily intensity (Deonna & Teroni, 2012). In response to these allegations, some perceptualists have argued that a less than perfect analogy does not discredit perceptual theories but only forces them to draw the analogy as more liberal than literal (Döring & Lutz, 2015). Other perceptualists have replied that some disanalogies are overrated. For instance, not all normal perception requires a dedicated sensory channel, as is illustrated by proprioception and time perception (Tappolet, 2016).

Appraisal theories in psychology are dual-system theories. This enables them to account for the same desiderata as judgmental and perceptual theories, while being unhindered by restrictions of a single System 2 or a single System 1. Moreover, appraisal theories do not see any harm in assuming that abstract appraisals can be automatic. They make this palatable by suggesting that abstract content does not have to take a verbal format and that learned or even innate stimulus–appraisal links can be activated via an associative operation. It deserves mention that some judgmental theorists have also gone down this road (e.g., Nussbaum, 2004; Solomon, 1993).

Zooming out, the birth of quasi-judgmental theories and perceptual theories reflects an alternate-components strategy (see Chapter 2) in which the cognitive component is reshaped from a judgment to a quasi-judgment or a perception. Appraisal theories, and the judgmental theories that have followed their trail, have been accused of using an elastic strategy, stretching the notion of appraisal or judgment to the point where it snaps (e.g., Scarantino, 2010; Scarantino & de Sousa, 2018). As I have argued elsewhere, however, appraisal is defined in terms of the types of information that it handles (which is not overinclusive) so that narrow commitments about format, operations, or operating conditions are not required (see Moors, 2013b). This solution may not fly for judgments, however.

Several appraisal theories and judgmental theories have also adopted an add-on strategy, which has turned them into mixed or multi-componential SETs. By planting components inside the emotion alongside stimulus evaluation, these theories have been able to account for some of the desiderata that pure judgmental theories could not. Indeed, SETs that include response and feeling components in the emotion can account for the bodily and phenomenal aspects of emotion, respectively. Theories that include action tendencies, moreover, can also account for the self-directed Intentionality of emotion and its motivating role in action. Note that the add-on strategy brings on board the component criterion.

The [S–R] criterion, finally, states that the transition from stimulus evaluation to action tendency happens via an innate [S–R] link or a learned equivalent of such a link. This criterion is available to causal-only and constituent-causal SETs (but seems to be also endorsed by constituent-only SETs, who take such links to be responsible for the influence of emotion on behavior; e.g., Tappolet, 2016). If the [S–R] links are innate, they fulfill the desiderata of ontogenetic and phylogenetic continuity. Both innate and learned [S–R] links, moreover, contribute to the automaticity of emotions.

Above all, the fact that [S–R] links are stimulus-driven processes that do not take current behavioral outcomes into account explains why emotions are practically irrational in the process-sense. In fact, it is the

[S–R] criterion in SETs (and all theories discussed until now) that accounts for cases of practically irrational emotions that are not theoretically irrational, such as those marked by costly aggression and arational actions. The reason why costly aggression and slamming the door out of anger are practically irrational is that they are caused by a stimulus-driven process that is blind to the consequences of these behaviors.

As always, practical irrationality in the process-sense has to be distinguished from practical irrationality in the output-sense. Any [S–R] link can still yield adaptive behavior, by accident, if the pre-given behavior happens to be also the most optimal one in the given circumstances. Thus, fleeing is adaptive for the threat posed by a fast-approaching car but not for the threat posed by a lack of time to catch one's flight. Similarly, fighting an opponent who is easy to control is adaptive if control is best exerted by fighting, such as when the opponent is weaker and there is no way to escape. However, it is maladaptive if control is best exerted by fleeing, such as when the opponent is stronger and escape is the only option. Ellsworth (2013, p. 127) argued that "[a]ppraisal theories can get us to the right branch of the emotional tree, but not to the right twig; to the neighborhood, but not to the street address." Appraisal tells you whether to avoid or to attack but not which concrete behaviors you should select to reach these abstract aims. This requires an additional goal-directed planning process that is highly sensitive to the specifics of the context. However, as the examples cited above suggest, appraisal can also get us to the wrong branch or the wrong neighborhood. If the threat of missing a flight first leads to the tendency to avoid, it does steer us to the wrong neighborhood.

FRUITFULNESS

The criteria of stimulus evaluation, evaluative property, and system endorsed by purely molar theories do not seem to generate unity within the set of emotions. Each emotion is characterized by a different evaluative property. Thus, generalizations cannot be made from one emotion to the next, or at least not more easily than from one emotion to a non-emotional evaluative phenomenon. Discoveries about fear (an emotion characterized by the evaluative property of danger) do not generalize more to sadness (an emotion characterized by the evaluative property of loss) than to trust (a non-emotional phenomenon characterized by the evaluative property of honesty). Adding system restrictions, such as the restriction in perceptual theories that the evaluative properties have to be generated by System 1, does not alleviate this problem.

Replacing the evaluative property criterion by the goal relevance criterion endorsed by molecular theories seems to make some progress in this regard in the sense that goal relevance is supposed to characterize all (or most) emotions. But how fruitful is it really to group episodes in which stimuli are appraised as goal relevant? A first thing to note is that goal relevance is a gradual notion that does not provide a sharp dividing line between emotional and non-emotional episodes. One might argue that fruitfulness is still possible in the weaker sense of allowing the classification of episodes as more or less emotional. A second worry is that the specific content of representations, here goal relevance, is not deep enough for grounding a scientific set. I am inclined to argue that the *abstract content* of a representation is deep enough for this, but not the *concrete content* of a representation. To clarify, it probably matters for future predictions whether we are dealing with stimulus or response representations, but perhaps not whether the content of a stimulus representation is an apple or a cherry. Just as it is plausible to assume that similar law-like regularities apply to thoughts of apples and cherries, it is plausible to assume that similar law-like regularities apply to evaluations of stimuli as more or less goal relevant. The action tendencies flowing from both types of evaluations may differ in intensity and control precedence, and in the extent to which they mobilize all organismic subsystems, but they may not differ in quality. In sum, the goal relevance criterion may provide a good approximation of how laypeople would rank episodes from more to less emotional, but it may not buy the set of emotions scientific status.

What about the [S–R] criterion? Adding this criterion helps to demarcate emotions from non-emotional phenomena that rely on goal-directed processes such as emotion regulation and instrumental behavior. But it does not allow us to demarcate them from non-emotional phenomena that are also supposed to be subserved by [S–R] links such as mundane instincts or reflexes and habits. In these cases, combining the [S–R] criterion with the criteria of stimulus evaluation and goal relevance may bring some relief as it could be argued that reflexes and habits are triggered by purely perceptual stimulus features that are not evaluated in terms of their goal relevance. For instance, the eyeblink reflex is triggered by an air puff and the habit of getting out of bed by the ringing of the alarm clock (but for a different view, see Moors et al., 2017; Moors & Boddez, 2017; see also Kruglanski & Szumowska, 2020; Chapter 7).

6.3.2 Divisio Definition

6.3.2.1 Criteria for Partitioning

Purely molar SETs favor a discrete view. Each evaluative property or molar appraisal value accords with one specific emotion type: danger with fear, offense with anger, loss with sadness, and so on. Purely molecular SETs, namely elemental appraisal theories, champion a dimensional view instead. In these theories, each molecular appraisal value has a separate downstream influence on the remaining components. The sum of changes in all components determines the quality and intensity of the emergent emotion. Add to this the fact that most appraisal variables are dimensional (i.e., stimuli are evaluated as more or less goal relevant, more or less goal congruent, more or less certain, more or less expected, and more or less controllable) and it is not difficult to see that this results in an infinite number of emotions. These are most easily organized in a multidimensional space in which the dimensions accord with appraisal variables and in which no emotion is more basic or fundamental than any other. Molar-molecular SETs, finally, although in principle compatible with both views, seem to favor a discrete view. Indeed, the molecular appraisals in biological appraisal theories, for instance, culminate in molar appraisals that fix discrete emotions.

SETs naturally account for individual and group differences (Roseman, 2011; Roseman & Smith, 2001; C. A. Smith & Pope, 1992). The lack of one-to-one relations between specific stimuli and specific emotions is explained by stating that the same stimulus is evaluated in different ways by different individuals and by the same individual on different occasions. In appraisal theories, inter-individual differences in evaluations are explained by stable person factors (i.e., traits, determined by learning and genetic predispositions). People differ in the goals they value, the expectations and control they have, and the attributions they make (Scherer & Brosch, 2009). Intra-individual differences in evaluations, on the other hand, are explained by context factors that influence variable person factors (i.e., states). A single person may vary in the goals she values, the expectations and control she has, and the causal attributions she makes across contexts.

Biological appraisal theories have long claimed that while the relation between stimuli and emotions is variable (because it is mediated by appraisal, which is influenced by person factors), the relation between appraisals and emotions is fixed (Roseman & Smith, 2001; Siemer et al., 2007): The same appraisals cause the same emotions; different appraisals cause different emotions. This fixity hypothesis cannot be maintained by elemental appraisal theories, however, at least not if they accept the plausible option that person factors not only influence appraisals but can also have a direct (non-appraisal-mediated) influence on the other components. For example, people may have different thresholds for the translation of particular action tendencies into overt actions, the firing of certain facial muscles, and the discharge of certain physiological responses (Frijda, 2009; Kuppens, 2013; Kuppens et al., 2007). This will translate in different thresholds for the feelings that feed on these components. It will also influence the quality of the emotion as a whole, understood as the emergent sum of all the components. It is not clear, however, whether all elemental appraisal theorists are ready to throw the fixity hypothesis completely overboard. Scherer and Ellsworth (2013), for instance, accepted a radical form of flexibility except on one count: They insisted on keeping a fixed link between appraisals and feelings. In their words: "although we allow for enormous variability in appraisals and in emotions, we do not allow the possibility that the exact same combination of appraisals can produce different feelings in different people or circumstances. If that were the case we might as well pack up and go home" (p. 191).

It should be clear that appraisal theories are openly interactionist (see Chapter 2). Appraisals reflect the way in which the world relates to the mind. Several judgmental and perceptual theories in philosophy, on the other hand, are more externalist. They seem stuck on the idea that evaluative properties are to be found in the world, thereby minimizing their subjective character. "People tend to feel happy when they are showered with attention, tend to experience fear when they are attacked from every direction, experience grief when dear ones die" (Deonna, 2006, p. 36). The externalist inclination of these theories is underscored by their concern for the theoretical rationality of emotions. People may differ in the emotions they have when facing the same events, but not each of these emotions is equally accurate. The externalist position has not gone uncriticized, however, even within philosophical ranks (Döring & Lutz, 2015; Lemaire, 2012, 2014; Lépine, 2016; Salmela, 2011).

6.3.2.2 Adequacy

APPARENT-SIMILARITY AND FRUITFULNESS

SETs with a discrete view are designed to vindicate vernacular emotion types. To this end, they postulate that each vernacular emotion has its own unique evaluative property (in purely molar SETs) or appraisal pattern (in molar-molecular SETs). This enables them to meet the apparent-similarity meta-criterion and at the same time to guarantee homogeneity within each emotion subset, preparing the grounds for fruitfulness. But the proof of the pudding is in the eating. It is up to empirical research to establish whether the hypothesized evaluative properties and appraisal patterns do indeed produce discrete emotions (see Section 6.4).

SETs with a dimensional view are skeptical about the scientific status of vernacular emotion types (see Moors, 2017a). They argue that molecular appraisal values combine into an infinite galaxy of emotions in which no star shines brighter than any other. Nevertheless, the question can be asked how the latter theories explain that some emotion types do carry a label in natural language while others do not. Scherer (1984, 1994b, 2001a), for instance, suggested that the emotions that fit best with the prototypical profiles of vernacular emotions may be more frequent than others, and called them "modal" emotions. In DS terminology, these modal emotions correspond to attractor pools in a multidimensional space. Frequency of occurrence adds contours to this space even if it does not count as a criterion for basicness. In contrast, Russell and Barrett (1999; Russell, 2005) suggested that prototypical profiles are not more frequent but rather more conspicuous and therefore more easily recognized and labeled (see Chapter 8). Barrett (2017a, 2017b; Hoemann et al., 2019) later argued that prototypical emotions are mere abstractions that need not even exist in nature. Putting aside the issue of modal emotions, the fruitfulness of partitioning the set of emotions with the help of appraisal dimensions hinges on whether the chosen dimensions do indeed cover the variety in emotions. This is again an empirical question.

6.4 Validation

I discuss five lines of empirical research that have addressed the hypotheses of SETs, and in particular those of appraisal theories.

(1) A first line of research examines the hypothesis of early appraisal theories (Lazarus, 1991; see Moors, 2013b) that appraisal is a necessary cause of emotions. Demonstrating that something is a necessary cause for an effect is not a realistic aim of research.[73] The best one can do is maintain that it is a necessary cause until proven otherwise. Several researchers have in fact tried to prove otherwise. The empirical arguments against the necessity of appraisal for emotion are collectively referred to as evidence for the "affective primacy" hypothesis, which is contrasted with the "cognitive primacy" hypothesis. However, not all of these arguments have been welcomed as convincing or relevant. A few examples are illuminating (for extensive discussions see the debate between Lazarus, 1982, 1984, 1991, 1999, and Zajonc, 1980, 1984; and later reviews by Clore & Ketelaar, 1997; Clore et al., 2005; Lähteenmäki et al., 2015; Leventhal & Scherer, 1987; Moors, 2007; Storbeck & Clore, 2007; Storbeck et al., 2006).

[73] If a cause is understood as an Insufficient but Necessary part of a set of conditions that are together Unnecessary but Sufficient to produce the effect (i.e., INUS condition; Mackie, 1974), then demonstrating that appraisal is not only a cause but a necessary cause of emotion would require examining all possible sets of conditions-that-are-together-sufficient-for-emotion and showing that appraisal is a necessary condition in each of these sets (see Moors 2013b).

https://doi.org/10.1017/9781107588882.010 Published online by Cambridge University Press

A first piece of evidence for affective primacy relies on behavioral methods. Kunst-Wilson and Zajonc (1980) found that mere exposure to stimuli that are not consciously perceived led to an increase in liking of those stimuli. The authors took this result as evidence for affect (understood as an incipient form of emotion) without cognition. Appraisal theorists, on the other hand, took it as evidence for unconscious appraisal (i.e., an evaluative type of cognition) without conscious identification (i.e., a non-evaluative type of cognition). S. T. Murphy and Zajonc (1993) further showed that the valence of prime stimuli influenced the evaluation of subsequent neutral target stimuli when primes were presented subliminally but not supraliminally whereas non-valenced features of the prime stimuli such as size, symmetry, and gender influenced the judgment of the target stimuli (in terms of these features) only when the primes were presented supraliminally but not subliminally. The authors took this result as evidence that affect precedes cognition. Appraisal theorists, on the other hand, took it as evidence that appraisal (i.e., an evaluative type of cognition) precedes the processing of non-valenced features (i.e., non-evaluative types of cognition).

A series of recent studies also called the effects themselves into question. Lähteenmäki et al. (2015) found evidence that the semantic categorization of valenced stimuli (animals, facial expressions) (i.e., a non-evaluative type of cognition) preceded their affective categorization (i.e., an evaluative type of cognition) (see also Storbeck & Clore, 2007). But whether or not appraisal is preceded by non-evaluative types of cognition is in fact not relevant for appraisal theories. Their claim is that appraisal precedes emotion, not that non-evaluative cognition precedes appraisal or emotion.

Another piece of evidence for affective primacy, already discussed in Chapter 5, is evidence that physical stimuli can directly cause emotions. This research aims to show that people experience emotions when taking drugs (de Montigny, 1989), when putting on a happy or sad face (Strack et al., 1988; but see Wagenmakers et al., 2016), and when engaging in coarse behavior (Cacioppo et al., 1993). Leaving the empirical status of some of this evidence to the side, evidence that physical stimuli can influence emotions is not in itself evidence that this influence is direct rather than mediated by appraisal (Arnold, 1960, p. 111; Clore & Centerbar, 2004; Clore et al., 2005; Moors, 2007).

A further piece of evidence for affective primacy relies on neuroanatomical methods. LeDoux et al. (1989) famously demonstrated that when the visual cortex of rats was lesioned, learning that a light predicted a shock remained intact, as they still showed fear-related behaviors to the

https://doi.org/10.1017/9781107588882.010 Published online by Cambridge University Press

light. This was taken as evidence for the existence of a fast route of activation from the hypothalamus to the amygdala, which is separate from a slow route of activation from the hypothalamus to the amygdala mediated by the cortex (LeDoux, 1996). Based on the assumption that the amygdala is the center of fear and the cortex is charged with cognitive processing, numerous theorists have interpreted this finding as the ultimate piece of evidence that emotion can be elicited without cognition. Again, appraisal theorists have argued that the amygdala can be considered as the locus of a primitive form of appraisal (Scherer, 1993a, p. 12), a possibility that was also contemplated by LeDoux (1989, p. 271). In addition, Storbeck et al. (2006) made the sobering observation that evidence for the fast route was obtained only when the conditioned stimuli required only a minimal amount of sensory preprocessing such as light vs. dark environments. Most stimuli that elicit emotions in humans and other animals are parasitic on more extensive semantic preprocessing and hence travel via the slow route. This was supported by the lack of evidence for a functionally independent fast route even in primates (Pessoa & Adolphs, 2010).

A final piece of evidence for emotions without appraisal is evidence that direct brain stimulation in the absence of an eliciting stimulus can trigger emotional responses and feelings (see reviews by Izard, 2007; Panksepp, 2007; but see Barrett, Lindquist, Bliss-Moreau, et al., 2007). Such evidence (if it were to exist) would indeed qualify as evidence for emotions that are not caused by appraisal, but it would not falsify the mechanism for emotion causation put forward by appraisal theories. Appraisal is considered to be a first step in the causal chain from stimulus to emotion, which is itself followed by the further step of the activation of an innate or learned [S–R] link (or set of those links). Appraisal can thus be considered as the remote cause of emotion and activation of the [S–R] as the proximal cause. If brain stimulation directly activates the proximal cause, this would not invalidate the claim that in normal circumstances the remote cause is necessary to activate the proximal cause. Nevertheless, arguments like this have persuaded some appraisal theorists that appraisal may not be a necessary but only a typical cause of emotions.

(2) A second line of research examines appraisal theories' hypothesis that appraisal can be automatic (Arnold, 1960; Lazarus, 2001; Oatley & Johnson-Laird, 1987). The automatic nature of appraisal is a precondition for maintaining that appraisal is a necessary or typical cause of emotions given that many of our emotions arise automatically. Research on the automaticity of appraisal branches out in two paths (Moors, 2010b).

A first path systematically examines the extent to which individual appraisal variables can be processed automatically. Such evidence has been obtained for the molecular appraisals of goal relevance (Mogg et al., 1995; Vogt et al., 2013; J. M. G. Williams et al., 1996), positive/negative valence (Fazio et al., 1986; Hermans et al., 2003), goal in/congruence (Moors & De Houwer, 2001; Moors et al., 2005), un/expectedness and novelty/familiarity (Berns et al., 1997; Gati & Ben-Shakhar, 1990; Oakes & Turner, 1986), dominance/submissiveness (which is closely related to difficulty/ease of control) (Moors & De Houwer, 2005), and agency or blame (Dik & Aarts, 2007; Hafri & Firestone, 2021; Hassin et al., 2005; Scholl & Tremoulet, 2000). Other researchers have tried to show that negativity (Cacioppo et al., 1999) or the molar appraisal of threat or danger (Öhman & Mineka, 2001; Wentura et al., 2000) have priority in the processing sequence.

A second path centers on the question of what level of vertical complexity can be handled under conditions of automaticity (Moors, 2010a, 2010b; see Chapter 1). Prior research suggested that appraisal can only be automatic if it is based on preset (innate or learned) evaluations, which require only the stimulus as input (i.e., single-input process) (Duckworth et al., 2002; see also Frijda, 2007b). Later research challenged this idea by showing that the processing of goal in/congruence, based on the online comparison between stimuli and goals (i.e., multiple-input process), can also be automatic (Moors et al., 2004). This finding dovetails with evidence for the malleability of automatic stimulus evaluation (Blair, 2002) as well as evidence that positive/negative evaluation only occurs for stimuli that are goal-relevant (Spruyt et al., 2018; Storbeck & Robinson, 2004; Winkielman et al., 2005, p. 132; see also Charland, 2005a).

(3) A third line of research tests hypotheses about the influence of appraisal on emotions or components. A first approach that fits best with appraisal theories with a discrete view examines the influence of appraisal on discrete emotions. A second approach that fits best with appraisal theories with a dimensional view examines the influence of appraisal on other emotional components, without linking these back to discrete emotions. The first approach faces a problem that was also mentioned in the context of the evolutionary theory: that research on distinctness collapses with research on concordance. To examine whether specific appraisals cause specific emotions, one should manipulate appraisals and measure emotions. Manipulating appraisals is relatively easy (even if it has to happen in an indirect way). Measuring emotions is more difficult, however, because it can only be done by measuring one or more emotional components. Eventually then, the first approach

https://doi.org/10.1017/9781107588882.010 Published online by Cambridge University Press

collapses with the second approach in that researchers have no choice but to study the influence of appraisal on other components. Take a typical study in the first approach in which participants are asked to recall instances of emotions and to rate these in terms of appraisal variables (Scherer, 1993b; Scherer & Meuleman, 2013; C. A. Smith & Ellsworth, 1985). As emotion words coincide with feeling words, this research could also be framed as research studying the relation between appraisal and the feeling component, which would fit in the second approach. The second approach also studies the relation between appraisals and (a) action tendencies or coarse behavior (e.g., Frijda et al., 1989), (b) somatic responses (Kreibig, 2010; Kreibig et al., 2012; Pecchinenda, 2001), and (c) subtle behavior in the facial channel (Mehu & Scherer, 2015; Scherer & Ellgring, 2007; Scherer et al., 2017, 2018), the vocal channel (Goudbeek & Scherer, 2010; Johnstone et al., 2001, 2005; Laukka & Elfenbein, 2012), and the gestural channel (Dael et al., 2013) (see review by Scherer & Moors, 2019). I illustrate the second approach by considering two appraisal hypotheses in more detail.

A first hypothesis concerns the influence of appraisal on feelings: *Goal-incongruent stimuli that are easy/difficult to control lead to feelings of anger/fear* (e.g., Ellsworth & Scherer, 2003). In some studies, participants receive the instruction to remember or imagine a goal-incongruent situation that was easy/difficult to control and to rate the extent to which they experienced anger/fear. In other studies, participants are instructed to remember an episode in which they felt anger/fear and to rate the extent to which they appraised the stimulus as goal-incongruent and easy/difficult to control. In the latter study, appraisals are not manipulated but measured and correlations are calculated between appraisals and feelings (hence not allowing for causal conclusions). A study by Scherer (1997a), for instance, confirmed that fear was more closely associated with difficult to control stimuli than anger. In still other studies, appraisals are manipulated using real stimuli that are expected to be appraised in a certain way. For instance, participants in a study by Wiech et al. (2006) received painful electric shocks. In a self-control condition, participants could decide when the stimulation would end. In an external-control condition, participants were led to believe that this was determined by another participant (to which they were yoked) or by a computer program. In reality, they received the same stimulation as they had chosen in the self-control condition. Participants reported less feelings of pain and fear in the self-control condition in which the goal-incongruent stimuli were easy to control than in the external-control condition in which these stimuli were impossible to control despite being exposed to the same duration and intensity of shocks.

A second hypothesis concerns the influence of appraisal on action tendencies: *Stimuli that are easy/difficult to control lead to the tendency to fight/flee*. Correlational studies in which emotions and action tendencies were measured via self-report have provided support for this hypothesis (Frijda et al., 1989). Such support was not obtained, however, in experimental studies in which appraisals were manipulated in a computer game and action tendencies were measured via (a) implicit behavioral methods (Fischer et al., 2022) and (b) MEPs enhanced by TMS (Moors et al., 2019; see Box 2.2). The behavioral study had the following procedure: Participants were cast in the role of a street musician, who collected money in a hat lying in front of them. On some trials, a thief would pass by and steal money (i.e., goal-incongruent stimulus). One thief was difficult to control in that nothing could be done to prevent a theft. Another thief was easy to control in that a theft could be prevented either by fleeing or fighting. On each trial, participants were instructed to either flee or fight. Based on appraisal theories' hypothesis, we predicted that participants would be faster to execute the instruction to flee from thieves that were difficult to control and to fight against thieves that would be easy to control. These predictions were not confirmed, however, as participants were faster to fight against thieves that were difficult to control (see also Donnerstein & Wilson, 1976; Richardson et al., 1986). A similar lack of support for this hypothesis was observed when action tendencies were measured with MEPs enhanced by TMS (Moors et al., 2019). The latter study will be discussed in Chapter 7 because it pits the predictions of appraisal theories against those of the goal-directed theory.

(4) A fourth line of research tries to understand the organization of the brain in terms of molecular appraisals. Instead of assuming that there are dedicated affect programs for fear, anger, sadness, and so on, appraisal theorists search for signatures for goal relevance, goal in/congruence, un/certainty, un/expectedness, ease/difficulty to control, and so on (see Sander et al., 2018). For instance, Sander et al. (2003; Brosch & Sander, 2013) provided evidence that the amygdala is not a fear hub (see Öhman & Mineka, 2001), but rather codes for the goal relevance of stimuli (see also Chapter 4). Sander et al. (2018) also reviewed research on the brain basis of the overarching components of appraisal, action tendencies, somatic responses, motor responses, and feelings.

(5) A fifth line of research examines the influence of emotions on the cognitive functions of perception, attention, memory, judgments, and decision making while specifically referring to the appraisals underlying these emotions as the driving force behind these influences (Brosch et al., 2013; Lerner et al., 2015; Lerner & Keltner, 2001; Lerner & Tiedens, 2006;

Liu et al., 2019; Montagrin et al., 2013, 2018; Pool et al., 2016; So et al., 2015).[74] I briefly illustrate this research line in the domain of decision making. The carry-over effect of incidental emotions (which are unrelated to the decision at hand; see Chapter 2) has been explained as the carry-over of a tendency to appraise the stimulus in a certain way. According to this logic, Raghunathan and Pham's (1999) finding that fear regarding the result of a medical exam led to less high-risk high-reward choices in a subsequent, unrelated decision task can be explained as the carry-over of appraisals of uncertainty and/or uncontrollability.

6.5 Embodied SETs

Two problems that haunt all SETs belonging to the evaluation-first brand (discussed so far) are worth emphasizing as they form the bridge to the embodied brand of SETs (to be discussed now). The first problem is that the stimulus evaluation process produces representations with abstract content, such as danger and offense, and that not all organisms to which we ascribe emotions have the capacity to represent such content. Let me call this the "abstract content" problem. This problem is most obvious for judgmental theories. The solution provided by perceptual theories to replace the conceptual representations with perceptual ones is unsatisfactory because it seems to involve a contradiction. How can abstract features ([eaS]) be represented by perceptual representations ([ecS]), given that the latter are defined as having concrete features? Dokic and Lemaire (2013) stated the problem as follows. Given that perception can only yield direct access to the particular object (i.e., concrete stimulus features), to arrive at the formal object (i.e., abstract evaluative stimulus features) requires enrichment from other, non-perceptual sources (but see Tappolet, 2016, 2020, for a reply).

The second problem that has puzzled some theorists is how one can go from a cold evaluation to a hot emotion (e.g., Dewey, 1895). Let me call this the "cold-hot transition" problem. Most theorists do not cast this as a problem at the subpersonal level of processing, however. At the subpersonal level, [S–R] links connecting afferent representations to efferent representations may simply have been wired in or learned. The problem is more a problem on the personal level, the level of experience. If evaluation of say, danger, can be done in cold blood, at what point does it ignite with the spark of emotional heat?

[74] It may be noted that this research no longer pertains to the first demarcation-explanation cycle but already belongs to a subsequent cycle in which the explanandum is the influence of emotion on cognitive functions.

https://doi.org/10.1017/9781107588882.010 Published online by Cambridge University Press

(1) A first solution to both problems is provided by J. J. Prinz's (2004a) neo-Jamesian theory, also known as the embodied appraisal theory. This theory relies on the teleosemantic theory of representation (Dretske, 1995), which requires that in order for A to represent B, A must reliably covary with B and A must have the function to represent B. Such a function can be set up by evolution, learning, or design (J. J. Prinz & Barsalou, 2000, p. 55). Thus, a fire alarm can represent fire because the fire alarm reliably covaries with fire and the fire alarm was set up to signal fire (in this case by design). Equipped with this theory of representation, the solution provided to the abstract content problem is that any concrete representation can represent abstract features provided that representation and features reliably covary and that the former has the function to represent the latter. J. J. Prinz (2004a) proposed the following mechanistic sequence for the emotional episode: A stimulus, after being perceived ([ecS]), activates an innate [S–R] link (conforming to evolutionary theories), which produces bodily responses (R) or at least mental simulations of such responses. Somatosensory feedback of these responses (iS) is in turn perceived ending up in a concrete representation ([icS]) (conforming to James, 1890b). Now, the concrete content of this representation (e.g., trembling, sweaty palms) signals the abstract meaning of the stimulus ([eaS]; e.g., danger) by virtue of their reliable covariation set up by evolution. In this sequence, stimulus evaluation is not first; it is preceded by bodily responses. Moreover, insofar as bodily representations are supposed to be hot, the theory also dodges the cold-hot transition problem.

(2) A second solution to the problems of abstract content and of cold-hot transition is offered by Deonna and Teroni's (2012, 2015, 2017, 2020) attitudinal theory. This theory makes two moves. First, it shifts the abstract evaluative properties from the *content* of representations to the *Attitude* towards representations. Applied to the case of fear, the theory states that the dangerousness of the stimulus is reflected in the Attitude that the organism adopts towards the representation of the stimulus rather than residing in its content. Deonna and Teroni (2012) proposed that emotions are evaluative Attitudes, with each emotion having its own sui generis Attitudinal mode, next to the Attitudinal modes of belief and desire (so they are not reduced to beliefs or desires). The same particular content can be believed, desired, feared, hoped for, rejected, and so on. The second move that the theory makes is to spell out Attitudes in terms of abstract action tendencies. For instance, fear is characterized by an Attitude of avoidance and anger by an Attitude of antagonism. This is the embodied aspect. These action tendencies, moreover, must be felt to

qualify as emotions. In this theory, the presumed mechanistic sequence goes as follows:[75] An external stimulus (eS) is first processed in non-evaluative terms by a cognitive process, called the cognitive base. The resulting stimulus representation ([eS]), which contains the particular object of the emotion, then activates an action tendency ([R]), presumably via an innate [S–R] link (conforming to evolutionary theory; Deonna & Teroni, 2012, p. 99). This action tendency is felt and forms the Attitude that retroactively gives meaning to the particular object. Note, moreover, that the action tendency is not felt centrally, but has to travel via its manifestation in bodily responses and the perception of somatosensory feedback in order to be felt peripherally (conforming to James, 1890b; for an alternative, see Mitchell, 2020b). So considered, Deonna and Teroni's (2012) theory does not differ much from that of J. J. Prinz (2004a). According to J. J. Prinz (2004a), the bodily feelings directly inform experience about the evaluative properties of the particular object by virtue of covariation. Bodily feelings typical for fear, for instance, signal that the stimulus is dangerous. Deonna and Teroni (2012), on the other hand, insist that bodily feelings first inform experience about an action tendency, which in turn, informs about the evaluative properties of the particular object. Bodily feelings typical for fear, for instance, signal that you want to escape and this in turn signals that the stimulus is dangerous. In any case, stimulus evaluation comes at the end of the mechanistic sequence in Deonna and Teroni's (2012) theory too. And given that this evaluation is based on bodily responses, which are assumed to be hot, this theory similarly sidesteps the cold-hot transition problem.[76]

(3) A third solution to the problems of abstract content and cold-hot transition comes from emotion theories that are molded after a species of embodied theories of cognition called grounded theories (Barsalou, 2008). Grounded theories propose that incoming stimuli do not give rise

[75] Deonna and Teroni (2012, p. 99) are only concerned with the personal level and do not provide a detailed mechanistic account on the subpersonal level. Yet they do mention evolutionary theory's assumption that basic emotions have distinct action tendencies and bodily response patterns.

[76] Dewey (1894, p. 24) foreshadowed several of the features of these embodied theories. For one thing, he argued that a stimulus activates an [S–R] link and that the action tendency ([R]) that is part of this link provides the object of the emotion as well as the value of that object. For another thing, he suggested that the action tendency itself is more than just a centrally located representation. It already innervates peripheral discharge, which provides the action tendency's heat.

https://doi.org/10.1017/9781107588882.010 Published online by Cambridge University Press

to pure stimulus representations, but rather to sensorimotor representations in which concrete stimulus and response features are bound together ([ecS–R]). The response features are often taken to be affordances, that is, the possibilities for action that are allowed by the stimulus (understood in the liberal sense proposed by Scarantino, 2003). For instance, the sight of a hammer not only activates a representation of the shape and size of the hammer, but also that the hammer can be used to knock. Activation of an affordance for action also leads to the partial re-enactment or mental simulation of the action, which can even spill over to overt bodily responses that prepare for these actions. Evidence for grounded cognition comes from studies showing that different types of graspable objects activate simulations of different types of grasping behavior (Ellis & Tucker, 2000; Grèzes et al., 2003).

Applying these ideas to emotions, Griffiths (2004b; see also Griffiths, 2003; Griffiths & Scarantino, 2005) proposed to extend the space of appraisal variables in appraisal theory with affordances. Thus, stimuli may not only be appraised in terms of how goal-relevant, goal-congruent, expected, and controllable they are, but also in terms of whether they afford specific actions such as flight, fight, and approach, that is, whether they are flee-able, fight-able, and approach-able. True enough, the original appraisal variable of controllability already refers to the potential for action, but it does so only in general terms. It captures *whether* there are action options to deal with a goal-incongruent stimulus, but not *which ones*. Affordances, on the other hand, express the specific action options at the organism's disposal. Because Griffiths's (2004b) affordance theory still includes the original "abstract" appraisal variables of appraisal theories (e.g., goal congruence, controllability), however, the abstract content problem is not alleviated. Yet a revised affordance theory that did away with the original appraisal variables and connected concrete stimulus features directly to affordance-related appraisals would circumvent the problems of both abstract content and cold-hot transition.

The next three solutions to these problems are provided by emotion theories that are grounded in a species of embodied theories of cognition called enactivist theories (Varela, 1991). These theories argue that the body is not merely instrumentally or causally related to cognition, but rather constitutively. The body is part of cognition (Hurley, 1998; P. Robbins & Aydede, 2009; but see J. J. Prinz, 2006b). Different scholars have interpreted the implications of enactivism for emotions in somewhat different ways, but they distinguish themselves from the previous

https://doi.org/10.1017/9781107588882.010 Published online by Cambridge University Press

embodied theories in that they take a more radical stance regarding the role of the body in emotions.

(4) The fourth solution is provided by the enactivist theory of Colombetti (2003, 2007, 2009; Colombetti & Thompson, 2007).[77] In Griffiths's (2004b) affordance theory discussed above, stimulus and response features were already bound up in the same sensorimotor representation ([S–R]). Colombetti (2009) pushed this idea further by arguing for the complete fusion of both types of features ([SR]). As discussed in Chapter 5, she made use of the DS framework, in which the mutual interactions and massive recurrence between stimulus and response representations grows so tight that they become indistinguishable. Note that by adopting the ideas of enactivist theories, Colombetti (2009) not only sought to solve the cold-hot transition problem on the personal level, but also on the subpersonal level (as there is no longer a gap between [S] and [R] in [SR]).

Apart from the promise of closing the mind–body gap, I wonder what the fourth solution buys us. Colombetti (2007) argued that previous applications of the DS framework to emotions have not been radical enough in that they still attached different functions to different components or subsystems. It seems to me, however, that the possession of separate functions is a good reason, perhaps the *only* good reason, to distinguish separate entities. And functional modularity does not entail commitment to neural modularity (see also D. Evans, 2008; J. J. Prinz & Barsalou, 2000). Distinct functions cannot be thrown out of the window lightly without at the same time losing any handles for prediction and control. Moreover, Colombetti (2009) does not seem to practice what she preaches as she keeps talking about the functions of evaluation and responding. Her own difficulty to shake off functional thinking may underscore exactly how vital it is.

(5) The fifth solution comes from another application of enactivism to emotion proposed by Shargel and Prinz (2018). Like Griffiths and Scarantino (2005), they put affordances center stage. The place and form of affordances is different, however. Shargel and Prinz (2018) abandoned J. J. Prinz's (2004a) earlier view that emotions are perceptions of bodily changes that indirectly track (and hence represent) the evaluative properties of stimuli. They replaced it with the view that emotions are the bodily

[77] This theory was already discussed in Chapter 5 as an example of a non-biological network theory, but it happens to fit here as well.

changes themselves that prepare for and bring forth the actions afforded by stimuli, but without representing anything. Griffiths and Scarantino (2005; see also Gibson, 1979) took affordances to be trait-dependent properties (i.e., dependent on the stable features of an agent) that are nevertheless state-independent (e.g., independent of the presence of the agent). Thus, a tree is climbable for a squirrel but not for a cow (trait dependence) but it remains climbable for squirrels regardless of whether a squirrel happens to be around to appreciate it (state independence). Shargel and Prinz (2018), by contrast, emphasized that the affordances that give rise to the bodily changes that are emotions are not state-independent stimulus features that lie in wait to be detected by the organism. Instead, they are state-dependent, and more specifically, they are dependent on the emotion that the person is already having. The emotion modifies the space of affordances for the person at a given time. Fear turns fleeing into a more achievable course of action. Anger turns aggression into a more likely strategy to drive one's point home in a discussion, and so forth. However, emotions do more than simply open up an altered space of possibilities for action. In their capacity of being overt bodily responses, emotions also exert a certain push or pull over the organism. Shargel and Prinz (2018) call this the imperatival nature of affordances. For the sake of terminological clarity, however, I think it would be best to reserve the term affordance for indicating possibilities for action and to separate this from the imperative for action, which amounts to an action tendency with control precedence (in line with Frijda, 1986; Scarantino, 2014). An affordance is a cognitive entity with a mind-to-world direction of fit; an action tendency is a conative entity with a world-to-mind direction of fit.

The mutual interactions between emotions (i.e., bodily responses) and affordances/action tendencies (i.e., possibilities/imperatives for actions) proposed by Shargel and Prinz (2018) are reminiscent of the principles of "recurrence" present in the DS framework discussed above (M. D. Lewis, 2005; Scherer, 2000, 2009a). Fear, understood as the bodily preparation for escape, energizes the body and in this way opens up the affordance for, but also pushes towards, escape. The bodily responses preparing for this escape, are again called fear. This is an example of what M. D. Lewis (2005) called a self-amplifying cycle (see Chapter 5). But the question that remains unanswered is this: What sparked the first cycle, if not an appraisal of the stimulus as dangerous or as one that requires escape before the first inkling of bodily mobilization towards escape (and hence fear) was in the air? It may be recalled that M. D. Lewis (2005) proposed recurrent cycles of appraisals and emotions that grew into appraisal-emotion amalgams over time. In Shargel and Prinz's (2018) theory, there is no initial appraisal that gets the ball rolling. And yet the descriptions they provide for emotional episodes suggest otherwise: "A swimmer far from shore begins to struggle with a strong undertow, and realizes that it will take every bit of strength she has to make it back. In virtue of entering a state of fear, her body mobilizes more forcefully to support strenuous action, helping her return safely" (p. 120). It seems justified to infer that the appraisal that a certain action with a certain strength is *needed* (not simply *afforded* or *pushed*) creates the first hint of fear, which is amplified in subsequent cycles. But the authors remain silent on this issue.

(6) The sixth solution, finally, is suggested by the radical enactivist theory of Hutto (2012; Hutto et al., 2018). Like radical enactivist theories of cognition (Hutto & Myin, 2017), Hutto (2012) rejects the idea of representations altogether. Shargel and Prinz (2018) already uttered the slogan that emotions do not *represent*, but rather *bring forth* new meaning. But what Shargel and Prinz (2018) especially wanted to sail against is the traditional notion of representation as the reproduction of pre-given meaning in the mind. Also it is not because emotions do not represent anything that they cannot be caused or partly be constituted by entities that do. Radical enactivists like Hutto (2012), on the other hand, take a radical anti-representational stance. They replace the teleosemantic account of Intentionality that J. J. Prinz (2004a) borrowed from Dretske (1995) with a teleosemiotic account. This account of Intentionality retains the directedness part while removing the representational part. If evolution sets up one entity (e.g., a certain brain state) to covary with another entity (e.g., a certain state in the world), the first entity is directed at the second entity, but the second entity thereby does not represent the first entity. According to Hutto (2012), this form of directedness is all that is required to understand the Intentionality of basic minds and emotions. While Hutto's (2012) theory can certainly be placed on the extreme end of the enactivist scale, it is unclear whether it still assumes a form of stimulus evaluation and hence whether it still qualifies as an SET.

Embodied theories are all the rage these days. Many scholars are attracted to them because they hold the promise of combining some form of Intentionality with bodily phenomenology, while at the same time providing a natural account of how emotions cause behavior. But embodied theories, perhaps because they are recent and have been presented by the authors often as sketches still to be worked out in full detail, have not been held against the same standards as their predecessors. Sure enough, J. J. Prinz's (2004a) embodied appraisal theory has had its share of criticism (e.g., Colombetti, 2014; Deonna & Teroni, 2012; Helm, 2010; Pineda, 2015a, 2015b; Roberts, 2013; Shargel, 2015; Shargel & Prinz, 2018) and suggestions for revisions (e.g., Barlassina & Newen, 2014). Likewise, Deonna and Teroni's (2012) attitudinal theory has received criticism (e.g., Dokic & Lemaire, 2015; Rossi & Tappolet, 2019; J. Smith, 2014) and suggestions for modifications (e.g., Hernández, 2020; Mitchell, 2020a, 2020b; Müller, 2018). The more radical embodied approaches of Colombetti (2007, 2009), Shargel and Prinz (2018), and Hutto (2012), however, seem more difficult to get a handle on. The lack of a detailed mechanistic story in these theories leaves the experimental researcher wondering which unique hypotheses can be distilled and put to empirical test. Colombetti (2007) suggested studies to examine whether bodily arousal (e.g., hormones like testosterone) would influence evaluations of stimuli (e.g., as more hateful or obstructive). In fact, such research already exists (e.g., research on the influence of peripheral feedback on evaluations and feelings; see Chapter 5), but I fail to see how this research can disambiguate between the interpretation of a (bidirectional) causal influence between separate evaluative and bodily entities and the interpretation that bodily entities are glued to or even constitutive of evaluative entities. Similar problems have arisen in research on embodied cognition (Barsalou, 2008). But while some theorists have found ways to reconcile non-embodied views of cognition with embodied views of moderate stripe, highlighting that they focus on different aspects and/or different levels of analysis (see J. J. Prinz & Barsalou, 2000; J. J. Prinz, 2006b), such a reconciliatory exercise seems much harder for embodied views with a more radical slant. If followed to the letter, these radical variants may turn out to be incompatible with a mechanistic approach more broadly speaking. They would ultimately also entail the bankruptcy of the experimental research approach.

Coming in from the cold, I will leave these fundamental debates to the side and I will return to a radically mechanistic approach. In such an approach, the first step is to build a mechanistic sketch with observable inputs and outputs and non-observable representations and operations in the middle. At the very minimum, representations are heuristic tools. They are functional entities, placeholders for certain types of information. How this information is implemented on lower levels of analysis, the exact format that this information is coded in, and how this information is experienced, are all interesting questions that, when they receive an answer, put meat on the bone and may help turn the mechanistic sketch into a full-blown mechanistic explanation (Piccinini & Craver, 2011). But before the meat can be added, the skeleton must be put into place. In the same vein, a mechanistic approach does not shun or ignore complexity in the form of recurrence (i.e., cycles), changes over time (i.e., diachronic

https://doi.org/10.1017/9781107588882.010 Published online by Cambridge University Press

complexity), competing mechanisms (i.e., synchronic complexity), or hierarchical levels of embeddedness (i.e., concentric complexity). But here, the approach is bottom-up, and progress is envisaged from lower to higher levels of complexity.

Against the backdrop of this mechanistic approach, I will now ask how serious the problems of abstract content and cold-hot transition that were raised against evaluation-first SETs and that motivated the emergence of the embodied SETs really are. After that, I will question the tight connection between stimulus and response features ([S] and [R]) that is uncritically assumed in both evaluation-first and embodied SETs, and in fact, in all emotion theories discussed until now.

(1) Evaluation-first SETs have been criticized for their assumption that the stimulus evaluation process produces representations with abstract content (e.g., danger, offense). The principal complaint is that babies and animals may not possess the abstract concept of danger or offense, and hence that they would have to be deprived of emotions. But how damaging is this objection to evaluation-first SETs? Part of the problem stems from a molar approach to stimulus evaluation. Maybe babies and animals do not possess the molar concepts of danger and offense, but it seems useful for any living organism to have some mechanism in place to detect the molecular values of goal relevance and goal in/congruence, that is, the extent to which goals and needs are un/fulfilled (M. Lewis & Ramsay, 2005). If this entails that the emotion repertoire of babies and animals is impoverished, then this only reflects the current consensus (M. Lewis, 2000). It does not show that these beings do not evaluate stimuli, only that they do so in a more primitive way.

Another part of the problem stems from the tacit assumption that the abstract output of stimulus evaluation takes on a verbal-like format. This is an unintended consequence of the fact that we cannot talk about abstract features without using words. Perceptual theorists nevertheless proposed that the output of stimulus evaluation takes on an image-like format. But the problem, stated earlier, is that images may be ill-suited to represent abstract, evaluative features (see also Dokic & Lemaire, 2013). We cannot paint or take a picture of peace in the same way that we can paint or take a picture of an apple. Tappolet (2016, p. 43, footnote 108) replied that animals must perceive opportunities and challenges such as good and bad things to eat. If we assume that animals are equipped with perceptual representations only, what else would these representations have to pick up other than what is functional for them? I agree that organisms should detect opportunities and challenges for goal satisfaction. Representing anything else (making exact copies of the outside world) would not fit in the functional view of representations endorsed here. If teleosemantics works for bodily responses, why would it not work for external stimuli? If an adrenaline rush can indirectly inform us about the danger of a sharp-toothed dog, then why would the sight of such a dog not be able to inform us directly? The reason why embodied theories like J. J. Prinz (2004a) and Deonna and Teroni (2012) forgo that solution is that they believe such a direct evaluation remains cold and detached. This leads them to posit a costly, time-consuming detour via the body.

From a mechanistic perspective, however, talk about the format of representations is of secondary interest. The priority should be to build a plausible mechanistic sketch, in which the type of information processed takes priority over the format in which it is represented. What if the verbal-like and image-like codes that are considered do not exhaust the range of possible formats, for instance? The point I want to make is that problems with format are not a free pass for fabricating solutions that seem far-fetched from a mechanistic point of view. Even if we remind ourselves that the proposals of several embodied theories pertain to the personal rather than the subpersonal level, the question remains why the personal level does not follow on the subpersonal level.

(2) Evaluation-first SETs have also been criticized for introducing the cold-hot transition problem. The complaint was that direct stimulus evaluations, whether coded in conceptual or perceptual format, are cold. One can judge or see that climate change and unprotected sex are dangerous yet feel no fear. Here again, the objection itself may be questioned. Stimulus evaluation is not cold (see also Helm, 2009; Lazarus, 2001, pp. 57–58; Moors, 2007). If heat is understood as valence plus intensity (see Chapter 2), then stimulus evaluation fits the bill. Stimuli are evaluated in terms of their implications for the fate of goals. Since goals are defined as valued outcomes, the fate of these valued outcomes should have value too. Thus, stimulus evaluation yields the valence or sign of emotions. This is captured in the appraisal variable of goal congruence, occasionally called "motivational valence" (Moors, 2010b). In addition to valence, stimulus evaluation can yield the intensity of emotions and other features that are part of the quantitative profile of emotions. This is captured in the appraisal variables of goal relevance and urgency (Frijda, 1986). Goal relevance determines the absolute level of the intensity of emotions; urgency determines how quickly they should be recruited and how steep the rise in their intensity should be. Both variables in turn are influenced by the value of the goal at stake. All else

equal, mismatches with more important goals are more goal relevant and often (although not always) more urgent.

The present reply rejects the fixed tie between heat and the body that so many philosophers take as gospel (e.g., Deonna & Teroni, 2012; see Chapter 1). If the heat of emotions stems from goal-related evaluations, as I have argued above, the link between heat and the body is no longer fixed. This is because the heat produced by goal-related evaluations does not always go hand in hand with the intensity of bodily reactions, even if it often does. To the extent that stimulus evaluation dictates an active action tendency, the body must prepare for the action, and somatic responses arise. The paradigm case is when a fast-approaching car, appraised as relevant and incongruent with one's goal for survival and requiring an urgent response, leads to intense and acute somatic responses. But not all stimulus evaluations with a goal-relevant output demand an active action tendency and hence bodily responses. Learning that one has not obtained a highly desired grant, for instance, may lead to intense disappointment manifested in spooking rumination but may lack the slightest bodily stirring because no urgent action is required.[78] The claim here is that feelings without bodily agitation can still be hot (see also Helm, 2009; Marks, 1993). This argument of "real mental feelings" fits with the idea proposed by contemporary appraisal theories that feelings spring from all components of the emotion, including the appraisal component. Each of the components, when present, colors the eventual experience. Moreover, some research suggests that appraisal even determines the lion's share of this experience. Scherer and Fontaine (2019) found that when participants were asked to classify emotion words (i.e., sedimented feelings), most of the variance was explained by appraisal variables, followed by action tendencies, somatic responses, and motor responses.

But what about cases in which evaluations also remain cold on the mental level, as in climate change and unprotected sex? People who evaluate that the future of the planet is in danger and that the unprotected sex they are about to have is risky most likely do value the state of the planet and their own health. Action may not be urgent in the case of

[78] Also in this case, there is no need to postulate an as-if loop in which a mental simulation of bodily responses is present (see Damasio, 1994; Deonna & Teroni, 2012; James, 1890b; J. J. Prinz, 2004a). In addition, as I will explain in Chapter 7, irrevocable loss does not require action (i.e., assimilation), but rather a reshuffling of priorities (i.e., accommodation) or a reinterpretation of the stimulus (i.e., immunization). The last two strategies do not require any world-directed action, not even in covert counterfactual form.

climate change but it may be so in the case of impending unprotected sex. The key may lie in the fact that these goals have to compete with other goals that have a higher value at the time of evaluation. Sure, climate change is important, but as it is still far in the future it may get snowed under by the hassles of everyday life. Likewise, the threat of unprotected sex fails to motivate avoidance when other more pressing short-term goals take the upper hand. The upshot is that cases in which evaluations remain cold are cases in which the goal at stake is not important enough at the time. The goal relevance criterion of appraisal theories states that stimuli can only elicit emotions if they are relevant for a goal with a sufficiently high value. Thus, cases of cold evaluation are predicted to be emotionally inert, and as such, they simply do not count as counterexamples to appraisal theories. Wrapping everything up, I have argued in this section that if the strict tie between heat and the body is dropped, the cold-hot transition problem disappears so that no theoretical artifices are required in which carts are put before horses, and stimulus evaluation can remain first.

(3) An objection that can be raised against both brands of SET (evaluation-first and embodied), and all other emotion theories discussed up until now, has to do with the tight connection that they assume between stimulus and response features ([S] and [R]) on the subpersonal level, an assumption that is hardly ever questioned. Appraisal theories may propose sophisticated processes to evaluate the stimulus, but the transition from appraisal outputs to the action tendencies is still mediated by the innate [S–R] link (in the biological version), or some average-based equivalent (in the non-biological versions). Even if theories frame this [S–R] link on an abstract level, thereby allowing flexibility on the input and output sides, the link itself is still rigid. Several concrete stimuli can be appraised as dangers, but dangers are still expected to lead to the tendency to avoid, whether this tendency is implemented in running, hiding, or averting one's gaze. Some theorists have framed the link on an even more abstract level, say, as the link between the threat to safety and the tendency to defend oneself or restore safety, which can be implemented in avoidance, but also in attack, and freezing. The increase in flexibility gained by this move, however, comes with the cost that the link becomes entirely eroded (Moors, 2017b, p. 71; Russell, 2009, p. 1278). Saying that a threat to safety leads to the tendency to restore safety amounts to little more than saying that the detection of a discrepancy leads to the tendency to reduce the discrepancy. It leaves us completely in the dark about the behavior that is to follow.

Also in the embodied theories of J. J. Prinz (2004a) and Deonna and Teroni (2012), the subpersonal mechanisms responsible for the causation of action tendencies and their bodily manifestations are innate [S–R] links borrowed from evolutionary theories. The more radical versions of embodied theories, called enactivist theories, tie the knot between [S] and [R] even more firmly. This is true for grounded theories with their sensorimotor representations (Griffiths, 2003). But it is even more true for Colombetti's (2009) enactivist theory in which [S] and [R] are completely fused.

In sum, virtually all theories discussed so far take a stimulus-driven process (i.e., innate or average-based [S–R] or [SR] link) as the backbone mechanism of emotion. Action tendencies are pre-given and lie in wait to be triggered by the right stimulus configuration. Once the key fits the lock, the door pops open. This still seems to be the implicit consensus, no matter how sophisticated the processes are that precede or follow this stimulus-driven process.

It is also worth reiterating that all of the theories discussed so far have contrasted the stimulus-driven process subserving emotion with goal-directed processes involved in emotion regulation and planning. The interplay they assume between these two types of processes is best captured by a default-interventionist architecture in which the stimulus-driven process is the default process and the goal-directed process can occasionally intervene. The goal-directed theory, to be discussed in the next chapter, calls this architecture into question and proposes instead a parallel-competitive architecture in which stimulus-driven and goal-directed processes operate in parallel and compete with one another and in which the goal-directed process (most often) wins the day.

CHAPTER 7

Response Evaluation Theories

SETs discussed in the previous chapter take a process of stimulus evaluation as central to emotions. They specify the types of information that organisms have to extract from the stimulus, the format in which this information is coded, the operations involved in the extraction, and the conditions under which the extraction can take place. In evaluation-first SETs, the extracted information dictates in a straightforward manner the action tendency and responses that must follow and the feelings that must be felt with the help of a stimulus-driven process. Embodied SETs place stimulus evaluation at the end of the causal chain, but they are equally committed to a stimulus-driven process. RETs to be discussed in the present chapter shift their focus to a process of response evaluation. I will concentrate here on my own goal-directed theory (Moors, 2017a, 2017b; Moors et al., 2017; Moors & Boddez, 2017). This theory minimizes the role of stimulus-driven processes while claiming a dominant role for goal-directed processes. Stimulus evaluation still has a place in this theory, but it is cast in a subservient role: (a) to determine whether there is a need for behavior (i.e., detection) and (b) as a source of information that helps select a specific behavior (i.e., selection). Before embarking on the goal-directed theory, I discuss a number of theories that can be considered as precursors and/or as kindred to the goal-directed theory: affordance theories of emotion and "strategic" social theories of emotion.

7.1 Precursors and/or Kindred Theories

7.1.1 Affordance Theories

Affordance theories of emotion (e.g., Griffiths, 2003, 2004b; Griffiths & Scarantino, 2005; Shargel & Prinz, 2018) were discussed in Chapter 6 as part of the family of SETs because, strictly speaking, affordances count as stimulus features. Stimuli are evaluated as flee-able, fight-able, approach-able, and so on. This remains true despite the fact that when affordances were imported in emotion theories they carried a somewhat different meaning than intended by Gibson (1979) who first coined the term. Gibson (1979) reserved the term for state-independent objective stimulus

https://doi.org/10.1017/9781107588882.011 Published online by Cambridge University Press

features, even if he considered them to be relative to the embodiment of an agent (trait-dependent). Thus, scissors are graspable for humans and trees are climbable for squirrels, regardless of whether humans and squirrels are around to evaluate these features. Affordance theories of emotion deviate from Gibson (1979) in that they consider affordances to be state-dependent subjective evaluations. Thus, certain stimuli only become flee-able or fight-able when the organism is around to make this evaluation. This does not change the fact that the evaluation of affordances is a *stimulus* evaluation process, that is, a process that takes a stimulus (or stimulus representation) as its input and produces a representation with response-related stimulus features as its output. This led Griffiths (2004b; see Chapter 6) to suggest that affordances can be considered as additional appraisal variables in appraisal theory. Thus, next to goal relevance, goal congruence, certainty, expectedness, controllability, and agency, he proposed to extend the list with specific options for control such as flee-ability, fight-ability, approach-ability, and so on. Frijda (2007b; Moors et al., 2013) also made suggestions in that direction.

It might be objected that evaluating the affordances of *stimuli* presupposes the evaluation of *responses*. The response evaluation process required here would be one that takes the representation of a response as its input and produces the representation of the response feature "feasible" or "possible" as its output. Based on this reasoning, one might argue that the issue of whether affordance theories count as SETs or RETs is merely a matter of semantics. It must be observed, however, that in the affordance theories discussed so far (Griffiths, 2003, 2004b; Griffiths & Scarantino, 2005; Shargel & Prinz, 2018), no mention was made of a response evaluation process embedded in the affordance evaluation process. So even if these theories were to implicitly assume a response evaluation process, they would qualify as proto-RETs at best.

7.1.2 Strategic Social Theories

Another set of theories that can be considered as kindred to the goal-directed theory are strategic social theories. These theories hold that emotions are shaped by scripts following the rules of society (e.g., Averill, 1982) and/or negotiated during online social interactions (Griffiths, 2003, 2004b; Parkinson, 1995). The social aspects of these theories will receive proper treatment in Chapter 9, which is dedicated to social theories. Here, I will zoom in on the strategic aspect of emotions that these theories emphasize because this is what they share with the goal-directed theory.

The notion of "strategic emotions" can be understood in at least two different ways. A first way is to consider emotions themselves as reactive or non-strategic but to argue that they can be used in a strategic way, as a means to reach a goal (Ainslie & Monterosso, 2005). This corresponds to cases of emotion regulation in which emotions are initiated (i.e., antecedent upregulation) or whipped up (i.e., consequent upregulation) (see Box 7.1). It is the goal-directed nature of emotion regulation that endows the emotion with a strategic character. Take a politician who has the superordinate goal to win a debate. Her knowledge that aggressive behavior will put her opponent out of balance may induce the subordinate goal to initiate aggressive behavior or even the goal to enter into a full-blown anger episode that includes angry feelings (perhaps because she believes that this will make her aggression appear more authentic). Various strategies are at the politician's disposal. She can reappraise her opponent as a really nasty person, but she can also start raising her voice in the hope that this will spread to her own feelings and back to her behavior (Parkinson & Simons, 2012, p. 473). A more sophisticated example is the strategic upregulation of emotional behavior as a way "to move socially sanctioned behavior into the realm of passive, involuntary, and thus excusable behavior" (Averill, 1980, p. 312; Griffiths, 2003, p. 61; Mesquita & Parkinson, 2022). An uncle at a family dinner, for instance, who realizes that he has gone too far, may drive his act so completely over the top that others recognize it as emotional and shake their heads instead of excommunicating him from the family. The notion of strategic emotions at stake in these examples is uncontroversial. All theories discussed until now hold that genuine emotions are non-strategic but that they can be faked as part of a strategy, and in doing so, may eventually even become real ("fake it till you make it").

A second way to understand the strategic nature of emotions is to consider them to be inherently strategic, already in non-regulated form. This is the view held by strategic social theories. It must be noted that several of these theories do not rule out the existence of reactive or non-strategic emotions. For instance, Averill (1982) and Griffiths (2003, 2004b) placed emotions on a continuum from more reactive to more strategic (see also Hinde, 1985a, 1985b). The crucial difference with the first notion of strategic emotions discussed in the previous paragraph, however, is that strategic emotions are not considered here to be less authentic than their reactive counterparts. The politician's tendency to engage in aggressive behavior as a strategy to knock out her opponent is genuinely emotional. Her feeling of anger is the feeling of that action tendency. Likewise, the uncle's tendency to upregulate his shouting at the family dinner is genuinely emotional.

https://doi.org/10.1017/9781107588882.011 Published online by Cambridge University Press

The strategic nature of emotions hints at the involvement of a response evaluation process but just like in affordance theories, strategic social theories do not give us much to work with. As I try to show below, the mechanisms put forward by strategic social theories do not fit the description of response evaluation processes, and they seem ultimately incapable of endowing strategic-ness to action tendencies. Motivated by the desire to make their theories also applicable to young children and animals, both Griffiths (2003, 2004b) and Parkinson (2007) stressed that the strategic nature of emotions does not require a complex decision process. Some social theories take as the preferred unit of analysis the unfolding social interaction, which is composed of a series of reciprocal actions among the interaction partners. Zooming in on the level of a single action from one partner to another, the mechanisms considered by Parkinson (2007, 2011, 2019a, 2019b; Parkinson & Manstead, 2015) are (a) emotional contagion, (b) social appraisal, and (c) direct adjustments to relational dynamics (see also Chapter 9). In the mechanism of emotional contagion, the nonverbal (e.g., facial) responses of another person are mimicked after which feedback of these responses induces a convergent emotion in the agent (Hatfield et al., 1994). Emotional contagion does not count as a response evaluation process and it does not yield strategic-ness. The agent simply adopts the response of another person without evaluating it and without meaning anything by it (but see Lakin et al., 2008; Parkinson, 2011; Chapter 9).

In the mechanism of social appraisal, an agent gets indirect access to the appraisal values of stimuli via the behavior of another person (Manstead & Fischer, 2001). This differs from the "personal" appraisal process discussed in Chapter 6 in which an agent gets direct access to the appraisal values of stimuli. Again, social appraisal does not count as a response evaluation process but rather as another indirect stimulus evaluation process (in addition to the indirect stimulus evaluation processes proposed by J. J. Prinz, 2004a, and Deonna & Teroni, 2012).

Finally, the mechanism of direct adjustment to relational dynamics is left rather vague, except that it is put forward as an alternative to personal appraisal (Parkinson, 2005, 2007). Illustrations are anger from hitting one's head against the kitchen cabinet and anger of an infant struggling from a tight embrace, which do not require an appraisal of external agency or other-blame (Frijda, 1986; Parkinson, 1999). Other illustrations are shame from being in the center of attention without an appraisal of self-devaluation, and the feeling of guilt from being falsely accused without an appraisal of internal agency or self-blame (Griffiths & Scarantino, 2005). In all these examples, however, a primitive form of personal appraisal, which minimally evaluates the goal incongruence of stimuli, seems hard to deny. What seems to be still missing from

Parkinson's (2007) theory, however, is how these primitive appraisals get translated into action tendencies and how these action tendencies get to be strategic.[79,80]

Griffiths and Scarantino (2005) likewise insist that complex, conceptual forms of appraisal are not involved in emotion causation. These authors, moreover, toy with the idea of extending appraisal theory with appraisal variables that code the value of having certain emotions. Spinning out their view further, however, it seems that the real mechanism they have in mind is no less than an affect program. They believe that a return to affect programs is justified because it enables animals to also have strategic emotions. The problem is that affect programs are unable to deliver on the strategic promise. They may have developed from processes that were "serviceable" at some point in our evolutionary history, as Darwin (1872) suggested, but once they crystallized into affect programs, they lost every power to adapt to the current outcomes of behavior. This does not seem to bother Griffiths (2003), as he writes that "a strategically sensitive emotion system might give rise to emotional episodes that appear self-serving and manipulative without the agent forming a plan to pursue their social interests or engage in manipulation" (p. 62). This comes down to saying that emotional behavior need not *be* strategic, but merely has to *appear* strategic. This weakens the meaning of strategic emotions up to a point where emotions are merely seen as functional in an evolutionary sense, marking a complete return to evolutionary theory. The return to affect programs not only robs emotions from their strategic potential, it is also not necessary to make sense of animal emotions. This is because decades of operant learning research have demonstrated that animals are

[79] These authors further argue that emotions are not individuated on the basis of their appraisal profiles but rather on the basis of their action tendency: remove the obstacle, disappear from sight, and repair the relationship. It is important to note, however, that all theories discussed so far include action tendencies. Here, they are simply disconnected from complex appraisal patterns, but not from simple appraisal patterns. In other words, all that the mechanism of direct adjustment to relational dynamics offers is a slimmed-down appraisal pattern.

[80] Note that the mere inclusion of action tendencies or the individuation of emotions on the basis of action tendencies (e.g., Frijda, 1986; Scarantino, 2014) does not render a theory strategic. All action tendencies are defined in terms of a valued outcome, that is, an end state by which they can be considered as successfully accomplished. Thus, the tendency to approach is defined in terms of distance reduction, the tendency to avoid in terms of distance increase, and the tendency to put on the light in terms of the light being on. Each of these action tendencies can in principle be caused by a stimulus-driven process ([positive-valence–approach]; [negative-valence–avoidance], [entering-the-room–putting-on-the-light]). They only count as strategic or instrumental when they are selected as strategies to reach a further goal (approach to eat food, avoidance to get to safety, and putting on the light to be able to read) (see Box 2.1).

perfectly capable of real instrumental behavior (e.g., Adams & Dickinson, 1981; Balleine & Dickinson, 1998; Dickinson & Balleine, 1994). In conclusion, social theories that highlight the strategic nature of emotions may only be able to provide apparent, but not real strategic-ness to emotions.

The goal-directed theory of emotion (Moors, 2017a; Moors et al., 2017) shares with its predecessors the aim to account for the strategic character of the phenomena called emotions. In contrast to its predecessors, however, the theory qualifies as a full-blown RET in that it proposes a detailed response evaluation process in which the steps and types of information that go into this process are unpacked. The goal-directed nature of this response evaluation process, moreover, makes it perfectly suitable to deliver on the strategic promise. Finally, the goal-directed theory of emotion holds that the processes causing emotional behavior are identical to the processes causing non-emotional instrumental behavior. In other words, the theory is a theory of behavior causation that can be applied to all sorts of behavior including emotional behavior. As will be explained below, the lack of a mechanistic distinction between emotional and non-emotional behavior will end up in the eliminativist conclusion that emotions are not a scientific set and that all we are left with are behavioral episodes that *appear* more or less emotional.

7.2 Constitutive and Mechanistic Explanations

The goal-directed theory of emotion is a dual-process theory. Although the emphasis is on goal-directed processes, it does not go so far as to rule out the existence of stimulus-driven processes. Most theories discussed so far are single-process theories when it comes to emotion – with a more or less sophisticated stimulus-driven process subserving emotion – but they do hold a dual-process view regarding the entire mental realm. The stimulus-driven process is for emotions and non-emotional habits; the goal-directed process is for emotion regulation and planning or instrumental behavior. The goal-directed theory, by contrast, proposes a dual-process view for emotion itself. The action tendencies in emotional episodes can be caused by stimulus-driven as well as goal-directed processes, although the bulk of the explanatory weight is on the goal-directed ones. Unlike other dual-process theories, however, the goal-directed theory is not a dual-system theory in that it resists stacking other dichotomies on top of the dichotomy between goal-directed and stimulus-driven processes.

The stimulus-driven process in the goal-directed theory takes the same shape as in the previous theories: A stimulus activates an innate or learned [S–R] link. The [S] part is a stimulus representation, with concrete

https://doi.org/10.1017/9781107588882.011 Published online by Cambridge University Press

([ecS]) or abstract ([eaS]) features. The [R] part is a response representation, which is an action tendency or intention.

The goal-directed process takes the shape of a cycle (see Box 2.1), which can roughly be split in three broad phases: (a) discrepancy detection, (b) strategy/behavior selection, and (c) feedback of the outcome of the behavior to the start of the cycle. The particular goal-directed theory proposed here shows partial overlap with other theories of emotion. The detection and feedback phases have overlap with SETs such as appraisal theories (Scherer, 1994a, 2001a) and causal belief-desire theories (Reisenzein, 2012), but also with control theories of affect (e.g. Carver, 2003, 2004, 2006; Carver & Scheier, 1990, 2008, 2013; Cochrane, 2019[81]) and reinforcement learning theories of emotion (e.g., Rolls, 1999, 2005). The selection phase has overlap with Duffy's (1941a, 1941b) early suggestions, and again with some reinforcement learning theories of emotion (Broekens et al., 2015; Mowrer, 1960; Rolls, 2005). But the goal-directed theory presented here is not identical to these other theories. Its unique features will become clear in a more detailed description of the goal-directed cycle to which I now turn (see Figure 7.1).

The goal-directed cycle starts with a comparison between the representation of a stimulus (i.e., an actual or anticipated state; [eS]) and a first goal (i.e., a desired state, i.e., the representation of a valued outcome; [O^v]) (Step 1; cognitive component). If a discrepancy is detected between the two elements, a second goal is activated, aimed at reducing this discrepancy ([*rd*]) (Step 2; motivational component). To achieve the second goal, a selection must be made between three broad strategies (Step 3; cognitive component). A first strategy is to act (i.e., *assimilation* of the stimulus to the goal, [R_{as}]). A second strategy is to change the first goal (i.e., *accommodation* of the goal to the stimulus, [R_{ac}]). A third strategy is to change the interpretation of the stimulus (i.e., *immunization*: changing the content of the stimulus representation, [R_{im}]).[82] To illustrate, if a student receives a low grade for an exam, which constitutes a discrepancy with the goal to be seen as a bright student, she may select the action to study hard for a retake or conceal her results from others (i.e., assimilation), she may drop the goal to be seen as a bright student and invest more in other parts of her personality (i.e., accommodation), or she may

[81] Cochrane (2019) proposes an emotion theory in which the core element is an [S–R] link combined with a negative feedback loop.

[82] Note that the strategy of immunization accords with Sartre's (1939) proposal that emotions are forms of magical consciousness. To illustrate, in the case of terror we do not act to solve the problem but we deny the stimulus, for instance, by fainting. This removes the stimulus from our consciousness.

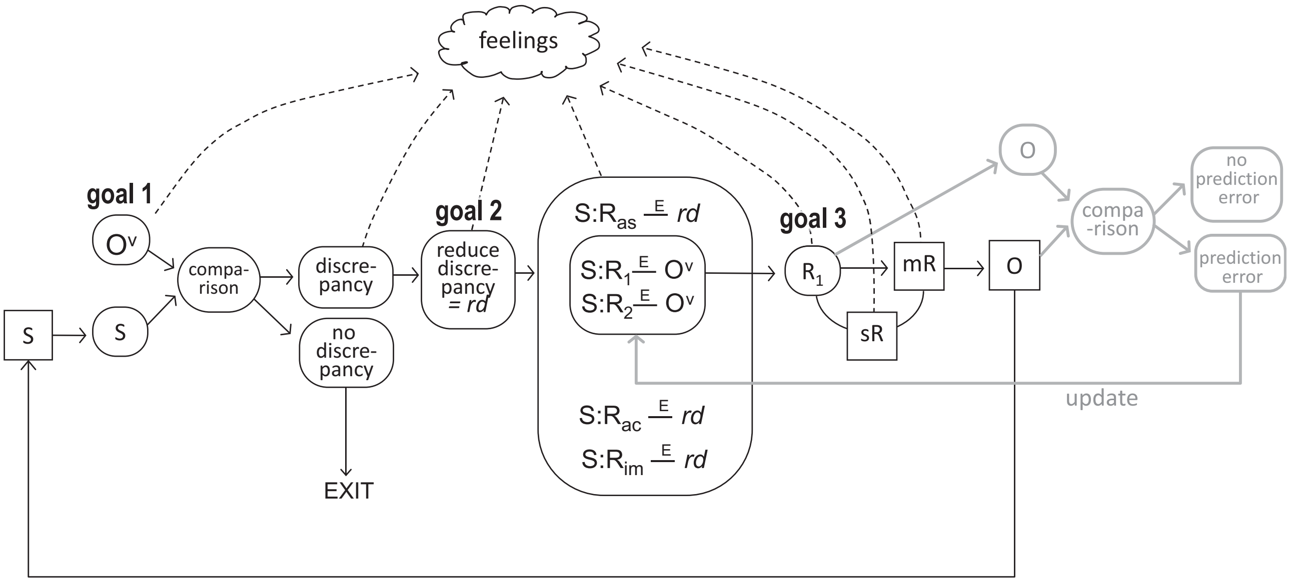


Figure 7.1 Goal-directed theory
Note: Adapted from *Opinion in Behavioral Sciences*, 39, Moors, Van de Cruys, & Pourtois, Comparison of the determinants for positive and negative affect proposed by appraisal theories, goal-directed theories, and predictive processing theories, 147–152. Copyright (2021), with permission from Elsevier.

attribute the low grade to fatigue and as not reflecting her capacities (i.e., immunization).

If assimilation is chosen, the system must still select a specific action option (Step 4; cognitive component). This is done by weighing up the expected utilities of the action options available in the system's repertoire and by selecting the option with the highest expected utility. The selected action option activates its corresponding action tendency ([R]). This is an intention to act and the third goal in the cycle (Step 5; motivational component). The action tendency is manifested in somatic responses (sR) that prepare and support overt behavior (mR) (Step 6; somatic and motor components). The overt behavior results in an actual outcome (O) (Step 7). This is a new stimulus (S) that is fed back to the start of the cycle (Step 8). Ideally, the cycle is repeated until there is no discrepancy left. (Brandstädter & Greve, 1994; Brandstädter & Rothermund, 2002; Sprangers & Schwartz, 1999).

Aspects of all steps in this goal-directed cycle may pervade into consciousness where they contribute to the content of feelings (feeling component). Positive and negative affect are generated in at least two of these steps. In Step 1, during the detection phase, the presence of a discrepancy produces negative affect whereas the absence of a discrepancy produces positive affect. In Step 4, during the action selection phase, the presence of action options with a sufficiently high expectancy for reducing the goal discrepancy and reaching goal satisfaction produces positive affect whereas the absence of such action options produces negative affect. Thus, positive and negative affect are not only generated by actual goal satisfaction but also by anticipated goal satisfaction (Moors et al., 2021).[83] Of course, feelings may extend beyond valence. Feelings are considered to add a conscious layer on all kinds of representations, including action tendencies. In this theory then, action tendencies can be felt centrally. Overt somatic and motor responses are not representations. They need to travel via the perception of somatosensory feedback to become the content of representations (iS→[iS]), which in turn can become conscious and hence felt (see Moors, 2017a, p. 12). Feelings can be labeled with an emotion word, but they can also remain unlabeled or raw.

It may be noted that even if the goal-directed theory covers all typical emotional components, it remains agnostic about which of these components should be part of a constitutive explanation of emotion. The theory has a skeptical outlook and merely tries to explain the phenomena

[83] I believe this account of affect is more parsimonious than that proposed by Carver and Scheier (2013).

https://doi.org/10.1017/9781107588882.011 Published online by Cambridge University Press

that people call emotions (see also Duffy, 1941a, 1941b; Meyer, 1933; Russell, 2003).

The complexity of the goal-directed process increases if we take into account multiple goals (i.e., multiple valued outcomes), which may either align or conflict with one another. For instance, shouting during a discussion with the aim of dominating another person in the short run may constitute a discrepancy with entertaining a stable relationship with this person in the long run.

A further source of complexity stems from the hierarchical organization of goals and subgoals (Pezzulo & Cisek, 2016; Powers, 1973a, 1973b; R. Smith et al., 2017). The same principles that guide the selection of an action tendency also guide the selection of the first goal in the cycle. Indeed, this first goal ([O_1^v]) can itself be considered as a subordinate goal that is instrumental for a superordinate goal ([O^{v*}]) (see Figure 7.2). For instance, a person may select the action tendency to hit an opponent as a means to win a battle but the goal to win a battle may itself have been selected as a means to gain status. Apart from our most superordinate goals, which are deemed to be innate and universal, the subordinate goals that we select depend on the reward structure of the environment. For instance, the superordinate goal to gain status prompts selection of the subordinate goal to win a battle if society rewards winners with more status.

The hierarchical organization not only extends upwards to higher-order goals but also downwards to lower-order goals. The goal to hit an opponent can give rise to a sequence of lower-order goals or motor programs for moving the torso, fists, and hands (Balleine et al., 2015; Mylopoulos & Pacherie, 2017). I now take a closer look at several steps in the goal-directed cycle, as this allows me to highlight some differences from related theories.

Steps 1 and 2

The first step detects the presence and magnitude of a discrepancy between the representation of a stimulus and a first goal. This maps perfectly onto the appraisal checks of goal in/congruence and goal relevance (e.g., Lazarus, 1991). Like all goals, this first goal is the representation of a valued outcome. The theory does not assume that the starting point of a cycle can be an anti-goal, that is, the representation of a disvalued outcome, as is the case in some feedback theories (e.g., Carver & Scheier, 2013) and in theories that distinguish between separate behavioral approach and inhibition systems (J. A. Gray, 1994).

The stimulus that is compared with the first goal can be either an actual stimulus or an anticipated stimulus. This creates two scenarios. The first

https://doi.org/10.1017/9781107588882.011 Published online by Cambridge University Press

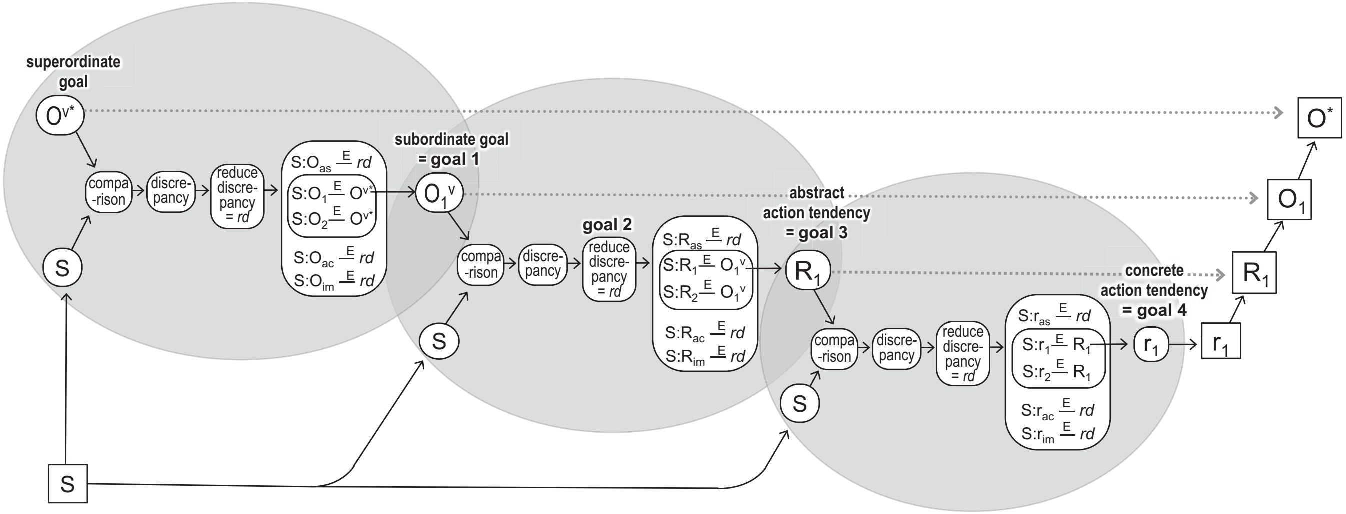


Figure 7.2 Hierarchy of goal-directed cycles

scenario is a discrepancy between a first goal and an actual stimulus, say, the goal to succeed on an exam and the news that one has failed. In this scenario, the second goal to reduce the discrepancy amounts to improving things and can be called a promotion goal (Higgins, 2018). The second scenario is a discrepancy between a first goal that is currently satisfied – and therefore equal to the actual stimulus – and an anticipated stimulus. The paradigm example is the case in which the first goal is to be safe, which is already satisfied, but there is a risk that this will not last. In this scenario, the second goal to reduce the discrepancy amounts to preventing the actual state from getting worse and can be called a prevention goal (Higgins, 2018). It is worth pointing out that promotion and prevention goals in Step 2 do not dictate the action tendencies that will be selected in Step 4. A promotion goal may lead to approach or avoidance and so may a prevention goal. If Sam wants to improve his relationship with Sunny after a fight, he may either try to talk to her or go for a walk. He may select the exact same actions if his aim is to prevent his relationship from getting worse.

The theory assumes that behavior only occurs when there is a need for behavior, that is, when there is a discrepancy between a stimulus and a goal. So how does the theory explain behavior in the case of positive emotions, such as those involved in appetitive approach and so-called broaden-and-build behavior (Fredrickson, 2004)? Again, two scenarios are possible. The first scenario is when the stimulus signals an imminent (or partial) match with a first goal that is not yet completely obtained (Moors, 2009). The remaining discrepancy still needs to be bridged via behavior, and this could be approach behavior. For instance, the chocolate cake on the table may signal an imminent match with my goal to taste it, but as long as it is still on the table, I still need to take action: approach it or ask for it. The second scenario is when there is a complete match with a first goal. When one goal is fulfilled, the person can open up to the pursuit of other goals. If there are no urgent discrepancies, she can engage in exploratory behavior to increase the probability of encountering new opportunities for goal satisfaction. This matches the broaden-and-build case.

The transition from the discrepancy detection (Step 1) to the second goal to reduce the discrepancy (Step 2) can be conceived of as an [S–R] link, cast on the highest level of abstraction. The [S] part is the representation of the abstract stimulus feature "goal discrepant"; the [R] part is the abstract tendency to reduce this goal discrepancy. One might even speculate that this abstract [S–R] link is innate. I guess it would not be too crazy to assume that when something is wrong, organisms have the innate tendency to fix it. Such a mechanism would still be general-purpose, however, in the sense that discrepancies may pertain to any possible goal.

Steps 3 to 6 (and 7 and 8)

Step 3 involves the selection of the broad strategies of assimilation (i.e., to act), accommodation (i.e., to replace the first goal), and immunization (i.e., to change interpretation of the stimulus). Assimilation entails overt or physical actions, whereas both accommodation and immunization entail covert or mental actions. If assimilation is selected, there is a further Step 4 in which a more narrow strategy or action option must be selected. The theory hypothesizes that not only the selection of an action option but also the selection of the broad strategy is based on a weighing up of expected utilities. If the action option of fighting turns out to have a low expectancy, the organism may switch gears and choose to flee instead (i.e., still assimilation but different action). If all actions turn out to be ineffective, however, the person may be forced to give up the first goal and choose a different goal (i.e., accommodation), or she may try to see the facts in a more favorable light (i.e., immunization).[84,85]

Strictly speaking, a process already counts as goal-directed if the expected utility of a single action option is processed. One action option can have multiple outcomes. Some are benefits, others are costs. In the current theory, costs are framed as low expectancies for reducing discrepancies with other goals. For instance, fleeing may have a high expectancy for reaching safety, but a low expectancy for saving energy. Fighting may have a high expectancy for winning a debate but a low expectancy for preserving the quality of a relationship. If the action repertoire – which is the set of action options or affordances that are accessible to the system at a certain time – contains more than one action option, these options need to be compared in order to select one of them (Morris et al., 2014).

The selection of action options and strategies relies on subjective estimates of values and expectancies. These estimates may rely on various operations. Values and expectancies can be computed online (i.e., tertiary process), but they will for a large part be pregiven by previous operant learning (i.e., secondary process). The updating of expectancies can happen as follows (Pessiglione et al., 2008; Wolpert et al., 1995). At the time that an action tendency is translated into actual behavior (Step 6), a representation of the expected outcome of the behavior is also produced (i.e., efferent copy; [O]). Once the behavior has occurred and has produced an actual outcome (O), the actual and expected outcomes are compared. In the case

[84] It is conceivable that Steps 3 and 4 are not sequential but integrated (see Figures 7.1 and 7.2).

[85] Accommodation and immunization are reminiscent of two types of emotion regulation strategies that have been classified under the heading of reappraisal. Here, they are not considered as strategies to control emotions, but simply as ways to decrease the stimulus-goal discrepancy.

of a mismatch (i.e., prediction error), the previously represented response-outcome expectancy can be adjusted. If, based on this adjustment, the system selects a different action next time, this corresponds to accommodation on the level of the action tendency. If the system persists in the same course of action, this counts as assimilation on that level.

Finally, some theorists (e.g., Bargh et al., 2010) have argued that the expected utility – also called the *desirability* – of the action options should be weighted by their *feasibility* or the extent to which a person expects to be able to execute the actions. For instance, fleeing may be desirable but not feasible because of a lack of capacity (e.g., the opponent is faster) or opportunity (e.g., there is no escape route). Expected utility or desirability is upward-looking (i.e., it tells how likely it is that an action will satisfy the first goal); feasibility is downward-looking (i.e., it tells how likely it is that an action will be implemented in a motor program). Other theorists assume that the action options in people's action repertoires only include the feasible ones, making a separate feasibility factor superfluous.

The goal-directed theory outlined here is a mechanistic sketch in which the types of information figuring in the contents of the representations dominate the agenda, and in which a liberal position is held regarding the format of the representations. Although expectancies and values are abstract pieces of information, the outcomes and behaviors to which they pertain may be represented in concrete, image-like format. The person can picture the outcome of different behavior options (during the selection phase; Step 4), but also what will happen if no action is undertaken (during the discrepancy detection phase; Step 1). For instance, the person can picture that the presence of a raging bull in a pasture will produce a bloody outcome, unless she decides to jump over the nearest fence. It could be hypothesized, moreover, that the vividness of these representations has a biasing influence on estimations of values and expectancies. This would be in line with the availability heuristic in the prospect theory of Tversky and Kahneman (1973) and the predictions of the construal level theory of Trope and Liberman (2010). The availability heuristic entails that events that come more readily to mind (and are therefore perhaps more vivid) are judged as more likely. Construal level theory proposes that nearby outcomes, which are more concrete (and are therefore perhaps more vivid) are judged as more likely.

Another assumption regarding the format of the representations in the cycle is that the abstract notions of values and expectancies are not propositions that need to be judged as true. The organism activates values and expectancies and treats them as true by default unless there is evidence to the contrary. No separate step of truth evaluation is thus required. This feature is important as it helps explain recalcitrant emotions as well as emotions from fiction and art. The mere imagination, construal, or perception of a goal-discrepant stimulus may be sufficient to trigger a goal-directed cycle in which possible action options and solutions are imagined. If the contents of these imagined representations break into consciousness, moreover, they are felt.

A further assumption of the theory is that all steps in the goal-directed cycle can be automatic. Several arguments make this plausible. First, goal-directed processes vary in complexity and at least the simple ones should be automatic. It seems reasonable to assume that the capacity limitations of the system constrain the number of action options and outcomes that can be processed simultaneously (see Pezzulo & Cisek, 2016). But even if the weighing of outcomes proceeds in a sequential manner, this may still happen rapidly (e.g., Berkman et al., 2017). Second, goal-directed processes do not have to rely on rule-based computation, but can also be based on the activation of preset knowledge via an associative operation. The latter operation is generally accepted as automatic. Third, the goal relevance of the stimuli in goal-directed processes may partly compensate for a lack of other operating conditions, such as a lack of time (Moors, 2016; see Chapter 1). This is supported by subliminal priming studies, in which goal-relevant stimuli broke into consciousness under shorter presentation durations than goal-irrelevant ones (e.g., Tapia et al., 2010). Fourth, evidence that animals are capable of goal-directed processing is ubiquitous (Heyes & Dickinson, 1990). This suggests that these processes do not require high-level cognition and should be fairly automatic. Fifth, evidence for automatic goal-directed processing in humans is growing (Hommel, 2013, 2017; Pessiglione et al., 2007; van Gaal et al., 2012).[86]

The theories discussed in the previous chapters differ from the goal-directed theory with regard to the interplay that they assume between stimulus-driven and goal-directed processes (see Box 2.1). The previous theories hold a default-interventionist architecture with stimulus-driven (i.e., emotional) processes as the default and goal-directed (i.e., emotion regulation) processes as the intervenor. This architecture builds on the widely assumed trade-off between automaticity and rationality (see

[86] Critics might object that if behavior is caused by a goal-directed process, it must be caused by an intention (i.e., action tendency or third goal) so that it cannot be automatic in the sense of unintentional. This is true. Even so, the intention need not be a conscious intention, and the process itself need not be slow, inefficient, conscious, and/or difficult to counteract. Thus, the behavior can still be automatic in various other ways. Note, moreover, that goal-directed processes are not more intentional than stimulus-driven ones, given that the response representation ([R]) that is part of the [S–R] link is also an action tendency or intention (see Box 2.1; see Chapter 9).

https://doi.org/10.1017/9781107588882.011 Published online by Cambridge University Press

Chapter 1). Stimulus-driven processes are seen as simple, which makes them (a) more automatic and therefore the default process but also (b) more rigid and therefore more error-prone. They are quick but dirty. Goal-directed processes, by contrast, are seen as complex, which makes them (a) more non-automatic but also (b) more flexible and therefore more likely to be accurate. The implications are that when operating conditions are poor, the agent has no other choice but to rely blindly on a stimulus-driven process. Only when operating conditions are ample, can a goal-directed process intervene to correct the initial action tendency elicited by the stimulus-driven process.

In sharp contrast with the previous theories, the goal-directed theory (Moors, 2017a; Moors et al., 2017) endorses a parallel-competitive architecture in which both processes are assumed to operate automatically so that they will often operate in parallel and enter into competition with each other. The goal-directed process, moreover, is hypothesized to often win the competition because it combines automaticity with rationality whereas the stimulus-driven process only furnishes automaticity, and the system should prioritize the process that yields the most benefits. Another argument that is not "from design" could be that the [S–R] associations in stimulus-driven processes tend to be weaker than [S:R–O^v] associations, and therefore often lose out (Fischer et al., 2022). The upshot is that in this architecture, the goal-directed process is the default determinant of behavior whereas the stimulus-driven process is the exception.

It may be worth pointing out that the competition between goal-directed and stimulus-driven processes will most often remain under the radar. If stimulus-driven processes are indeed weak, they will not reach awareness and no tension with a goal-directed process will be felt. The tension that we do often experience in self-regulation conflicts (within or outside of the emotional realm) is framed in this theory as stemming from the competition between two goal-directed processes that are each at the service of a different goal.

The theory considers at least two exceptions in which a stimulus-driven process could still determine behavior, although alternative goal-directed explanations remain possible in both cases. A first exception is when the system is faced with two equally attractive action options. In Buridan's thought experiment, a donkey placed in front of two identical stacks of hay at equal distance cannot decide between them and consequently dies of hunger. In such a case, a pre-existing [S–R] link including one of the response options could bias selection in favor of that response option. Thus, if the donkey previously received hay on the right side, a stored [S–walk-right] link could tip the balance. According to an alternative, goal-directed explanation, however, previous experience with hay on the right side could also have altered the value of the right haystack (or its associated responses), for instance, via the principle of mere exposure (Zajonc, 1968).

A second exception is when the action option that is selected is cast at a too high level of abstraction so that it does not enter into competition with an existing [S–R] link cast at a lower level of abstraction. In this case, the [S–R] link could still determine behavior because it does not suffer from competition. To illustrate, if my office was on the third floor for years, this might have formed an [elevator–press-3] link. Suppose my office recently moved to the fifth floor. If upon entering the elevator I merely activate the action tendency to go to my office but do not implement this in the more concrete action tendency to press 5, the stored [elevator–press-3] link does not suffer from competition and can freely determine behavior. According to an alternative, goal-directed explanation, however, the system never stored an [elevator–press-3] link to begin with but rather an [elevator: press-3–access-to-office] link. The novel [elevator: press-5–access-to-office] link, which is now correct, is too weak at present to override the old [elevator: press-3–access-to-office] link, which is now incorrect (see Buabang, Köster, et al., 2021).

7.3 Scientific Definitions

7.3.1 Intensional Definition

This section considers whether the causal-mechanistic explanation of the phenomena called emotions provided by the goal-directed theory allows us to distill criteria that could potentially serve to demarcate the set of emotions from others sets. These criteria are first evaluated in terms of the meta-criterion of apparent-similarity, that is, whether they can account for the desiderata listed in the working definition, and then in terms of the meta-criterion of fruitfulness, that is, whether they can deliver a common denominator for the members of the set that is deep enough to allow for scientific extrapolation. My conclusion will be that apparent-similarity is guaranteed, making the goal-directed explanation a suitable explanation for alleged emotions, but that fruitfulness is lacking, preventing us from elevating the folk set of emotions to a scientific set.

7.3.1.1 Criteria for Demarcation

The goal-directed theory suggests that what people recognize as emotions are goal-directed cycles in which a discrepancy is either present, anticipated, or absent/solved, and in which the person selects a physical or mental action based on a cost-benefit analysis, or no action if none can be found. This is the first criterion, which I call the "goal-directedness"

https://doi.org/10.1017/9781107588882.011 Published online by Cambridge University Press

criterion. The second criterion, which I call the "value" criterion, is that the value of the goal at stake (first goal) must be sufficiently high. Thus, the loss of a friend and the loss of the soap under the shower initiate a processing cycle with the same steps, but the reason why people think losing a friend leads to an emotional cycle and losing the soap does not, or less so, is because the goal to keep the friend has a high value whereas the goal to hold on to the soap does not, or less so. This is a gradual criterion that matches with the "goal relevance" criterion put forward by molecular SETs (see Chapter 6).

7.3.1.2 Adequacy

APPARENT-SIMILARITY

Equipped with these two criteria for demarcation, let us now examine how the theory fares in the light of the desiderata from the working definition.

(1) The goal-directedness criterion accounts for the world-directed Intentionality of emotions because the goal-directed process is made up of representations that contain information. The goal-directed cycle combines a (minimal) stimulus evaluation process (in the detection phase, Step 1) followed by a response evaluation process (in the selection phase, Steps 3 and 4). The stimulus evaluation process assesses whether and which goal is violated (or threatened) or satisfied. The response evaluation process assesses which overt or covert action is best undertaken. Together this yields a richer form of Intentionality than the stimulus evaluation process in SETs. Appraisal of a stimulus as a loss, threat, or offense (i.e., molar values) does not tell us yet which goal is lost, threatened, or violated or what can be done about it. Likewise, appraisal of a stimulus as goal-incongruent and easy to control (i.e., molecular values) does not tells us which goal is at stake or in what sense (via which specific behavior) the stimulus is easy to control.

In addition to providing a richer form of world-directed Intentionality, the goal-directed theory also provides a unique account of the double direction of fit that emotions display according to some authors (Helm, 2001, 2009; Roberts, 2013; see Chapter 2). In fact, all theories discussed so far that include both a stimulus evaluation process and an action tendency into their constituent explanation of emotion (i.e., SETs and hybrids of this theory with evolutionary and network theories) could account for this double direction of fit: Stimulus evaluation is a cognitive entity, which has a mind-to-world direction of fit, ensuring the world-directedness of emotions; an action tendency is a conative entity, which

https://doi.org/10.1017/9781107588882.011 Published online by Cambridge University Press

has a world-to-mind direction of fit, ensuring the self-directedness of emotions. Stimulus evaluation is accurate when it conforms to the world. Evaluating a snake as dangerous is accurate for instance, if the snake is indeed dangerous. An action tendency is satisfied when the world conforms to it. The tendency to avoid, for instance, is satisfied when the end state that defines the action tendency is reached, that is, when the distance from an object is increased.

Some authors have suggested that action tendencies in fact already display a double direction of fit (like other conative entities such as desires; Lauria, 2017; Oddie, 2005). They are not only directed at reaching the end state that defines them so that they can be judged as more or less satisfied; they can also be judged as more or less accurate ways for reaching a superordinate goal. The tendency to avoid, for instance, can be judged as more or less satisfied, but also as more or less accurate for reaching the goal of safety. Other authors, however, have rejected the possibility that a single mental entity can have a double direction of fit (e.g., Döring, 2007; M. Smith, 1987). As M. Smith (1987) puts it, conative entities, when satisfied, should go out of existence, whereas cognitive entities, when accurate, should endure. This would make the idea of a single mental entity with both directions of fit incoherent.

The goal-directed theory provides an elegant solution to this puzzle. It can handle the intuition that action tendencies can be evaluated as accurate or inaccurate for reaching a goal without having to endow them with a mind-to-world direction of fit. Indeed, the goal-directed theory postulates that action tendencies are preceded by a selection phase in which the expected utilities of action options are processed. This involves the subjective estimation of values and expectancies, and it is these that can be judged as more or less accurate, that is, as matching more or less with objective values and probabilities. Thus, it is the selection phase that has a mind-to-world direction of fit, but this fit need not be transferred to the action tendency that follows on this phase.

If next to assimilation (i.e., strategy to act), we also consider the strategies of accommodation (i.e., strategy to replace the first goal) and immunization (i.e., strategy to reinterpret the stimulus), the story can be extended as follows. In accommodation, the goals of the mind adapt to the world. In immunization, the world is adapted to the goals of the mind but only in a mental sense, which comes down to (the afferent part of) the mind adapting to (the efferent part of) the mind. For both strategies, it can also be judged whether the selection of the strategy was accurate or not.

All steps in the goal-directed cycle that have a world-to-mind direction of fit can be inaccurate. This not only allows us to explain cases of theoretically irrational emotions, but also cases of practically irrational

https://doi.org/10.1017/9781107588882.011 Published online by Cambridge University Press

emotions. In the previous theories, we saw that the theoretical irrationality of emotions can account for some cases of practically irrational emotions, such as when my fear of house spiders prevents me from sleeping, but not for other cases, such as costly aggression and arational actions. To explain the latter cases, these theories have to appeal to the stimulus-driven processes that they hold responsible for emotions. Since the goal-directed theory does not ascribe a substantial role to stimulus-driven processes, this *deus ex machina* solution is not available for this theory. Yet in this theory, a much stronger relation is assumed between theoretical and practical rationality. I will first consider the implications of this assumption for the issue of practical irrationality before turning to standard cases of theoretical rationality.

At first blush, the goal-directedness criterion flies in the face of the apparent practical irrationality of emotions. The goal-directed process is the process par excellence to deliver practical rationality because it can flexibly adapt to changes in the current outcomes of behaviors. This is precisely the reason why most emotion scholars dismiss a goal-directed explanation of emotions outright and why they advocate dual-system models in which emotional behavior is placed in sharp contrast with instrumental behavior (Ainslie, 2001, pp. 48–70).

Two replies can be formulated. First, some instances of emotional behavior *appear* irrational but are in fact rational. This is because people have multiple goals, and the goal driving their behavior may be hidden, not only to observers, but also to the individual. Take the case of an aggressive outburst. This behavior is costly in the sense that it can ruin a relationship, but at the same time, it may help restore the person's social status (as someone who cannot be messed with). The social status goal may have a higher value than the relationship goal, but it may also be hidden because it is not such a noble goal to have.

Second, the goal-directed process can also produce *real* irrational behavior. Several steps in the goal-directed cycle carry risks for errors. A person may assign the wrong values to goals; she may fail to detect the discrepancy between a stimulus and a goal; her action repertoire may not be equipped with the most adaptive action options; she may over- or underestimate the expected utilities and feasibilities of different action options; and finally, she may fail to have a suitable motor program to implement the chosen action tendency (Köster et al., 2021; Moors, 2019). For instance, Sunny may have the goal to maintain high social status because she estimates this to have a high expectancy for being happy. She may be right about this but she may also be wrong. Once the goal for social status is in place, the insulting remark of her colleague Silvester may form a discrepancy with this goal. Sunny may estimate that acting is a good way to reduce this discrepancy, and she may select the specific action tendency to insult him back. Again, she may be right that this is a good way to restore her social status but she may also be wrong. It could be that directing a fixed stare at Silvester is a better option, but since she has never practiced this option before, it is not part of her action repertoire or only at the bottom of the list. It could also be that acting was not worth the trouble and that she would do better to seek social status elsewhere. As this example illustrates, the practical irrationality of goal-directed processes hinges on their theoretical irrationality. Thus, even if goal-directed processes qualify as practically rational in the process-sense, the subjective estimations inherent in them make them vulnerable to errors. This is why goal-directed processes may not always lead to goal satisfaction and may not always be rational in the output-sense.

According to the goal-directed theory, cases of theoretically irrational "emotions" can be understood as cases of apparent or real irrational "decisions." Fear of a house spider, for instance, can be reformulated as the decision to flee, which activates somatic responses preparing for flight and perhaps concomitant facial expressions and feelings. The inaccuracy of this fear can be reframed as the inaccuracy of the decision to flee based on false estimations about the magnitude of the discrepancy of the house spider with one's goal for safety or false estimations of the expectancy that fleeing will procure safety.

But what about recalcitrant emotions? If emotions are decisions, how can a person believe one decision to be right, yet act according to another? Again, the key is to realize that different cycles at the service of different goals may be in competition with one another. The goals of some cycles may be more hidden whereas those of others may be more in the foreground. *Believing* that spiders are not dangerous can be considered as a mental act, and *saying that one believes* that spiders are not dangerous is a verbal act. These acts must be weighed against the physical act of *running away* from the spider. Each of these acts can have a number of benefits and costs. For instance, the mental act of believing that spiders are not dangerous has a high expectancy for the goal to be consistent with one's other knowledge (i.e., an epistemic goal) and the verbal act of saying that one believes that spiders are not dangerous has a high expectancy for the goal to make a smart impression (i.e., impression management goal). Neither of these acts seems to have major costs. The physical act of running away may have a high expectancy for the goal to keep spiders away from bodily orifices (i.e., bodily integrity goal). Even if this behavior comes at the cost of epistemic consistency and impression management, the goal for bodily integrity may have a higher value and win the competition. Another element of importance is that the behaviors in this example – believing, saying, and fleeing – are not technically incompatible with each other. The implication is that as long as believing or saying one thing and doing another does not come with a massive cost, people may not be motivated to solve this inconsistency.

But there are more options from a goal-directed perspective. As suggested earlier, all the representations in goal-directed action cycles can be activated automatically and without assenting to their truth. The representations in some cycles may be truly endorsed whereas those in others may be merely activated. In the case of the spider, a discrepancy with the goal for safety may be activated even if the person does not believe it for a fact or if she believes there is only a tiny chance that she is in danger (not all spiders are harmless) but nevertheless chooses to be on the safe side. A similar story can be told about fear of heights in virtual reality or fear of a shark in a movie theatre. The goal-directed cycle, which was put in place for goal-directed behavior in real life, may be hijacked by salient fictional stimuli and give way to an intense action tendency. The belief that the height or the shark is not real may activate a competing goal-directed cycle, perhaps serving an energy preservation goal, which may dictate the tendency to remain passive. Yet it is possible that this belief lacks the vividness to make the latter tendency win the competition. Even so, an occasional foregrounding of reality may still temporarily succeed in pulling the viewer out of her immersion.

The goal-directedness criterion can easily be reconciled with the notions of ontogenetic and phylogenetic continuity. The assumption that the goal-directed process does not require conceptual or propositional representations makes it able to account for the emotions of babies and animals. Assuming that these beings are limited in the amount of goals they have and the operations they can draw upon to estimate expected utilities, the goal-directed processes that they engage in should be less complex, and so should their emotional lives.

(2) The value criterion – that the goal at stake should have a high value – can account for a host of other apparent properties of emotions. If the value of the goal at stake is high, the discrepancy with the stimulus should have a high magnitude, which requires a big and perhaps urgent reduction. This may translate into an intense action tendency that takes control precedence, followed by an intense mobilization of the physiological apparatus and intense behavior. If all these components are intense, they are also likely to reach consciousness, leaving a salient phenomenological trace. The fact that these episodes pop out in the stream of consciousness probably explains why people started noticing them and invented words to label them. In any case, if this intensity is combined with valence from the detection and selection phases, the heat of emotions is amply secured.

Another downstream consequence of the value criterion – the requirement that the goal at stake should have a high value – is that it will be difficult to come up with another goal that can compete with it. This explains why emotional behavior seems so hard to control. For instance, if a discrepancy with the goal for social status leads to the tendency to fight and the goal for social status is situated at the high end of a person's goal hierarchy, it will be difficult to counteract this tendency because there will be few other goals that can beat this goal. In the goal-directed theory, the uncontrollable nature of emotions does not so much stem from the fact that they arise too quickly to be stopped but more from the fact that they touch on important matters.

FRUITFULNESS

The goal-directedness criterion does a poor job in terms of fruitfulness because the same goal-directed and (in exceptional cases stimulus-driven) mechanisms are put forward to account for emotional and non-emotional episodes. Adding the value criterion does not help much to improve this. As mentioned, the value criterion is identical to the goal relevance criterion in appraisal theories. In Chapter 6, I argued that this criterion qualifies as a descriptive criterion for ranking episodes from more to less emotional but that it may not be deep enough to be considered as a scientific criterion. Nevertheless, it remains an interesting exercise to outline the various factors that influence the value of goals, and/or the intensity of behavior and feelings. This can best be done by abandoning the narrow focus on so-called emotional behaviors such as fight and flight and so-called emotional feelings such as anger and fear, and to broaden the scope to all kinds of intense behavior and feelings. Thus vigorous fighting and vigorously working on the writing of a book may have more in common than meets the eye.

The conclusion that emotions are produced by a general-purpose rather than a special-purpose mechanism is shared by other skeptical theorists such as James (1890b), Duffy (1934, 1941a, 1941b), Meyer (1933), and Russell (2003). James (1890b) differs from the goal-directed theory, however, in that the general-purpose mechanisms he proposed were instincts and habits, not goal-directed processes. Duffy's (1941b) proposal is strikingly similar to the goal-directed theory presented here. Although worked out in less detail, her proposal is also that the goal-directed processes normally found in the non-emotional realm are the very same processes that underlie the phenomena that people call emotional, and that the only difference between emotional and non-emotional phenomena lies in their intensity, which stems from the goal relevance of the events. Following in the footsteps of James (1890b) and Duffy (1941b), Russell (2003; see Chapter 8) also argued that general-purpose instead of special-purpose processes are involved in alleged emotional episodes. Rather than specifying what these processes are, he suggested to hand this task over to each of the domains that can be linked to each of the alleged components of emotions: cognition to cognitive psychology, motivation to motivation psychology, somatic responses to physiologists, and so forth. Duffy (1941b) and Russell (2003) agree that emotion does not rise beyond the status of being a folk concept. Duffy (1941b) probably took the most radical position when she wrote that "[f]or many years the writer has been of the opinion that 'emotion' as a scientific concept, is worse than useless" (p. 283). Her message was that now that psychology had come of age (in the first half of the twentieth century!), the discipline should rid itself of folk concepts such as emotions and replace them with concepts such as energy and direction (see Arnold, 1960, for a critique).

7.3.2 Divisio Definition

7.3.2.1 Criteria for Partitioning and Adequacy

APPARENT-SIMILARITY AND FRUITFULNESS

If the set of emotions is itself not a scientific set, there is no use in searching for the ultimate way to carve it up into subsets. Nevertheless, there may still be merit in figuring out how the goal-directed theory can make sense of vernacular emotion types. This would be a way to meet the apparent-similarity meta-criterion. For instance, anger and disappointment could be linked to the discrepancy between a goal and an actual stimulus, fear to the discrepancy between a goal and an anticipated stimulus, satisfaction to the absence of a discrepancy, joy to the reduction of a previous discrepancy, and relief to the successful reduction of a previous anticipated discrepancy (see also Broekens et al., 2015; Reisenzein, 2012; Rolls, 2005). Further differentiation between anger and disappointment could be achieved, for instance, by saying that in anger, the person tries to repair the discrepancy by choosing to harm another person, whereas in the case of disappointment, she tries to repair it in a non-violent way (Bossuyt et al., 2014a). It must be kept in mind, however, that these attempts do not reify these subsets as scientific subsets. They would only show that the theory has the power to explain the subsets that people take to be emotions. These vernacular subsets, moreover, may not be the only or even the best way to carve up the set. In other words, the subsets may not necessarily be fruitful. Further, there may be other, more fruitful ways of carving up the realm of goal-directed phenomena. For instance, in addition to distinguishing between different types of stimulus-goal discrepancies (pertaining to the detection phase, Step 1),

https://doi.org/10.1017/9781107588882.011 Published online by Cambridge University Press

one could distill four highly abstract tendencies (pertaining to the selection phase in Step 3: the tendency to assimilate, accommodate, immunize, and remain passive – or in doubt about what to do) (Moors, 2017b). I suspect that these distinctions will turn out to be more fruitful in the long run than vernacular emotion categories. This may not only be true within the confines of scientific discourse, but also in therapy. Analyzing people's "emotional" problems not in terms of emotions but in terms of problems in one or more steps of the goal-directed cycle may provide more concrete entry points for initiating change or acceptance. Instead of asking clients to turn inward and get in touch with "how they feel" or to increase their capacity for "emotional granularity" (using finer-grained emotion labels; Barrett et al., 2001; Smidt & Suvak, 2015), therapists could help them discover which goals are dominating their lives and devise less costly ways to achieve these goals (see Moors & Boddez, 2021).

7.4 Validation

Research aimed at testing the goal-directed theory of emotion is still in its infancy, but at least three lines of research seem to emerge.

(1) The first research line starts with a search for instances of "emotional" behavior that are traditionally explained in stimulus-driven terms. The aim is to re-examine these instances to see if they can be explained by a goal-directed process instead. Examining whether behavior is caused by a goal-directed process can be done by measuring values and expectancies for naturally occurring behavior, but ideally it is done in experimental research in which values and expectancies are manipulated and the resulting action tendencies or behavior are measured. The manipulation of values and expectancies can be done between groups and/or across time. If it is done across time, this matches the format of outcome devaluation studies or contingency degradation studies (see Box 2.1).

In a behavioral study, Fischer, Kuppens, and Moors (2020) re-examined the stimulus-driven hypothesis that ostracism leads to antisocial behavior (Leary et al., 2006). In recent years, this stimulus-driven hypothesis was refined by adding as a moderator of this effect the type of goal that is violated by social exclusion (K. D. Williams, 2007). The refined hypothesis stated that social exclusion leads to antisocial behavior if the goal for self-esteem is violated but to prosocial behavior if the goal for belonging is violated. This already suggested a role for the value of the outcome of the behavior. Antisocial behavior may be selected as a means to repair a discrepancy with the goal for self-esteem; prosocial behavior may be selected as a means to repair a discrepancy with the goal for belonging.

https://doi.org/10.1017/9781107588882.011 Published online by Cambridge University Press

But in fact, this refined hypothesis could still be framed as a stimulus-driven hypothesis in which the abstract stimulus feature "incongruence with the goal for self-esteem" ([S]) triggers an antisocial action tendency ([R]) and the abstract stimulus feature "incongruence with the goal for belonging" ([S]) triggers a prosocial action tendency ([R]). The goal-directed hypothesis rejects the idea of a fixed link between the violation of a specific goal and a specific response. The behavior that is selected not only depends on the value of the envisaged outcome but also crucially on the expectancy that the behavior will help reach this outcome. Thus, demonstrating the goal-directed nature of a behavior also requires demonstrating the role of expectancies. To this end, Fischer, Kuppens, and Moors (2020) manipulated the number of behavior options available to socially excluded participants, and hence the relative expectancies of the behaviors for repairing the thwarted goals. One group had the choice between antisocial behavior (i.e., sending aggressive messages), prosocial behavior (i.e., sending friendly messages), and moralizing behavior (i.e., sending moralizing messages) towards their perpetrators, whereas another group could only choose between anti- and prosocial behavior. Results showed that when the moralizing option was present (which had more benefits and less costs than the other options), participants preferred this option and chose less antisocial behavior than when this option was absent. This result cannot be explained by differences in the emotion-eliciting stimulus (both groups were socially excluded), nor by differences in the goals that were thwarted (both groups reported similar violated goals), but rather by differences in the relative expectancies of the available action options.

(2) Behavioral research in which expected utilities of behaviors are manipulated mark an important step towards testing the goal-directed theory. However, studies that measure overt behavior as the dependent variable all suffer from the following disadvantage. They do not allow disambiguating between (a) a default-interventionist scenario in which the behavior was caused by an initial stimulus-driven action tendency but was later corrected by a goal-directed process, and (b) a parallel-competitive scenario in which the behavior was caused by an initial goal-directed action tendency. The second line of research tries to disambiguate between these two scenarios. The crucial difference between them is that each holds a different process responsible for early action tendencies: a stimulus-driven process in the default-interventionist scenario and a goal-directed process in the parallel-competitive scenario. Thus, disambiguating between the two scenarios can only be done by measuring early action tendencies.

One way to measure early action tendencies is by registering MEPs enhanced by TMS (Moors et al., 2019) (see also Box 2.2). Using this method, we pitted (a) the stimulus-driven hypothesis from appraisal theory (see Chapter 6) that *goal-incongruent stimuli that are easy/difficult to control lead to the tendency to fight/flee* against (b) the goal-directed hypothesis that *the tendency to fight/flee is elicited when fighting/fleeing has the highest expected utility*. Participants took part in a multiple-trial computer game in which they were street musicians with their hat lying in front of them to collect money while three thieves could pass by. Participants were first trained to move the right index finger to fight and the left index finger to flee (or vice versa). After that, they learned that one thief was difficult (or rather impossible) to control (i.e., both fighting and fleeing had zero expectancy to prevent a theft), another thief was easy to control by fighting him (i.e., fighting had the highest expectancy to prevent a theft), and still another opponent was easy to control by fleeing from him (i.e., fleeing had the highest expectancy to prevent a theft). Both accounts predicted that the thief who was easy to control by fighting would cause a tendency to fight. The stimulus-driven account predicted this based on the link easy-to-control–fight. The goal-directed account predicted this based on the fact that fighting had the highest expected utility. The two accounts made different predictions regarding the two other thieves. For the thief who was easy to control by fleeing, the stimulus-driven account predicted a tendency to fight (based on the link easy-to-control–fight), whereas the goal-directed account predicted a tendency to flee (because the expected utility for fleeing was highest). For the thief who was difficult to control, the stimulus-driven account predicted a tendency to flee (based on the link difficult-to-control–flee), whereas the goal-directed account predicted a tendency to be passive (because there was no action option with an above-zero expected utility). During a test phase, participants were randomly presented with the three thieves while they had to merely observe them. At 400 ms post-stimulus onset, a single TMS pulse was delivered to the primary motor cortex (M1) and MEPs were continuously measured on the muscles of the "fight" and "flight" fingers. The TMS pulse was used here to boost the action tendency elicited by the stimuli so that it became measurable from the muscles via MEPs. In this way, comparison of the MEP amplitudes for the fight and flight fingers allowed us to infer which action tendency was elicited by the stimuli. MEPs that were higher for the fight finger indicated a tendency to fight whereas those that were higher for the flight finger indicated a tendency to flee. The results provided support for the goal-directed hypothesis (although only when the "fight" finger belonged to the dominant hand). Borrowing Ellsworth's (2013) metaphor again, these results suggest that unlike the appraisal process, the goal-directed process does get us straight to the right twig or the right street address.

(3) A third research line tries to test other, more detailed assumptions of the goal-directed theory, such as the hypothesis that when stimulus-driven and goal-directed processes enter into competition, the goal-directed process is the winner. Fischer, Fini, et al. (2020) installed a goal-directed process by rewarding participants for avoiding positive stimuli and for approaching negative stimuli and they expected this process to compete with a potentially pre-existing stimulus-driven process in which positive stimuli lead to an approach tendency and negative stimuli to an avoidance tendency. Early action tendencies were measured via TMS-enhanced MEPs. Results showed that the goal-directed process determined early action tendencies.

(4) A fourth research line examines the key assumption of the goal-directed theory that goal-directed processes can proceed in an automatic way. The TMS/MEP studies discussed above provide evidence for this assumption. The TMS pulse was administered at 400 ms post-stimulus onset. This suggests that the processing of expected utilities preceding the action tendencies was automatic in the sense of fast. This evidence can be added to the evidence for goal-directed processing in "non-emotional" decision contexts cited earlier (e.g., van Gaal et al., 2012). Future research could further examine the limits of the complexity that can be handled under conditions of automaticity.

(5) A fifth line of research investigates how the building blocks of the goal-directed cycle are implemented in the brain. Research carried out by appraisal researchers about the way in which the brain implements goal relevance and goal congruence (see Sander et al., 2018) can easily be recuperated by the goal-directed theory. In addition, several studies provide evidence that value and risk are separately encoded in the brain and directly feed into action-guiding decision processes (Fiorillo et al., 2003; Grabenhorst & Rolls, 2011; Quartz, 2007; Tobler et al., 2006; Touroutoglou et al., 2020; see Railton, 2017). Pezzulo and Cisek (2016) review brain mechanisms involved in hierarchical feedback loops. Brain research is also conducted in the realm of "emotional" decision processes, involved in flight, fight, and freezing behavior (LeDoux & Daw, 2018).

(6) A sixth line of research re-examines evidence for the influence of emotions on the cognitive functions of perception, attention, memory, judgment, and decision making (Moors & Fischer, 2019). The emotion theories discussed in the previous chapters accept the scientific status of emotions and invoke them as explanantia for changes in these cognitive

functions. They induce positive/negative mood or specific emotions and they measure the influence on these functions (e.g., Brosch et al., 2013; Lerner et al., 2015). The goal-directed theory does not accord a scientific status to emotions, thereby rendering them powerless as explanantia for the observed changes. Instead, the goal-directed theory offers alternative explanations for the findings in terms of discrepancies, values, and expectancies. To illustrate, let us reconsider the influence of *integral* and *incidental* emotions on decision making (see Chapters 2 and 6). To explain why participants in an ultimatum game engage in costly aggression after receiving an unfair offer (e.g., Pillutla & Murnighan, 1996; Sanfey et al., 2003), there is no need to invoke anger as a mediator. A possible goal-directed explanation would be that the unfair offer is discrepant with a participant's goal to be treated fairly, which is more important than maximizing profit, and that aggressive behavior has the highest expectancy for reducing this discrepancy (Moors & Fischer, 2019). Likewise, to explain why participants avoid choosing the high-risk high-reward option during a decision-making task after being submitted to a fear induction procedure (Raghunathan & Pham, 1999; see Chapter 6), there is no need to assume that fear was indeed induced and got carried over to the decision-making task. An alternative, goal-directed explanation would be to assume that the fear induction procedure temporarily increased the value of the goal for safety and/or the expectancy that risk avoidance helps reaching this goal. Instead of fear – or the tendency to appraise stimuli as dangerous or uncertain – getting carried over (Lerner & Keltner, 2001), it might instead be the tendency to evaluate action options in certain ways. Another possible goal-directed explanation is that the action option chosen in the first episode remained highly accessible in the participants' action repertoire and was therefore more easily selected a second time. These reinterpretations bypass the notions of fear and/or appraisals of danger and uncertainty (Moors & Fischer, 2019). Recent research has started to test these and other goal-directed explanations (e.g., Civai et al., 2010; Matarazzo et al., 2020; Turillo et al., 2002).

To summarize, while (evaluation-first) SETs (discussed in the previous chapter) saw stimulus evaluation as the prime source from which every other emotional component flows, RETs (discussed in the current chapter) shifted the weight to response evaluation. Although SETs were key to account for the world-directed Intentionality of emotions, the transition from stimulus evaluation to action tendencies was still a rigid affair. RETs are the first theories that allow for genuine flexibility in the relation between stimulus evaluation and action tendencies. The current chapter worked out the details of my own goal-directed theory as one example of an RET (see Moors, 2017a, 2017b; Moors et al., 2017; Moors & Boddez, 2017, for further reading). This theory proposed that an automatic goal-directed cycle, in which discrepancy detection is alternated with behavior selection, can account for all components that are seen as part of both emotional and non-emotional episodes. Although the difference between both types of episodes, namely, the value of the goal at stake, may not be deep enough to grant the set of emotions scientific status, it helps us to account for a long list of intuitions about emotions (e.g., intensity, control precedence, uncontrollability, apparent irrationality) together with the possibility for mistakes (which explains true irrationality). The theory combines its broad scope with parsimony, in that it postulates a shared system for emotional and non-emotional episodes. Breaking with the long-standing conviction that emotion stands in sharp contrast with instrumental action, the theory helps us appreciate the idea that emotions are instrumental to our most cherished goals and that there is no reason for organisms to switch to a dumb system when these goals are at stake.

BOX 7.1 Emotion Regulation

Emotion regulation has been invoked by evolutionary theorists as one possible explanation for the lack of concordance among components that is sometimes observed. These theorists take dissociations among components as the explanandum and put forward emotion regulation as the explanans. In the current section, emotion regulation is treated as the explanandum. After crafting working definitions in intensional and divisio format, I present constitutive and mechanistic explanations, and try to formulate a scientific definition that demarcates emotion regulation from emotion.

Let us embark on an intensional working definition of emotion regulation. If *regulation* means control, *emotion regulation* means control over an emotion. A person has control over an emotion when three conditions are fulfilled: (a) she has a goal to influence the emotion; (b) the emotion is influenced (i.e., the effect); and (c) the goal and the effect are causally related.[1] The term emotion regulation can be used to refer to the attempt to regulate one's emotion, in which case only the first condition is fulfilled. It can also be used to refer to successful emotion regulation, in which case all three conditions are fulfilled. The first condition does not presuppose the existence of an emotion; a person may want to influence what she *believes* is an emotion. The second and third conditions, on the other hand, do presuppose the existence of an emotion. This entails that people may attempt to regulate their emotions but that they can only be successful in the eyes of vindicator theorists of emotion.

Turning to a working definition in divisio format, different types of emotion regulation can be distinguished based on (a) the agent exerting the control, (b)

https://doi.org/10.1017/9781107588882.011 Published online by Cambridge University Press

types of goals corresponding to types of influence, and (c) types of targets corresponding to types of emotions or emotion components (Gross & Thompson, 2007). To clarify, emotion regulation is intrapersonal when the goal to influence the emotion is held by the emoter; it is interpersonal when the goal to influence the emotion is held by someone else. Examples of interpersonal emotion regulation are when parents try to influence the emotions of their children, and when romantic partners try to influence each other's emotions (Mesquita et al., 2014; Parkinson & Simons, 2012).

The goal to influence an emotion may be the goal to maintain an emotion but typically it will be the goal to change an emotion, either in the sense of upregulation or downregulation. People typically want to upregulate positive emotions and downregulate negative ones, but they can also downregulate positive emotions and upregulate negative ones.

The target of emotion regulation is an emotion. This can be unpacked for multi-componential theories in which an emotion has several components (Moors, 2013c). A person may have the goal to influence her entire emotion, but she may also have the goal to influence only one of the emotion's components. For instance, a person can have the goal to influence her feelings – which is what most people have in mind when they talk about emotion regulation – but she can also have the goal to influence any of the other components. A violin player, for instance, who wants to stop trembling targets the somatic component. A person who wants to conceal her disappointment when receiving a gift targets the expressive part of the motor component. And a person who wants to refrain from hitting her boss targets the coarse part of the motor component (or perhaps even the motivational component).

Let us now turn to a mechanistic explanation of emotion regulation. Most emotion theories contend that emotion regulation relies on a goal-directed process. The goal is to influence the emotion, or one of its components (i.e., the target of regulation), and this requires the selection of a strategy. Emotion regulation strategies can be considered as subordinate goals that are at the service of the superordinate goal to influence the emotion or one of its components. One way to organize these strategies is according to their targets (for other taxonomies see Gross, 1998a, 1998b; Gross & Thompson, 2007; Koole, 2009; R. J. Larsen & Prizmic, 2004; Parkinson & Totterdell, 1999; Quoidbach et al., 2010). Possible targets are again each of the components plus the emotion-eliciting stimulus. Targets of emotion regulation strategies (i.e., subordinate goals) should not be confused with targets of the emotion regulation itself (i.e., superordinate goals). A person can have the goal to change her feelings (i.e., superordinate goal) for which she may select the strategy to change her facial expression (i.e., subordinate goal).

Examples of strategies that target the emotion-eliciting stimulus are ones that try to change the presence of the stimulus either (a) physically, via stimulus selection or problem solving or (b) mentally, via distraction (averting gaze or attention). The strategy to influence the appraisal component, called reappraisal, can focus on each possible appraisal variable. Examples are attempts to (a)

https://doi.org/10.1017/9781107588882.011 Published online by Cambridge University Press

downplay the goal relevance of an emotion-eliciting stimulus by putting things in perspective (e.g., in order to downregulate sadness), (b) overestimate the controllability of an emotion-eliciting stimulus (e.g., in order to downregulate fear), and (c) decrease the external attribution of the emotion-eliciting stimulus (e.g., in order to downregulate anger). The strategy to influence the somatic component may involve administering chemical substances (e.g., pills and alcohol). The strategy to influence the motor component (or motivational component) includes examples such as putting on a happy face, venting by slamming doors or by gossiping, and suppressing the tendency to hit by counting to ten or the tendency to yell by biting one's lip. Finally, the strategy to influence the feeling component can take the shape of suppression or denial in the case of downregulation and of savoring or indulgence in the case of upregulation. Strategies that target the stimulus and appraisal are grouped under the label of antecedent-focused strategies and they are contrasted with response-focused strategies that target responses and feelings (e.g., Gross & Thompson, 2007). The notion of antecedent-focused regulation strategies makes it clear that emotion regulation not only pertains to emotions that are ongoing, but can also pertain to future emotions. Thus, emotion regulation is not limited to changing the intensity of existing emotions but can also conjure up new emotions or nip them in the bud.

A large number of studies have compared the effectiveness of different emotion-regulation strategies (see reviews and meta-analyses in Koole, 2009; Vishkin et al., 2020; Webb, Miles, & Sheeran, 2012). In previous work (Moors, 2013c), I have argued that the effectiveness of a strategy depends on at least two factors: (a) the direct controllability of the stimulus/component that is the target of the strategy (i.e., the subordinate goal) and (b) the extent to which this said stimulus/component in turn influences the component that is the target of the emotion regulation itself (i.e., the superordinate goal). If a person has the goal to improve her feelings and selects the strategy to put on a happy face, the effectiveness of that strategy will depend on whether the person directly controls her facial expression, and if so, whether this in turn influences her feelings. Different emotion theories make different predictions about possible directions of influence among components (i.e., "b"). Research into these influences is also relevant for emotion regulation (see Moors, 2013c). But the entire causal chain is only as strong as its weakest link, and so it is also important to consider the direct controllability of stimuli/components that are the target of a strategy (i.e., "a"). Direct control over stimuli (i.e., solving problems) may or may not be possible. Direct control over appraisal may not always be easy, either because it is automatic or because the goal at stake is too important. Direct control over somatic responses seems possible via the administration of physical substances. Direct control over behavior may seem easy at first sight (e.g., putting on a happy face), but any attempt to change one's behavior may suffer from competition from tendencies to engage in other behavior (e.g., to pout). Finally, direct control over feelings seems virtually impossible, at least if feelings are understood as the passive reflection of the content of representations into consciousness. In fact, this is why people turn to the other strategies in the first place.

But the mechanistic story does not end here. The goal-directed process involved in emotion regulation can be embedded in a broader goal-directed cycle (see Figure 7.3). The goal to regulate an emotion can itself be considered as a subordinate goal that is at the service of a superordinate goal. Examples of superordinate goals are hedonic goals, social goals, performance goals, epistemic goals, and eudaimonic goals (e.g., Niedenthal et al., 2006; Tamir, 2016; Tamir & Millgram, 2017). People may suppress their anger to feel better, to get social approval, to be able to focus better on work, to learn what happens if they don't get angry, and to prove they are in control. The broader goal-directed cycle starts with the detection of a discrepancy between an actual or anticipated emotion and a first goal. For instance, a person is angry or anticipates being angry and this is discrepant with her goal to get social approval. Once a discrepancy is detected, the person develops a second goal to reduce the discrepancy. To achieve this, (a) she can develop a further goal to change (or otherwise influence) the emotion (i.e., assimilation of the emotion to the first goal), but she can also (b) abandon the goal for social approval (i.e., accommodation of the first goal to the emotion), or (c) reinterpret the emotion as being not discrepant with the goal for social approval (i.e., immunization). Which of these broad options will be chosen depends on their relative expected utilities. The person will not try to change her emotion if she expects that her attempts will not be effective. If the person does choose to change her emotion, she still has to select a specific regulation strategy. This selection again rests on weighing up expected utilities of the strategies that the person has in her repertoire. If reappraisal has a higher expected utility than taking a pill, the person will choose to reappraise. Once chosen, the strategy must still be implemented. If this is done and the strategy is successful, the emotion changes. If not, the emotion may still change spontaneously, but it may also remain unchanged. The new state of the emotion is thereafter fed back as the input into a new cycle where it is again compared with the first goal. By embedding the emotion regulation process within a broader goal-directed cycle, it becomes clear that adaptive emotion regulation not only depends on the effectiveness of the chosen regulation strategy, but also on the detection of the need to regulate the emotion, on the presence of effective regulation strategies in the person's repertoire, and on the timely implementation of the chosen strategy (see also Gross, 2015; Tamir et al., 2020; Webb, Schweiger Gallo, et al., 2012).

The received view, consistent with classic dual-system logic, is that the goal-directed processes involved in emotion regulation are non-automatic, but that extensive training can turn these goal-directed processes into habits, which are automatic (e.g., Eder, 2011; Ekman, 1972; Gyurak et al., 2011; Hopp et al., 2011).[2] The alternative dual-process view that I defend in Box 2.1, however, rejects the idea that extensive training turns goal-directed processes into habitual ones even if it does make them more automatic. Regardless of the underlying process, empirical evidence for automatic emotion regulation is growing (e.g., Bargh & Williams, 2007; Eder, 2011; Gross, 1999; Jostmann et al., 2005; Mauss, Bunge, & Gross, 2007; Mauss, Cook, & Gross, 2007; L. E. Williams et al., 2009).

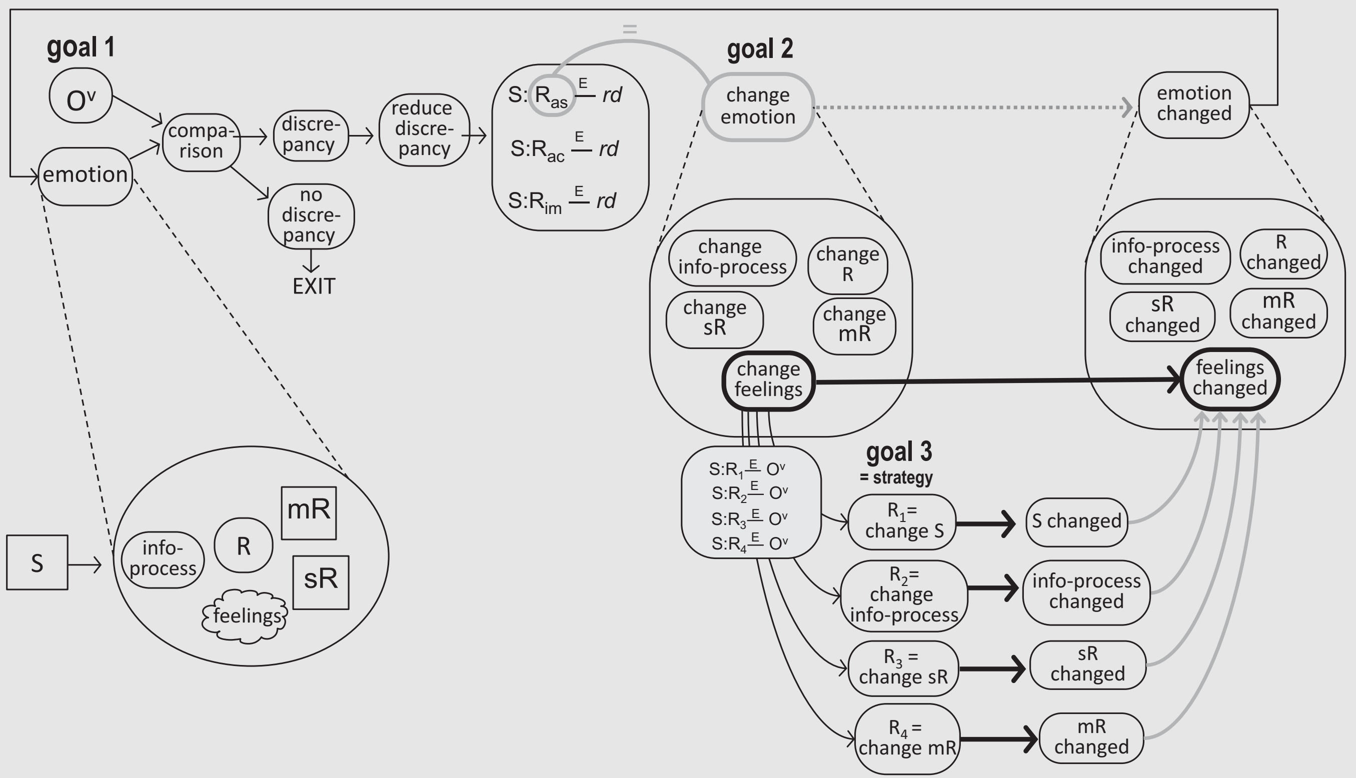


Figure 7.3 Emotion regulation

Moving on to the scientific intensional definition of emotion regulation, one question that has puzzled many scholars is whether and how emotion regulation can be demarcated from emotion (see Frijda & Mesquita, 2007; Gross & Barrett, 2011; Kappas, 2008, 2011). For vindicator theories, the answer is simple. Emotion relies on a special-purpose stimulus-driven mechanism whereas emotion regulation relies on a goal-directed mechanism. Even if extensive training turns the goal-directed mechanism into a habitual – and hence stimulus-driven – process, it is still different from the stimulus-driven process involved in emotion.

But what about skeptical theories such as the goal-directed theory? At first blush, the idea voiced in the goal-directed theory that both emotion and emotion regulation rely on a goal-directed process suggests that the two phenomena do not differ from each other. This can be expressed in two ways. First, emotion can be considered as a form of *regulation*: In an emotional episode, an organism tries to control the environment by selecting action tendencies to resolve a discrepancy. For instance, a spider phobic flees from a house spider in an attempt to reduce the discrepancy that the spider poses for her goal for physical integrity. Second, emotion regulation can be considered as a form of *emotion* (see also Campos et al., 2004; Frijda & Mesquita, 2007). The goal to influence an emotion is itself an abstract tendency that stems from the detection of a discrepancy. For instance, the fear of the house spider may pose a discrepancy with one's goal for social approval, and this may prompt the tendency to suppress one's fear.

Despite the similarity in the mechanisms involved in emotion and emotion regulation, however, I believe it is useful to keep a conceptual separation between these two phenomena on the grounds that they each have a different target. Emotion is a form of control that takes the *stimulus* as its target whereas emotion regulation is a form of control that takes the *emotion* as its target. Emotion could indeed be considered as a form of regulation but it would qualify as *stimulus* regulation rather than emotion regulation. Conversely, emotion regulation could also be considered as a form of emotion but it would qualify as an emotion about an *emotion* (a meta-emotion) whereas emotion in the normal sense is emotion about a stimulus.

Moreover, keeping two concepts in the loop also helps distinguish emotion regulation from emotion competition. In emotion regulation, the emotion cycle is hierarchically nested in the emotion regulation cycle. In emotion competition, on the other hand, the same stimulus gives rise to two competing action tendencies. This could be because the stimulus is discrepant with two different goals and solving their discrepancies may call for two different actions. For instance, social exclusion can be discrepant with one's goal to maintain status, which one may try to reduce via aggressive behavior, but at the same time, it can also be discrepant with one's goal for social inclusion, which one may try to reduce via friendly behavior (see Fischer, Kuppens, & Moors, 2020). The competition may end in one of the two emotions winning the competition, or in some kind of compromise.

Of course, it is not because I recommend keeping separate concepts in the loop, that I am attached to the concepts of "emotion," "emotion regulation," and "emotion competition." Given that I do not grant scientific status to "emotion" in the first place, better terms might be "first-order regulation," "second-order or meta-regulation," and "regulation competition."

[1] This definition of control requires an entity that can have goals (see also Moors & De Houwer, 2006a, footnote 7). Theorists who define control in the broader sense as mere causation or influence have widened the meaning of emotion regulation to also include emotion self-regulation, where emotion regulates itself via behavior that is part of or a consequence of the emotion (e.g., if fear leads to avoidance, it cancels itself) (e.g., Carver & Scheier, 2013). I do not consider this to be a form of emotion regulation, but as the unregulated dynamics of emotion over time. By asking for an entity that can have goals, I do not imply that the goals must be self-chosen. They may also be implanted in an organism by evolution or design. In that case, they would nevertheless count as the entity's own goals (see Bechtel, 2008; for a contrasting view, see Wiegman, 2020).

[2] Both Gyurak et al. (2011) and Hopp et al. (2011) suggest automatic goal-pursuit as a mechanism underlying automatic goal regulation. However, on closer inspection of their proposals, it seems that the heavy lifting is still done by habitual processes. Gyurak et al. (2011) and Eder (2011), moreover, refer to automatic goal-pursuit in the form of implementation intentions. The common view, however, is that implementation intentions are temporally set up [S–R] links (but see Moors et al., 2017, p. 312; Sheeran et al., 2005).

https://doi.org/10.1017/9781107588882.011 Published online by Cambridge University Press

CHAPTER 8

Psychological Constructionist Theories

8.1 Precursors

James's (1890b) theory is an important precursor of many theories including psychological constructionist theories (henceforth, PCTs). To reiterate, James (1890b) proposed that after a stimulus is perceived (Step 1), it activates an [S–R] link (instinct or habit) (Step 2). This in turn produces bodily responses (Step 3). The bodily responses are again perceived and hence felt (Step 4). These bodily feelings are the emotions.

Critics complained that James's (1890b) bodily feelings were not sufficient to account for the variety of emotions in terms of discrete emotion types. Failed attempts to empirically identify visceral response patterns for specific emotions led to the conviction that bodily feelings are undifferentiated (see Reisenzein, 1983). The complaint was not so much that all bodily feelings are the same, but rather that their variety did not map onto the vernacular emotion types that some theorists wanted to distinguish. In other words, the differentiation between them was unruly or diffuse. Thus, while James's (1890b) bodily feelings delivered the quantity of emotions, they were unable to deliver the emotion-specific quality of these emotions. Several theories proposed adjustments to James's (1890b) theory with the aim of solving this presumed lacuna. Two theoretical families proposed adding a cognitive process into the mix. SETs – at least the evaluation-first brand – inserted a stimulus evaluation process in between the stimulus and the [S–R] link (James's Step 1). PCTs, on the other hand, added a construction process near the end (after James's Step 4), in which diffuse bodily feelings are interpreted in terms of specific emotions (e.g., fear) by binding them to a stimulus (e.g., tiger). The processes proposed by these two theories not only differ in terms of their timing but also in terms of their inputs and outputs. In the stimulus evaluation process in SETs, the input is an external stimulus or the concrete representation thereof (e.g., tiger) and the output is an appraisal or evaluative property, which is an abstract representation of the stimulus (e.g., danger). In the construction process in PCTs, the input consists of the conscious representation of an internal stimulus (i.e., bodily feelings) and the representation of an external stimulus (e.g., tiger) and the output is the conscious representation of a specific emotion (e.g., fear).

The solution proposed by PCTs was presaged by the early theories and empirical work of Wundt (1897/1998), Marañon (1924), Dunlap (1932), Harlow and Stagner (1932), Young (1943), Ruckmick (1936), and Hunt (1941) (see Gendron & Barrett, 2009, for a historical overview). I will limit my discussion to the well-known PCTs of Schachter (1964), Barrett (2006b, 2012, 2017b), and Russell (2003, 2009, 2012), in that order.

8.2 Constitutive and Mechanistic Explanations

PCTs reject the view of evolutionary theories that the phenomena called emotions are caused by special-purpose mechanisms. Instead, they take them to be caused by the combination of two or more general-purpose mechanisms (**Axis A**: two vs. more). The PCTs of Schachter (1964) and Barrett (2006b) are two-factor theories: A first factor consists of diffuse bodily responses (i.e., somatic component; sR), the immediate outcomes of which are consciously perceived (iS → [iS]) and hence felt (i.e., raw feeling component). A second factor is a construction process (i.e., cognitive component) that binds these felt bodily responses to perceived external stimuli ([eS]) and produces the conscious representation of a discrete emotion (i.e., labeled feeling component; [iaS]). For example, the encounter with a crouching tiger vs. a cherished person leads to diffuse bodily feelings, which when bounded to the crouching tiger vs. the cherished person, turn into representations of fear vs. love. Russell's (2003) PCT is more appropriately described as a multi-factor theory. In addition to diffuse bodily feelings and the construction process, a host of other components can figure in the phenomena called emotions.

But the difference between Schachter's (1964) and Barrett's (2006b) theories, on the one hand, and Russell's (2003) theory, on the other hand, is more profound. Schachter (1964) and Barrett (2006b) believe that the conscious representation of a discrete emotion, which is the output of the construction process, *is* an emotion. Russell (2003, 2009, 2014), on the other hand, believes that the conscious representation of a discrete emotion is just what it is: a conscious representation *of* an emotion, not an emotion per se (**Axis B**). Thus, Schachter (1964) and Barrett (2006b) believe that once the bodily feelings are constructed as fear, the person *has* the emotion of fear. Russell (2003, 2009, 2014), on the other hand, believes that once the bodily feelings are constructed as fear, the person merely *perceives herself as having* the emotion of fear. And if this perception is conscious, she experiences herself as having this emotion. This is a meta-experience.

Schachter (1964) and Barrett (2006b) identify emotions with emotional experience. This means that they hold a narrow constitutive explanation, in which emotion is restricted to the feeling component. Russell (2003) identifies the phenomena called emotions with the entire emotional episode, including, but not restricted to, stimulus evaluation, action tendencies, somatic and motor responses, and feelings (**Axis C; Axis 5c**). Both Schachter (1964) and Barrett (2006b, 2017b) do pay some attention to the behavior that follows on the construction process. Russell (2003, 2009) considers behavior that occurs both prior to and after the construction process (**Axis D**). Other differences between the three theories will be highlighted below.

8.2.1 Schachter

In Schachter's (1964; Nisbett & Schachter, 1966; Schachter & Singer, 1962) theory, the first factor is undifferentiated physiological activity called "arousal" and this arousal is felt. The construction process in the second factor interprets the arousal in the light of external stimuli, resulting in the labeling of the arousal in terms of a specific emotion. This labeled emotion is again felt. According to one popular reading of the theory, the construction process is more precisely understood as a causal attribution process (see Reisenzein, 1983). Attribution of arousal to a tiger vs. a cherished person, for instance, generates the labels of fear vs. love.

The fact that the arousal in the first factor is part of the input of the construction process in the second factor suggests that the two factors proceed in a sequential manner. At the same time, Schachter (1964) emphasized that the arousal of the first factor and the labels that are part of the second factor have separate contributions that combine in a multiplicative way. If either of them is lacking, there is no emotion. In addition, the first factor takes care of the quantity of the emotion while the second factor takes care of its quality.

The fact that both factors have separate contributions, however, does not imply that they must have completely independent causal origins. Nisbett and Schachter (1966) distinguished between a typical and an atypical route (see Figure 8.1(a) and (b)). In the typical route, a stimulus is first evaluated (e.g., as dangerous). This evaluation causes both the arousal and it suggests an emotion label (e.g., fear) to be used in the construction process. Here the two factors share the same causal root: the evaluation of a single stimulus. The stimulus that causes the arousal is immediately recognized by the person as the source of the arousal.

In the atypical route, arousal is caused by internal activity or by a stimulus that is either not in the focus of attention or that does not explain the arousal well enough (e.g., because the arousal is too intense). The

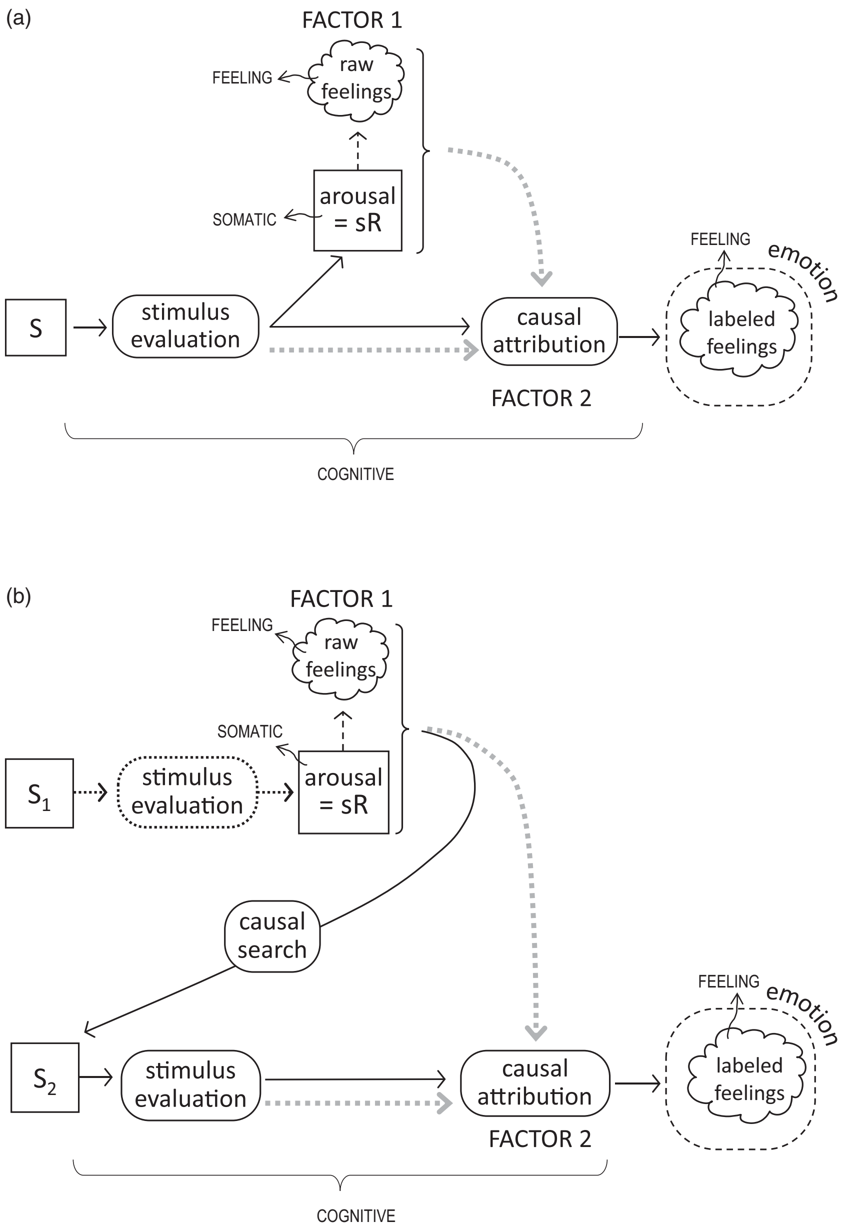


Figure 8.1 Schachter's theory: (a) typical path; (b) atypical path with misattribution

https://doi.org/10.1017/9781107588882.012 Published online by Cambridge University Press

need to understand this arousal instigates a search for its cause. This search may end in the correct attribution of the arousal to the stimulus that did cause the arousal or in the misattribution to another, salient stimulus. After a cause is found, evaluation of this cause again yields a label, which fixes the identity of the emotion. Although the latter route is the atypical route, it nevertheless formed the inspiration for subsequent PCTs.

It is tempting to think that Schachter's (1964) typical route corresponds to the mechanism proposed in appraisal theories (e.g., Deonna & Scherer, 2010). After all, both theories assume the operation of a stimulus evaluation process. In appraisal theories, however, stimulus evaluation determines the quality of the other components and hence of the emotion. In Schachter's (1964) typical route, on the other hand, stimulus evaluation produces a somatic component that makes no contribution to the quality of the emotion, and an emotion label that completely determines this quality. The crucial difference between these two theories is that Schachter (1964) held that there is no emotion-specific quality prior to labeling, whereas appraisal theories hold that appraisal already produces emotion-specific quality prior to labeling and hence that labeling is optional.

8.2.2 Barrett

Barrett's (2006b, 2012, 2017b) PCT stands on the shoulders of Schachter's (1964) PCT to which she has added a growing number of adjustments over the years. In the early version of her theory (Barrett, 2006b), the two factors can easily be recognized. The first factor, however, is not restricted to arousal (i.e., intensity) but also includes valence (i.e., positive/negative quality). The combination of arousal and valence is called "core affect" (see also Russell, 2003, 2005; Russell & Barrett, 1999). Core affect is neurophysiological activity reflected in the conscious experience of arousal and valence. Core affect is ubiquitous. We always feel a certain degree of activation and we always feel good or bad or something in between. Core affect can be caused by external stimuli (eS) such as failing an exam or the weather, but also by internal stimuli (iS) such as pain, fatigue, hormones, or diurnal cycles (Thayer, 1989). Failing an exam and being in pain may produce high arousal and negative valence; bad weather and fatigue may produce low arousal and negative valence, and so on. Barrett (2006b, p. 31) mentioned as the mechanism that translates external stimuli into core affect, an automatic appraisal process in which stimuli are evaluated as more/less goal-relevant (yielding arousal) and goal-in/congruent (yielding valence). Internal stimuli, on the other hand, can be translated into core affect via purely physical

https://doi.org/10.1017/9781107588882.012 Published online by Cambridge University Press

mechanisms. More recently, however, Barrett (e.g., Hutchinson & Barrett, 2019, p. 281) seems to have banished appraisal as a causal mechanism for core affect, so that it is now only a matter of the body (i.e., caused by internal stimuli; iS). This is reflected in the fact that the term core affect is used interchangeably with the term physical sensations. Nevertheless, these physical sensations are still characterized by arousal and valence.

The construction process in the second factor is a categorization process, also referred to as a "conceptual act." Barrett's (2006b) categorization process is similar to Schachter's (1964) attribution-and-labeling process in terms of the pieces of information that go into it. The categorization process takes as its input the core affect from the first factor together with category knowledge tailored to the immediate situation and produces as its output the representation of a specific emotion. The categorization process itself proceeds automatically but the output is conscious, and hence felt. Like in Schachter's (1964) theory, this feeling *is* the emotion.

Barrett (2006b) adds that the categorization of one's own internal state as an emotion is similar to the categorization of another person's external state as an emotion. The difference lies in the input: In self-categorization it is core affect; in other-categorization it is the other person's observable behavior. Emotions are in the eye of the beholder. Organisms that are not able to self-categorize as being in an emotion, such as young children and non-human animals, can only have an emotion from a third-person perspective.

The categorization process is presented as a form of perception that is influenced by category knowledge. Barrett (2006b) clarifies this insight by drawing an analogy with color perception. The eyes receive light from various wavelengths. Although the spectrum of wavelengths is a continuum, we perceive discrete colors (red, green, yellow, blue) depending on how we have learned to parse the continuum. Emotion perception works in the same way. Whether we perceive our core affect as anger or fear depends on category knowledge.

An emotion category is a population of context-specific emotion scripts stored in memory.[87] For instance, there could be different scripts for anger in a competition context, anger in traffic, anger during a conflict at home, and anger about global injustice. Each script contains information about a specific type of situation in which a discrete emotion label was used to categorize an instance of core affect, together with

[87] Barrett (2006b, 2022) distinguished between emotion categories in the world and emotion categories in people's minds, using the term "emotion concepts" to refer to the latter. For ease of communication, I use the term emotion category for both.

https://doi.org/10.1017/9781107588882.012 Published online by Cambridge University Press

information about the specific action tendencies and behavior that occurred, and specific ways in which to regulate the emotion. In addition to their context-specificity or situatedness, the representations in the script have a multimodal sensory and embodied format (Wilson-Mendenhall et al., 2011). Novel instances of core affect can be categorized via the retrieval (i.e., an associative operation) of these stored emotion scripts. Activation of the motor parts leads to their partial re-enactment or simulation.

Although the high-level description presented above of Barrett's (2006b) categorization process as one that takes core affect as its input and produces a specific emotion as its output suggests that core affect is logically prior to categorization, Barrett (2006b) explicitly resisted a sequential handling of the two factors. Core affect and categorization run in parallel and they constrain each other mutually as they evolve over time until a stable "emotional solution" or Gestalt emerges. The principles of parallel-constraint-satisfaction (Kunda & Thagard, 1996) at work here are compatible with the principles of the DS approach (see also Chapters 5 and 6): Both factors influence each other in parallel in iterative cycles. Barrett (2006b), moreover, advocated a strong form of emergence, which entails that the whole is more than the sum of the parts. Just like H_2O has qualities (liquid) that H and O do not have on their own, the emotion has qualities that cannot be reduced to the qualities of the components.

It may be noted that the simultaneous handling of the two factors also suggests a shift in thinking about the factors themselves. Instead of seeing core affect as one factor and the construction process as another, Barrett (2012) started treating the construction process as the overarching process that draws on various sources of information such as core affect (i.e., internal stimuli), external stimuli, and category knowledge. It is these three sources that can be regarded as the factors that mutually constrain each other *within* the construction process. This entails a complex associative operation. Such a mechanism is in fact highly reminiscent of the mechanism proposed by network theories in Chapter 5. Indeed, the activation of a script can be considered as the activation of a network. Two differences are perhaps that in Barrett's (2012) theory (a) each emotion does not correspond to a single network but rather to a collection of networks and (b) each network is simultaneously activated via multiple entries: bodily responses activating the response nodes and an external stimulus activating the stimulus node.

The categorization of core affect as emotion is supposed to have several consequences. First, it generates meaning or Intentionality. Although core affect is caused by a stimulus, it is not experienced as about anything until it gets categorized as an emotion. "[D]uring the categorization process, core affect becomes bound to the object that we believe to have caused the affective state in the first place" (Barrett, 2006b, p. 36). Although Barrett

(2012) robbed the appraisal process of its causal powers, she still left a role for appraisal as a description of the content of emotional experience. This accords with "purely descriptive" appraisal theories (Clore & Ortony, 2008, 2013; Ortony & Clore, 2015), which consider appraisal patterns to be descriptions of situations as well as descriptions of the content of emotional experience, but which are agnostic about how the features of the situation end up in the emotional experience. These appraisal theories do not count as SETs strictly speaking because they do not postulate a stimulus evaluation process either as a cause or a component of the emotion.

A second consequence of categorization is that it allows the person to infer how to act, as the script suggests ways for dealing with the stimulus as well as with one's own emotional state. A person who perceives herself as angry vs. disappointed in a competition context, for instance, may engage in retaliation vs. withdrawal. In addition, a person may also be treated differently by others depending on whether she is perceived as angry vs. disappointed in this context. Angry competitors may be excluded whereas disappointed competitors may receive consolation. So once physical changes are categorized as emotions, they acquire social functions they did not have before (Barrett, 2012). Here, Barrett (2012) draws an analogy with the categorization of plants as weeds or flowers: Weeds are eradicated whereas flowers are cultivated.

Recently, Barrett (2017b) has expanded and sharpened her theory in two ways (see Figure 8.2). A first novel element is the presentation of the construction process as an active inference process guided by prior predictions (Barrett, 2017b; Barrett & Finlay, 2018; Barrett & Simmons, 2015; Barrett et al., 2015; Hutchinson & Barrett, 2019; Mobbs et al., 2019). This is modeled after predictive processing theories of cognition and action (A. Clark, 2013; Friston, 2003). It may be noted that while other scholars already applied predictive processing principles to explain affect (i.e., Factor 1; e.g., Cunningham et al., 2013; Seth, 2013; Seth et al., 2012; Van de Cruys, 2017), Barrett (2017b) applies it more to the construction process (i.e., Factor 2; see Velasco & Loev, 2020, for a comparison). The basic principle of predictive processing theories is that the system makes predictions about incoming input. When a prediction is compared with actual input and is found to deviate from that prediction, a prediction error is generated. The prediction error can be reduced by adjusting the prediction to better match the input (i.e., assimilation) or by engaging in behavior to make the world more conform to one's predictions (i.e., accommodation) (Barrett & Simmons, 2015).[88] So instead of seeing the

[88] The cycle in predictive processing theories is reminiscent of the goal-directed cycle in the goal-directed theory (Chapter 7). However, the crucial difference is that in the former theories, the stimulus at the start of the cycle is compared to an expectation or

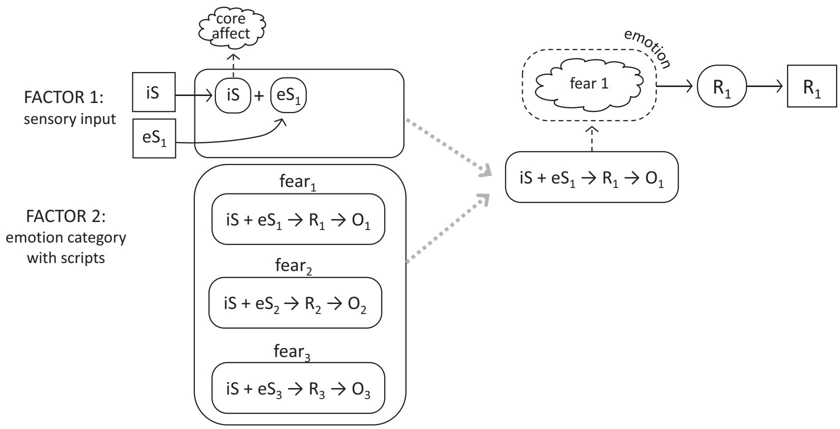


Figure 8.2 Barrett's theory (recent version)

categorization process as something that comes into force after the sensory input (i.e., core affect and the stimulus) is registered, the sensory input is already anticipated or predicted. This is possible because the system is equipped with emotion categories. Once the sensory input occurs, it is compared to the predicted input. An important point to note, however, is that an emotion category (e.g., anger) is still (as before) conceived of as a population of context-specific scripts. This entails that the comparison process operates to reduce the possible scripts to the one that best resembles the current sensory input.

A second novel element has to do with the content of the emotion scripts and the basis for their individuation (Barrett & Finlay, 2018). Each individual emotion script is said to contain information about a motor action that was executed, the sensory input causing this action, and the sensory outcome of this action. This entails that prediction in this theory takes the form of a series of hypotheses or guesses about which actions to take and what their causes and sensory outcomes will be. Thus, settling on a specific member of an emotion category enables the brain to make inferences about which action to take and how to label the emotion. The members (scripts) of an emotion category (e.g., anger) are individuated on the basis of their function or action goal. "[I]nstances of anger can be

prediction (i.e., representation of an expected outcome, which has a mind-to-world direction of fit), whereas in the latter theory, it is compared to a goal (i.e., representation of a valued outcome, which has a world-to-mind direction of fit). Expectations and goals can be seen as orthogonal: A person can expect things that she wants or does not want, and she can want things that she does or does not expect (see Moors et al., 2021).

https://doi.org/10.1017/9781107588882.012 Published online by Cambridge University Press

associated with the goal to overcome an obstacle (particularly when the obstacle is another person), to protect against a threat, to signal social dominance or appear powerful, to affiliate and repair social connections, to enhance performance to win a competition or a negotiation, or to enhance self-insight" (Hoemann et al., 2019, p. 1833). Given this within-category variety, the question arises what holds the various members of any particular emotion category (e.g., anger) together. In Barrett's (2017b) PCT at least, there seems to be nothing else left than the emotion word. Children learn to group scripts together under the denominator of anger, because adults have labeled them in this way.

8.2.3 Russell

Russell (2003, 2005, 2006, 2009, 2012, 2014, 2017) adopts a skeptical approach. Rather than trying to explain the phenomena called emotions with the aim of vindicating them, he expects that an explanation of these phenomena will reduce them to other phenomena and in this way effectively explain them away. An alleged emotion is an episode that includes various components such as core affect, all kinds of information processes (e.g., perception, appraisal, attribution, thoughts), action tendencies, physiological responses, subtle and coarse behavior, and emotional meta-experience (see Figure 8.3). Core affect is only one component in this episode next to, or perhaps preying on, the other components. In Russell's (2009, p. 1267) view, psychological construction means that the components in each token emotional episode "are cobbled together on the fly." Each component is generated by its own set of mechanisms, forming a mini-theory of their own. There is not a single underlying mechanism that glues them together. It is only if we zero in on the component of emotional meta-experience – the experience of having an emotion – that

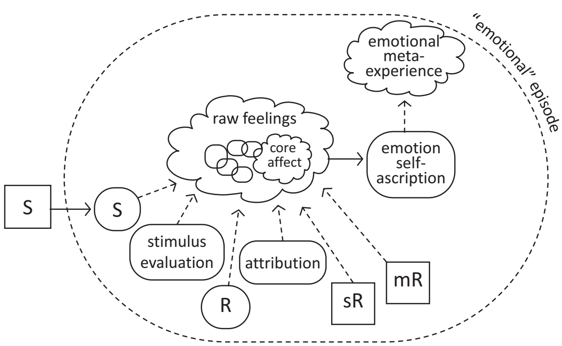


Figure 8.3 Russell's theory

https://doi.org/10.1017/9781107588882.012 Published online by Cambridge University Press

the two-factor idea of Schachter (1964) can be recognized. This experience results from the categorization of all current components in terms of a specific emotion. This categorization process does not go on all the time (as per Barrett, 2006b; Lindquist & Barrett, 2008) but is typically prompted by a salient change in core affect or when people are asked how they feel. The categorization process is based on recognizing the resemblance between one's current state and the mental script for a particular emotion (Fehr & Russell, 1984). This mental script takes the form of a prototype. Resemblance with the prototype is a matter of degree and a threshold must be passed before an emotional meta-experience occurs. No single component counts as a strictly necessary part of this script, not even a change in core affect. An example of a script without a change in core affect would be the script for cold anger (Russell, 2005). An important point to reiterate is that Russell (2003, 2014) does not take the result of this categorization process to be an emotion but a meta-experience. If a person has the experience of having an emotion, all that she has is the representation of the experienced aspects of all current components categorized as an emotion or a specific emotion, and this representation can itself be conscious and hence be experienced.

8.3 Scientific Definitions

8.3.1 Intensional Definition

8.3.1.1 Criteria for Demarcation

The constitutive explanations of Schachter (1964) and Barrett (2006b) identify emotion with the experience of a specific emotion or labeled feeling. This yields a first necessary criterion, which I call the "labeled feeling" criterion. The causal-mechanistic explanations in both theories, moreover, entail that this experience results from a mechanism that combines two factors: (a) the presence of diffuse bodily responses projected into bodily feelings (i.e., the umbrella term for Schachter's arousal and Barrett's core affect) and (b) a construction process that binds these bodily feelings to external stimulus input and produces the representation of a specific emotion. This yields two additional necessary criteria, which I call the "bodily feeling" criterion and the "construction" criterion, respectively.

A few remarks about the bodily feeling criterion are in order. First, in Schachter's (1964) theory, bodily feelings were understood as conscious representations of actual bodily responses. Later theorists left the possibility open that actual bodily responses were not necessary but could be replaced by information about such responses (presented via false feedback; Valins, 1966) or that they could be mentally simulated (Barrett et al., 2015; MacCormack & Lindquist, 2017; see also James, 1890b; Damasio,

https://doi.org/10.1017/9781107588882.012 Published online by Cambridge University Press

1994; J. J. Prinz, 2004a). Second, in Schachter's (1964) theory, the bodily feeling criterion must be qualified in the sense that it involves a noticeable *change* in bodily feelings (arousal). This does not seem to be an explicit requirement in Barrett's (2006b) theory. Third, in Barrett's (2006b) theory, the bodily feeling criterion is required for a person to *have* an emotion, but not for a person to *perceive* an emotion in others. In the latter case, the categorization happens on a different input, namely the behavior of others.

Russell (2003) does not subscribe to any of these criteria. He maintains a skeptical position, according to which there simply are no necessary and sufficient criteria for an episode to count as an emotional episode. Once the components in an alleged emotional episode are effectively explained, there is nothing left to explain (see Sauter & Russell, 2022). In the next sections, I evaluate the adequacy of these proposals in terms of the meta-criteria of apparent-similarity and fruitfulness.

8.3.1.2 Adequacy

APPARENT-SIMILARITY

I first evaluate the candidate scientific definitions of Schachter (1964) and Barrett (2006b). The bodily feeling criterion, which stems from bodily responses, already accounts for two desiderata: the bodily aspect of emotions and part of their phenomenal aspect. The arousal and/or valence dimensions of these bodily feelings, moreover, combine to deliver the heat of emotions. The labeled feeling criterion adds further specific quality to the phenomenal aspect.

The construction criterion furnishes the world-directed Intentionality of emotions. As mentioned, the bodily feelings are objectless but bind to an object when they get interpreted as a specific emotion (Barrett, 2006b). Deonna and Scherer (2010, p. 46) objected that since the categorization of a person's state is optional, the Intentionality must be optional as well. Thus, they argue, "if, to produce the experience of strong feelings, it is sufficient to create strong arousal (by exercise or drugs) and provide environmental cues that suggest an emotional state to justify an otherwise incomprehensible arousal, then the [I]ntentional object is certainly not required, but superfluous or at best optional." This ties in with Mandler's (1975) characterization of Schachter's (1964) theory as a jukebox theory: "It suffices to drop a coin into the slot to get the machine running (producing arousal) and that the individual would then choose the most appropriate record by construing meaning from situational cues" (p. 46). Hoemann et al. (2019, p. 1841) insisted, however, that the category is not arbitrarily chosen but is instead constrained along cultural lines.

Barrett's (2006b) proposal that categorization is a form of perception should, at first blush, yield all the benefits (and costs) of the perceptual

https://doi.org/10.1017/9781107588882.012 Published online by Cambridge University Press

theories discussed in Chapter 6. On second thought, however, a few important differences can be noted. In normal perception, the object exists independent of the observer (Pober, 2018). The same goes for emotion perception in perceptual theories (at least those with an externalist view; e.g., Deonna, 2006). According to these theories, danger and offense can be objectively determined. In Barrett's (2006b) categorization process, on the other hand, the object (i.e., the emotion) gets created. Given that there is no objective reality against which the categorization can be evaluated, the categorizer can never be said to be wrong. This has led critics to conclude that the categorization process is unable to account for the intuition that people can have the wrong emotion, that is, cases of theoretically irrational emotions (e.g., Deonna & Teroni, 2012; Pober, 2018). A possible way out for Barrett (2006b), however, would be to argue that theoretical irrationality is only an apparent property of emotions, and that so-called cases of inaccurate emotions are simply cases in which the categorization of the emoter dissociates from the categorization of outside observers. Hoemann et al. (2019, p. 1840) seem to accept this implication as they declare that an emotion cannot be objectively right or wrong but only more or less appropriate according to cultural rules (see De Leersnyder et al., 2014).

What about practical rationality? Barrett (2006, 2012, 2017b) holds that categorization influences the type of behavior that emoters generate. In the early version of the theory, the mechanism underlying the transition from categorization to behavior was barely specified. In the newer version, the brain makes a number of guesses for how to deal with the situation. After the sensory input is in, the brain settles on one of these guesses and suggests the best way for action. The term "best" suggests that the brain produces adaptive behavior. Although the theory explains how actions get selected – the script that most resembles the sensory input gets activated – it is unclear how the system makes sure that this action is indeed the best. The scripts do contain information about sensory inputs, responses, and sensory outcomes, but it is unclear whether this also includes information about response-outcome expectancies and the values of outcomes, including non-sensory outcomes. As long as these two elements are missing, the process proposed by Barrett (2017b) does not qualify as a goal-directed process in the strict sense of the term used here, and lacks the adaptive power that comes with such a process.

An often raised objection against Barrett's (2006b) theory is that it cannot account for the intuition that infants and animals also feel emotions, given that they do not possess the emotion categories required to do so (i.e., they do not meet the construction criterion) (Deonna & Scherer, 2010; Deonna & Teroni, 2012; Pober, 2018). Several replies are possible. In an earlier version of the theory, the claim was indeed made that infants and animals do not experience emotions but can only be said to have

https://doi.org/10.1017/9781107588882.012 Published online by Cambridge University Press

them from a third-person perspective. Thus, the first reply would be to argue that ontogenetic and phylogenetic continuity are merely apparent properties. In a later version of the theory, the requirement for the possession of linguistic emotion labels was softened, and it was argued that infants and animals do already possess proto-conceptual primitives (Hoemann et al., 2019). These primitives correspond to the various situation-specific scripts discussed earlier. Infants and animals can form ad hoc scripts to give meaning to the situation and to predict ways for responding to them without labeling them with words.

Turning to Russell's (2003) theory, consideration of the issue of apparent-similarity may need to be postponed until the mini-theories for all the components have been fully developed. That being said, several component-specific theories already exist and continue to develop. For instance, there is plenty of work on general-purpose mechanisms involved in facial movements (e.g., Fernández-Dols, 2017; Fridlund, 1994; Sauter & Russell, 2022; Scarantino, 2017b), physiological activity (Gendolla & Richter, 2010; Kreibig, 2012; J. T. Larsen et al., 2008), action tendencies and behavior (LeDoux, 2012b; LeDoux & Daw, 2018; Moors et al., 2017; Pezzulo & Cisek, 2016), and consciousness (Frijda, 2005, 2009; Lambie & Marcel, 2002; LeDoux & Brown, 2017; LeDoux & Hofmann, 2018).

Note that my own goal-directed theory (Moors, 2017a) can easily be fitted in the program set out by Russell (2009) as it proposes a theory of behavior causation that is general-purpose (not specific to emotions) and shares the same skeptical outlook. The only point of disagreement perhaps is that Russell (2005, 2009) portrays the different components in alleged emotional episodes as more loosely connected than I do (see also Moors, 2012). In the goal-directed theory, all components are causally related rather than being explained by separate mini-theories.

Another theory that squares well with Russell's (2009) program is the two-factor theory of LeDoux (2012b) in which mechanisms for behavior causation are decoupled from mechanisms for generating feelings. Behavior is caused by survival circuits. These are innate [S–R] links that are not specific to emotions, but rather to fundamental physiological needs such as defense, maintenance of energy, fluid balance, thermoregulation, and reproduction.[89] Survival circuits are triggered by USs (or their CS counterparts) and produce reflex behavior or fixed reaction patterns. For instance, the defense circuit in rodents is triggered by the

[89] This is the point where LeDoux (2012b) departs from Panksepp (2012), who did align his set of subcortical circuits with emotion types (even if these emotion types did not map neatly onto the vernacular ones). It is precisely because LeDoux's (2012b) survival circuits are not specific to emotions that a construction process is needed to provide the experience of specific emotions.

https://doi.org/10.1017/9781107588882.012 Published online by Cambridge University Press

smell of a predator. Feelings, on the other hand, correspond to what we normally take to be human emotions. A subset of these feelings are constructed by the conscious interpretation and labeling of the activity or manifestation of a survival circuit in the presence of a particular kind of challenge or opportunity. Survival circuits do not align well with feeling types, however. Anger could be based on signs of aggression, for instance, but aggression can occur in relation to different survival circuits, as is reflected in the existence of different varieties of aggression such as defensive aggression (defense circuit), conspecific aggression (reproduction circuit), and predatory aggression (maintenance of energy circuit). LeDoux and Daw (2018) recently extended the set of mechanisms of behavior causation to also include habits and goal-directed processes. Although this draws their theory somewhat closer to the goal-directed theory (Moors, 2017a), the explanatory weight that they confer on innate and habitual processes far exceeds that granted by the goal-directed theory.

FRUITFULNESS

Barrett (2006b, 2014, 2017a, 2017b; Barrett et al., 2015) emphasized that Factors 1 and 2 are both generated by general-purpose mechanisms. Neither core affect nor categorization are unique to emotions. We always occupy some point in the two-dimensional space of core affect, and categorization is required for nearly every mental process. I agree that categorization, understood in the generic sense, is not specific to emotion, but this does not seem to hold for the categorization process proposed in Barrett's (2006b) theory. This is a particular kind of categorization process, namely one in which a particular set of inputs is sorted under the category of an emotion (e.g., anger) or a situation-specific subset of this emotion (e.g., anger in a competition context). This categorization process is in fact specific to emotions, the more so because Barrett (2006b) takes the output of this categorization process to be a real emotion, and not just a mere self-ascription.

Returning to Russell (2003, 2009), the mechanisms responsible for all the components in alleged emotional episodes are truly general-purpose. Specifically, in the component of emotional meta-experience, the categorization process is considered to be nothing more than the perception or self-ascription of an emotion. The object of this perception does not need to correspond to any scientific reality. To drive this point home, Russell (2003, 2009) drew an analogy with the recognition of stellar constellations like the Big Dipper. The stars involved are real, but the constellation itself is not a scientific entity that explains the presence of the stars or generates any interesting predictions. Based on this analogy, Russell (2003, 2009) rejects the fruitfulness of the set of emotions and denies it any scientific status.

8.3.2 Divisio Definition

8.3.2.1 Criteria for Partitioning

Both Schachter (1964) and Barrett (2006b; Barrett et al., 2009) combine a dimensional with a discrete view. The mechanism involved in Factor 1 yields the dimensions of arousal and/or valence, which fits with a dimensional view. The mechanism involved in Factor 2 yields specific emotions, which fits with a discrete view.

In Russell's (2009, 2012) PCT, the variety within the folk set of emotions can be organized according to any dimension or category that describes the components in presumed emotional episodes. There is no a priori best way for doing this; it depends on the researcher's aims (see also James, 1890b). Because Russell (1979) has analyzed the meaning of emotion words using the dimensions of valence and arousal, he has been misunderstood as claiming that the set of emotions themselves is best organized according to these two dimensions.

8.3.2.2 Adequacy

APPARENT-SIMILARITY AND FRUITFULNESS

The construction process in Schachter's (1964) and Barrett's (2006b) theories can easily provide any of the vernacular emotion types from any working definition imaginable. How fruitful are emotion subsets that correspond to these vernacular emotion types? It may seem ironic that after denying a biological essence to vernacular emotions, Barrett (2006b) ends up postulating a process that returns to these vernacular emotions again (Fridlund, 2017). To avoid the risk that these emotion categories would invite any mental essences, however, Barrett (2017b) treats them instead as populations of context-dependent action-goal scripts that are held together by an emotion label. However, this presents us with a dilemma, neither horn of which seems very attractive. One horn of the dilemma is to conceive of the label as a mere frivolity that does not capture anything meaningful. In this case, all we are left with is a set of action-goal scripts for specific situations. Yet this raises the question: Why insist that these scripts are emotional? What would be lost if the scripts were simply called situation-specific action-goal scripts? If the brain is capable of selecting the best action to deal with the situation, such as fleeing, what does the brain gain from labeling the bodily responses that prepare for this action as fear?

The other horn of the dilemma considers the emotion categories individuated by linguistic labels as important organizing factors. If this option is taken, the question can be asked: How fruitful is this organizing factor? It may be recalled that we are not allowed, according to Barrett (2017b), to search for any commonality that is shared by the members of a particular emotion category (e.g., anger), not even in terms of action goals, as these too are supposed to be context-dependent (Hoemann et al., 2019; see above). If there is no basis for homogeneity, making generalizations from one member of a set to the next (extrapolation) seems prohibited, and the conclusion must be that emotion categories are not fruitful subsets. But perhaps the claim that the members of a category have nothing in common except their label is an overstatement. Their commonality could, for instance, be based on statistical similarity across many features, and therefore be difficult to capture in words (Barsalou, 1985; Hélie, Waldschmidt, & Ash, 2010; Voorspoels et al., 2011; see Lindquist, MacCormack, & Shablack, 2015; but see Lindquist et al., 2012, p. 125). This does not erase the fruitfulness problem, however. Subsets with non-verbalizable boundaries do not promise to be very fruitful. For one thing, they do not allow scientists to determine membership of instances to these subsets. For instance, how should a researcher approach the question of whether anger is beneficial for well-being if there is no way to figure out what qualifies as anger in the first place? Seeking a commonality like an *abstract* action goal for each emotion category, such as the goal to harm for anger and the goal to seek safety for fear, does not seem to be the path that PCTs would want to take, as this would involve positing an essence again in line with evolutionary theories.

Let us finally return to Russell's (2009, 2012) theory. Due to the lack of recommendations for organizing the folk set of emotions, there is nothing to evaluate in terms of adequacy. The theory is nevertheless able to make sense of vernacular emotions. As with other PCTs, this relies on categorization. People can self-ascribe any emotion for which they have a label in their language.

8.4 Validation

8.4.1 Empirical Research

Research conducted under the flag of PCTs can be split into (a) a deconstructive program that collects evidence and arguments against the existence of affect programs and (b) a constructive program designed to yield positive evidence for the hypotheses of PCTs.

The deconstructive program was already reviewed in the validation section of Chapter 4. Evidence against discrete emotions is crucial for PCTs because the first factor in these theories is supposed to yield diffuse bodily feelings only. If the first factor were to already produce emotion-specific quality, the second factor, which is supposed to provide this emotion-specific quality would become superfluous. The role of the

https://doi.org/10.1017/9781107588882.012 Published online by Cambridge University Press

second factor would then have to be reduced to providing verbal labels for entities that are already differentiated in themselves.

The constructive program should ideally try to show that the two factors in PCTs, bodily feelings (Factor 1) and construction in the form of attribution or categorization (Factor 2), are both necessary for the experience of specific emotions. This involves showing that both factors together produce a discrete emotion, but each factor alone does not. A common approach to test this is via misattribution studies. In these studies, bodily feelings and stimuli to which these feelings can be attributed are manipulated independent of each other. The bodily feelings are manipulated via the manipulation of bodily responses. This is done either via (a) an indirect method, by presenting stimuli that produce more vs. less bodily responses or (b) a direct method, by injecting chemical substances vs. a placebo solution. The content of the attribution process is manipulated via the presentation of different salient stimuli in the environment that are likely to be picked up as causes of the arousal. An example of a study that used the indirect method for manipulating bodily responses and hence bodily feelings – here called arousal – is a field study by Dutton and Aron (1974) in which male passers-by were approached by an attractive female experimenter (salient stimulus) as they crossed either a high wobbly bridge vs. a low safe bridge (more vs. less arousal-inducing stimulus). Whenever a man walked across one of the bridges, the experimenter walked up to him and asked him to make up a story about a photo while they were standing on the bridge. Afterwards, the experimenter gave her telephone number in case the participant wanted more information about the study outcomes. Results showed that participants on the high bridge made up more stories with a sexually oriented content and called the experimenter more often than participants on the low bridge. This suggests that arousal caused by one stimulus (the high bridge) can be misattributed to another, salient stimulus (the experimenter) (but see Allen et al., 1989).[90]

An example of a study that used the direct method for manipulating arousal is the famous experiment of Schachter and Singer (1962) in which participants were either injected with epinephrine (which induces arousal) or a placebo solution. Stimuli were a combination of types of information (correct, incorrect, or no information) and the presence of a pseudo-participant who was either euphoric or angry. Participants' emotions were measured directly via self-report (rated as euphoric/angry) and indirectly by observing their behavior during a waiting period

[90] Allen et al. (1989) examined alternative explanations in terms of negative reinforcement (i.e., participants on the high bridge are under threat so that the presence of the experimenter is more rewarding) and general response facilitation (arousal makes participants more proactive) and found evidence for the latter.

https://doi.org/10.1017/9781107588882.012 Published online by Cambridge University Press

(classified as euphoric/angry). The hypotheses were that the placebo group would not report or show any emotions because the first factor (arousal) was missing. The epinephrine groups were all exposed to the first factor, but the nature of the second factor, that is, the salient stimulus to which they could attribute their arousal, varied depending on the types of information they had received. It was hypothesized that the correctly informed group would not suffer from misattribution as they were given a plausible explanation for their increased arousal, and hence that they would not report or show euphoric/angry emotions. The groups that received incorrect or no information, on the other hand, would misattribute their arousal to the presence of the euphoric/angry pseudo-participants and would therefore report and show matching euphoric/angry emotions. The results of this study were inconclusive, however, and turned out to be difficult to replicate (e.g., Marshall & Zimbardo, 1979). The modest conclusion was that misattribution of inexplicable arousal leads to the intensification of emotions but not to a change in their quality (Reisenzein, 1983). Another complaint was that evidence for the phenomenon of misattribution does not show that attribution of arousal is the standard mechanism for emotion elicitation (Deonna & Scherer, 2010).

While the previous studies induced actual bodily responses to create bodily feelings, other studies have provided participants with bogus feedback about their bodily responses. In the pioneering study of Valins (1966), participants in the experimental group watched erotic pictures while overhearing what they believed was their own heartbeat but which was in fact a prerecorded tape. Half of the pictures were accompanied by an increase in heart rate and the other half by a steady heartbeat. Participants in the control group watched the same pictures accompanied by the same sounds but these were framed as extraneous noise. As predicted, participants in the experimental group rated the former type of pictures as more attractive than the latter type whereas participants in the control group did not. Valins (1966) interpreted this result as support for the idea that mere information about a change in arousal combined with information about a potential cause of this change (the erotic stimulus presented alongside this change) is sufficient to produce emotional experience. This interpretation is known as the two-cognition theory. Subsequent research was able to replicate the effect with both positive and negative stimuli, but also pointed at the limits of this effect (see review by Parkinson, 1985). In addition to the risk of demand effects, differences between the experimental and control groups regarding the attention paid to the sounds, and the fact that participants' preferences may not count as genuine emotional experiences, it is likely that all stimuli already elicited a substantial degree of arousal to begin with. This imposes the more modest conclusion that the impact of already arousing stimuli on preferences can be modified (rather than installed)

https://doi.org/10.1017/9781107588882.012 Published online by Cambridge University Press

by bogus feedback (Parkinson, 1985; but see M. A. Gray et al., 2007). It is further possible that the false feedback of bodily responses was more than a cold piece of information but induced actual bodily responses via a mimicking mechanism (D. Goldstein et al., 1972; but see Parkinson, 1985) or that it led to the mental simulation of these bodily responses, in line with the suggestion of Barrett et al. (2015; see above).

Recent research revisited the questions previously addressed by Schachter (1964). One line of research focused on the role of Factor 1: the influence of bodily feelings – understood here as core affect – on the intensity of emotions. MacCormack and Lindquist (2017) reviewed both neuroscientific and behavioral studies. Contrary to what MacCormack and Lindquist (2017) claimed, however, evidence for the influence of core affect on the intensity of emotions is not uniquely consistent with PCTs. Such evidence would also be consistent with appraisal theories. These theories hold that all components in an emotional episode directly project onto the feeling component and indirectly also via appraisal.

Another line of research focused on the role of Factor 2: the influence of categorization (rather than attribution) on the emotion-specific quality of emotions (Barrett, Lindquist, & Gendron, 2007; Lindquist, 2017; Lindquist, MacCormack, & Shablack, 2015; Lindquist, Satpute, & Gendron, 2015). Here again, some studies rely on neuroscientific methods and provide mostly correlational evidence. Other studies employ behavioral methods (see also Box 2.2). The aim of the behavioral studies is to manipulate the accessibility of emotion categories and to measure the influence of this manipulation on (a) participants' own emotional *experience* (Kassam & Mendes, 2013; Lindquist & Barrett, 2008) and (b) their *perception* of other people's emotions (Lindquist et al., 2006). I provide an illustration of each.

Lindquist and Barrett (2008) conducted an *experience* study. They manipulated the accessibility of anger vs. fear by showing participants a picture of an interaction between an angry and a fearful man and by asking them to tell a story about the angry vs. the fearful man. Core affect was manipulated by asking participants to imagine a high-arousal negative event in the presence of high-arousing negative music vs. a neutral event in the presence of neutral music. Participants' fear experience was measured indirectly via self-reported risk aversion. Results showed more risk aversion in the group primed with fear and exposed to high-arousal negative stimulation than in the other groups. This was interpreted as evidence that emotion categories shape diffuse core affect into specific emotions. It is noteworthy that Lindquist and Barrett (2008) measured emotional experience via self-reported risk aversion. If, as they claim, there is little concordance among the components of emotions, it is doubtful that one component (feelings) can be measured indirectly via another component (behavior).

Lindquist et al. (2006) conducted a *perception* study. They manipulated the accessibility of emotion categories by manipulating the accessibility of emotion words. In one group, they increased the accessibility via a priming procedure (by presenting the emotion word three times). In another group, they decreased the accessibility via a satiation procedure (by presenting the emotion word 30 times). Participants had to recognize stereotyped facial expressions that either matched or mismatched with the primed/satiated emotion category. Results showed that only in the match condition were participants slower to judge the facial expression when the emotion category was satiated. This provides support for the role of categorization in the perception of facial expressions.

Evidence for the influence of emotion categorization or labeling on emotional *experience* and emotion *perception* is uncontroversial. Many of the other theories discussed so far would not deny that categorizing emotions after the fact can influence one's subsequent experience. If it is accepted that feelings are determined by representations that become conscious during an emotional episode, then representations of specific emotions or emotion labels should also be able to color these feelings. PCTs make the bolder claim that without categorization, an episode has no emotion-specific quality. As argued earlier, however, demonstrating the necessity of something is not a realistic purpose of investigation. It can only be provisionally maintained until proven otherwise. Such a piece of counterevidence was provided by Sauter et al. (2011). They showed that Yucatec Mayan participants, who have a single word for anger and disgust, performed just as well as German participants on a task in which they had to perceptually categorize morphed faces on a continuum from anger to disgust. This suggests that emotion concepts are not necessary to at least perceive different emotions.

8.4.2 *Internal Consistency*

In addition to empirical validation, PCTs have also been evaluated in terms of internal consistency. Deonna and Scherer (2010, p. 48) raised a first objection to Lindquist and Barrett's (2008, p. 902; Barrett, 2006b; but see Barrett, 2022) assumption that people constantly categorize their core affect, and not just when the source of the bodily changes is ambiguous (as per Schachter, 1964, and Russell, 2003). If people indeed constantly categorize core affect to determine the emotions they are in, this would put a heavy burden on the system in terms of computational resources or metabolic costs. In addition, it would also generate noise that detracts from the person's first priority, which is to deal with stimuli in the service of goal satisfaction.

A second objection raised by Deonna and Scherer (2010; Deonna & Teroni, 2012) was that Barrett (2006b) modeled her mechanism for

https://doi.org/10.1017/9781107588882.012 Published online by Cambridge University Press

emotion causation on Schachter's (1964) atypical mechanism for emotion causation. Even Schachter (1964) acknowledged that in typical cases, stimulus evaluation is the source of both the bodily changes and the label to categorize these bodily changes. This led these critics to propose that the PCT of Barrett (2006b) is not a theory of emotion causation, but rather a theory of "how we come to know our emotions." So rather than dismissing the theory, they reframed it as having an alternative explanandum. Rather than being a necessary factor in emotion causation, emotion categorization could be a way to discover or come to know one's emotion, when that emotion is already caused by other processes. Reframing the theory in this way, they argued, allows us to explain the phenomena that motivated the invention of PCTs in the first place. These are cases in which we are uncertain about our emotions because the stimuli that caused them are not consciously accessible or require conscious search before they are accessed.

In the skeptical spirit of Russell's (2003) PCT and my own goal-directed theory (Moors, 2017a), I am inclined to argue that categorization may not so much be *a way to come to know one's emotion*, but rather *a way to talk oneself into having an emotion*. Barrett (2006b) assumed that representing oneself as having an emotion is sufficient for having the emotion. To show what could be wrong with this idea, consider the analogy with the self-ascription of having supernatural powers (Moors, 2017b). True enough, such a self-ascription may influence behavior and communication and in this sense acquire social functions that the person's state would otherwise not have. Yet it is not because a person *thinks she has* supernatural powers, that she actually *has* them or that supernatural powers even exist. This will become painfully clear when she decides to act on her presumed supernatural powers by trying to fly out of the window.

To summarize, Schachter (1964) articulated a two-factor theory of emotion causation in which diffuse bodily responses (Factor 1) have to be combined with a construction process that interprets these bodily feelings in terms of discrete emotions (Factor 2). In an effort to keep the two-factor theory of emotion causation palatable to current taste, Barrett (2006b, 2012, 2017b) reshaped the processes involved, in sync with recent developments in cognitive (neuro)science (i.e., embodied and situated cognition; Barsalou, 2008; constraint satisfaction; Kunda & Thagard, 1996; predictive processing; A. Clark, 2013). Russell (2003, 2009, 2012) took a different path. He rejected the idea that Schachter's (1964) two-factor process is a mechanism for emotion causation and gave it the more modest role of a mechanism for emotion self-ascription.

CHAPTER 9

Social Theories

In all of the emotion theories discussed so far, the primary focus has fallen on the mechanisms that take place inside the individual. All these "personal" theories nevertheless assume that emotions often, if not most of the time, occur in a social context. Social theories go beyond assuming this trivial fact, however, by working out its detailed implications. Many social theories start out from, and ride on, the mechanistic explanations provided by personal theories. Application of personal mechanisms in a social context has led to the consideration of phenomena on larger scales, both in terms of the number of individuals involved (i.e., numerosity scale) and time (i.e., temporal scale). This, in turn, has led to the proposal of novel mechanisms of emotion causation that are specifically social in nature.

A few remarks regarding scale expansion are in order. The numerosity scale expands the scope from (a) the individual to (b) dyadic interactions, (c) group interactions, and (d) cultural contexts (Boiger & Mesquita, 2012; Keltner & Haidt, 1999). On closer scrutiny, however, this scale combines two partly independent factors. A first factor is social presence, or better, *social interaction*, that is, whether a person is alone or interacts with one (in dyads) or several others (in groups). The last step on the scale (cultural context) does not follow this logic, however. Although a person is always embedded in at least one cultural context, she cannot literally interact with all the people in her culture at once. Thus, a second factor is the use of *socio-cultural information*, that is, representations including others, such as representations of descriptive norms (i.e., what others do) or injunctive norms (i.e., what others think one should do; Lapinski & Rimal, 2005). This information is not only available in dyads and groups, but may also be consulted when an individual is alone.

The temporal scale, in turn, expands the scope from a single interaction unit to a series of interaction units over time. Such an expansion is natural in dyads and groups given that most social interactions are composed of a series of reciprocal behaviors from sender to receiver(s) and back.

My discussion of social theories will be organized as follows. For each of the personal theories discussed in the previous chapter, I start by

https://doi.org/10.1017/9781107588882.013 Published online by Cambridge University Press

reiterating the role they already grant to *socio-cultural information* (i.e., second factor of the numerosity scale), that is, the degree of social infusion that they already permit in their mechanisms. Next, I present ways in which social versions of these theories have detailed and elaborated on the influence of this socio-cultural information. After that, I mention novel mechanisms for emotion causation that social theories have added to the repertoire when they extended their scope to *social interactions* (i.e., first factor of the numerosity scale; temporal scale). Unlike the previous chapters, the present chapter is not organized according to the stages in the demarcation-explanation cycle, the reason being that the mechanisms discussed here build on those that were already discussed in earlier chapters. Before turning to the theories, I present a brief preview of the types of mechanisms that are at stake in *social interactions* and I list a few important axes on which social theories can be placed and compared.

Mechanisms in Social Interactions

One set of social mechanisms already becomes available when we zero in on the smallest unit of the social interaction: the behavior of one sender in the presence of one receiver. The mechanisms in this micro-interaction can be organized into ones operating in the sender and ones operating in the receiver. From the angle of the sender, it is worth distinguishing between different types of behavior with the help of a taxonomy that rests on various dichotomies such as (a) social vs. non-social, (b) coarse vs. subtle, (c) direct vs. indirect, (d) communicative vs. non-communicative, (e) verbal vs. nonverbal, (f) intentional vs. unintentional, and (g) stimulus-driven vs. goal-directed (see Figure 9.1).

A first type of behavior is non-social behavior. This may include both coarse behavior, which extends beyond the body (e.g., approaching food), and subtle behavior, which is confined to the body such as facial and vocal behavior (e.g., scrunching the nose, widening the eyes, sighing). Non-social behavior can be unintentional, but will often be intentional. This means nothing more than that it is caused by the intention or action tendency to engage in the behavior. It by no means implies that the action tendency is also conscious.

A non-social action tendency may be stimulus-driven, in which case it stands for the [R] in an innate or learned [S–R] link. It may also be goal-directed, which means that it is selected as the best strategy to reach a certain goal. To illustrate, approaching food may be caused by an instinct or habit activated by the sight of food or it may be selected as the best way to get access to this food. Likewise, certain facial behaviors are caused by innate protective reflexes such as the eyeblink reflex. Other facial behavior may be caused by a goal-directed process. Scrunching the

https://doi.org/10.1017/9781107588882.013 Published online by Cambridge University Press

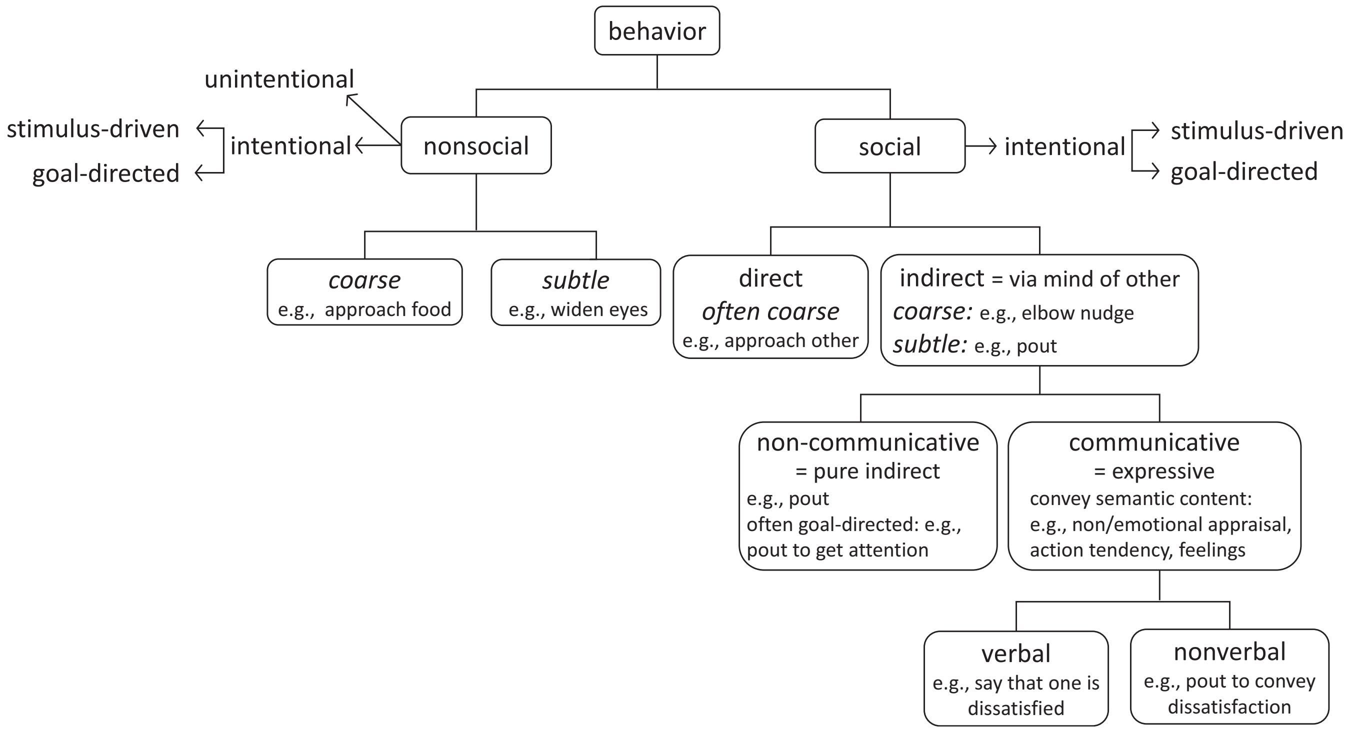


Figure 9.1 Taxonomy of behavior of a sender in the presence of a receiver

nose may be selected because it is the best way to reduce air intake from a bad smell, widening the eyes may be selected because it is the best way to increase information intake, and sighing may be selected because it is the best way to increase oxygen intake. Note that the goals that are served by non-social behavior (e.g., hunting) may themselves be non-social (e.g., get food) or social (e.g., feed the family).

A second type of behavior is social behavior. This is behavior that a sender directs towards a receiver. Understood in this sense, social behavior is intentional in the sense that it is caused by the tendency to influence the receiver in a particular way. Again, there is no need for this tendency to be conscious. Social behavior can be split into direct social behavior and indirect social behavior. The direct variant targets the receiver directly, which will often require coarse behavior. For instance, a sender can attack, avoid, push, or pull the receiver. The indirect variant targets the receiver indirectly, by first exerting an influence on the "mind" of the receiver. This can happen via coarse behavior (e.g., nudging somebody with the elbow) or subtle behavior (e.g., pouting).

Within the subclass of indirect social behavior, a distinction can further be drawn between communicative and non-communicative behavior. In the communicative or expressive variant, the sender has a representation with a certain semantic content and has the tendency to convey this content to the receiver. Anything that can be represented by the sender can figure in the content of a message. Examples are evaluations, action tendencies, and feelings, which can be "emotional" or "non-emotional" (for theories that adhere to the distinction). Further, communicative behavior can take a verbal or nonverbal (e.g., facial, vocal, gestural) form. A person can simply say that she is dissatisfied, or she can pout to convey the same message. In the non-communicative or pure variant, on the other hand, the sender simply wants to influence the mind of the receiver in a particular way, but she does not have a representation with semantic content that she wants to convey to the receiver. For instance, a toddler may pout to a caregiver (e.g., with the further aim of getting attention) without trying to send a semantic message (see Fridlund, 1994, below).

All three types of social behavior – direct, communicative indirect, and non-communicative or pure indirect – are intentional in the sense that they are caused by the tendency to influence the receiver in a particular way (e.g., to approach her, to send her a message, or to pout at her). This tendency can itself be stimulus-driven if it is part of an innate or learned [S–R] link. The action tendency can also be goal-directed, in which case it is selected as the best strategy to reach a goal. For instance, approaching another person (i.e., direct social behavior) may be done to obtain safety; communicating one's sadness (i.e., communicative indirect social behavior) may be done to get consolation; and a toddler's pouting (i.e., pure indirect social behavior) may be done to get the caregiver's attention. The goals that are served by social behavior may themselves be social or non-social. Pouting may be done to get affection or to get a cookie.

Social behavior can be more or less successful in influencing the receiver, and the communicative variant of social behavior can be more or less successful in transmitting a message to the receiver. But all types of behavior, including non-communicative and even non-social behavior, may be used as a source of information by the receiver. A person who scrunches her nose to reduce air intake may unintentionally alert a bystander that something is fishy. Likewise, a sender who wants to attack may prepare by clenching her teeth and holding the receiver in a fixed stare. Even if this facial behavior was not caused by the intention to communicate anything, the receiver may still infer that the sender is offended (i.e., appraisal), that she wants to attack (i.e., action tendency), or that she feels angry (i.e., feeling).

In addition to mechanisms that operate in the sender and the inferences that receivers are able to draw from the sender's behavior, social theories have also proposed mechanisms that produce emotions in the receiver in response to the behavior of the sender. Examples of mechanisms in the receiver are emotional contagion and social appraisal (see also Chapter 7). These will be discussed in more detail later.

If we gradually build up from the smallest interaction unit to multiple interaction units across time (i.e., on the temporal scale), another novel set of social mechanisms becomes available that can be captured under the heading of distributed mechanisms. The idea is that processes do not happen in a single person but are spread out or "negotiated" among the interaction partners. Increasing the number of interaction partners from dyads to groups (i.e., on the numerosity scale) further increases the complexity and calls for macro-level social theories. Both types of complexity can benefit from a DS approach in which higher-order emergent patterns result from lower-level reciprocal interactions (Parkinson, 2012, pp. 295–296; Mesquita & Parkinson, 2022). Note that in previous applications of the DS approach, self-organization took place inside the individual whereas here it takes place among individuals.

Axes

Social theories differ in whether they are loyal to the mechanisms of a single personal theory or whether they integrate the mechanisms of several personal theories (**Axis A**: pure vs. mixed). Social theories also differ with regard to the levels they address on the scales of numerosity (dyads vs. groups vs. cultures) and time (single vs. multiple interaction units) (**Axis B**). Theories that focus on socio-cultural information qualify as *socio-cultural* theories (Frijda & Mesquita, 1994; Harré, 1986; Mesquita et al., 2017). Theories that include mechanisms for single interactions have been dubbed *communicative* theories (Parkinson, 1995). Following the taxonomy laid out above, however, they would be more appropriately characterized as *social behavior* theories. Finally, theories that expand to multiple interaction units across time and that propose distributed mechanisms go by the name of *transactional* theories (Boiger & Mesquita, 2012; Parkinson, 1996). It may be noted that the theories of individual scholars often combine several of these qualities (e.g., Boiger & Mesquita, 2012; Parkinson, 1996, 2012).

9.1 Evolutionary Theories

9.1.1 Personal Version

Evolutionary theories take it as a given that socio-cultural information rides on the back of secondary (learning) processes, but does not touch primary (innate and hence universal) processes. That being said, according to Darwin's (1872) teleological principle, primary processes originated in the learning of our evolutionary ancestors, hence this learning must have been sensitive to socio-cultural information at some point. In line with this, Ekman (1992a) stressed that the fundamental life tasks (challenges and opportunities) that ground the affect programs for basic emotions were first and foremost social in nature. He declared that affect programs first developed for interpersonal encounters (e.g., facing competition, escaping enemies) and only later became triggered by non-social stimuli (e.g., escaping thunder, loss of physical support, music). Once these affect programs became cemented in the brain, however, plasticity got reduced and novel socio-cultural information (e.g., a change in cultural norms) was no longer able to touch the innate process. To invoke the avocado pear metaphor again (see Chapter 4), current socio-cultural information can touch the soft flesh of emotion but not the hard core (Faucher & Tappolet, 2008). The soft flesh is situated on the input side, where classical conditioning can expand the set of stimuli that can trigger affect programs, as well as on the output side, where emotion regulation and planning can alter and refine the manifestation of emotions in line with cultural display rules (Matsumoto, Yoo, et al., 2008).

9.1.2 Social Version

A first social extension of the evolutionary theory works out the implications of expanding the numerosity level from the individual to dyads, groups, and cultures. Keltner and Haidt (1999; Keltner et al., 2006) asked what the function of emotions could be on all these levels. A few examples. At the individual level, emotions serve to alert the individual about social stimuli and prepare a response. For instance, anger provides information about the unfairness of a transaction and suggests using aggressive behavior as a solution. At the dyad level, emotions of a sender help a receiver to coordinate with the sender by producing reciprocal or complementary emotions. For instance, anger from a sender may lead to anger or fear in a receiver. In turn, the emotions of both interaction partners influence the nature of the interaction. Anger increases the probability of conflict, happiness the probability of cooperation, and sadness the probability of disengagement (Oatley & Johnson-Laird, 2011). At the group level, shared emotions establish group identities whereas different emotions help establish group roles. Shared indignation about a societal problem can be the instigator of a new civil movement, for instance, and anger is for leaders whereas admiration is for followers. At the cultural level, finally, emotions build cultural identities and are the lubricants for transmitting cultural values and norms. In their current state, however, the "functions" identified on all these levels do not seem to pass the stage of being just-so stories (see Gould, 1978; Holcomb, 1996). Emotions may have certain effects (or so it may seem), but this does not mean that these effects reflect the functions of emotions. Some effects may be accidental, not having been set up by evolution to benefit the emoter or the group.

A second social extension of evolutionary theory (e.g., Keltner, Sauter, et al., 2019; Shariff & Tracy, 2011a) focuses on the evolutionary origin of communicative behavior, in particular facial expressions. Based on their reading of Darwin (1872), evolutionary theorists proposed that facial behavior first evolved as (a) adaptations and later as (b) exaptations. They argued that in a first stage of evolution, facial behaviors had the physiological function of preparing the organism for emotion-specific behavior. An example is baring the teeth as a preparation to bite an opponent in anger, widening the eyes to increase visual input in fear, and scrunching the nose to reduce air intake in disgust. Although these facial behaviors did not initially evolve to allow the sender to communicate information, they nevertheless allowed receivers to infer information about the sender's emotion. Such facial behaviors are called "cues" (Shariff & Tracy, 2011a). In a later stage of evolution then, some of these cues were co-opted to serve the novel function of communication. Facial behaviors that evolved for the purpose of communication are called "signals." During the transition from cues to signals, moreover, facial behaviors underwent a process of ritualization in which they became more exaggerated and stereotyped, so that they would be recognized more easily. In fact, evolutionary theorists took the ritualized character of facial behaviors as support for the existence of signals.

It may be recalled, however, that Darwin (1872) emphasized that facial behaviors were remnants of ancestral instrumental behavior rather than adaptations (Crivelli & Fridlund, 2019; Fridlund, 1994; Fridlund & Russell, 2022; see Chapter 3). Some of these vestigial behaviors indeed got co-opted to fulfill a new function, namely, to express a mental state. But here too, there is a risk of misreading Darwin (1872) when he wrote that facial behaviors "express" mental states. For it seems that what he had in mind was the *pure* form of *indirect* social behavior rather than the *communicative* form (see Figure 9.1). Thus he wrote: "Even such words as that 'certain movements serve as a means of expression' are apt to mislead as they imply that this was their primary purpose or object. This, however, seems rarely or never have been the case [. . .]. An infant may scream either intentionally or instinctively to show that it wants food; but it has no wish or intention to draw its features into the peculiar form which so plainly indicates misery" (Darwin, 1872, p. 357).

Regardless of what Darwin (1872) had in mind, evolutionary theorists abide by their evolutionary story that facial behaviors were originally adaptations for the preparation of emotional behavior that were later exapted for the communication or expression of emotions. Moreover, they consider it an open question whether or not the original adaptive function of these facial behaviors still applies today (Lee & Anderson, 2017; see Sauter & Russell, 2022).

Evolutionary theorists acknowledge that facial behaviors can be influenced by, or communicate, other things than emotions, such as attempts to regulate emotions via display rules, paralanguage (speech-related movements; e.g., raising eyebrow to accentuate a word), and protolanguage (symbolic faces; e.g., wink). Facial behaviors may also communicate emotions that a person is not having by faking emotional expressions (Ekman, 1997; Kappas, 2003; Scarantino, 2017b, 2019; Scherer et al., 2013). Evolutionary theorists remain convinced, however, that true and false emotional expressions can be distinguished based on their morphology. A true smile, called a Duchenne smile, would be characterized by the contraction of both the zygomaticus major (mouth corners up) and the orbicularis oculi (eye wrinkles) whereas a fake smile only by the contraction of the zygomaticus major (e.g., Ekman et al., 1980; Ekman & Friesen 1982). Recent research, however, shows that Duchenne smiles are easy to fake and occur when they are theoretically unexpected such as in the presence of intense but unpleasant stimuli (see reviews by Crivelli & Fridlund, 2019; Girard et al., 2021). The latter finding is in line with Fridlund's (1994) suggestion that eye wrinkles are not a marker of positive stimulus valence but rather of response intensity.

In addition to morphological differences, evolutionary theorists hold the assumption that true emotional expressions are innate and therefore automatic whereas emotion regulation including social pretense is goal-directed and therefore non-automatic, unless it has become habitual, in which case it is again automatic. Thus, as long as it is not habitual, emotion regulation can only occur under ample operating conditions, such as when there is abundant time and attentional capacity and/or sufficient motivation. Even if motivation to fake is high, lapses in these regulatory attempts can produce occasional leakage of the true expression in the form of micro-expressions. Ekman (1973, pp. 215–218) reported an unpublished study with North American and Japanese students, who watched a stress-inducing film when they were either alone or while they were being interviewed by an experimenter. Both groups showed similar negative facial expressions when they were alone, but Japanese students showed more polite smiles during the interview, presumably to mask their negative facial expressions. The conclusion was that facial behaviors may be partially determined by social pretense, but not entirely, otherwise they should have been absent when students were alone (but see Fridlund, 1994).[91]

Let me already note that the goal-directed theory rejects the assumption that spontaneous facial behaviors are emotional whereas deliberate ones are fabricated. Both spontaneous and deliberate facial behaviors are goal-directed and neither of them is emotional in the scientific sense of the term. The tension between alleged "emotional" and "non-emotional" facial behaviors is nothing more than the competition between two goal-directed processes that are each at the service of a different goal (see also Fridlund, 1994, 2017; see below).

9.2 Network Theories

9.2.1 Personal Version

Network theories agree with evolutionary theories that socio-cultural information travels with learning (i.e., secondary processes), but they allow a more drastic role for learning in the sense that what is learned is not merely tagged on to the emotion but rather a proper part of it. Socio-cultural information can lead to a far-reaching reshaping of the

[91] Fridlund (1994; Crivelli & Fridlund, 2019) pointed out that the results of this study were oversold because Ekman (1973) consistently omitted a third condition in which participants watched the film in the mere presence of the experimenter and in which no difference with the alone condition was observed.

emotional repertoire, and this idea is more pronounced in the non-biological than in the biological version of network theories.

9.2.2 Social Version

Social extensions of network theories have proposed "emotional contagion" as a novel social mechanism for emotion causation that can occur during social interaction. Emotional contagion is an umbrella term for various mechanisms that explain interpersonal emotion transfer without the mediation of inferential processes (Elfenbein, 2014; Parkinson & Manstead, 2015). One such mechanism is *overt* emotional contagion (Hatfield et al., 1994). This mechanism is composed of a first step in which a receiver overtly mimics the emotional behavior (facial expression, posture) of a sender and a second step in which proprioceptive feedback from this behavior leads to the same emotion in the receiver. It is especially this second step that fits with network theories, more in particular, with their assumption that emotion networks can be activated via the response side (see Chapter 5). An alternative mechanism is *covert* emotional contagion. This refers to the direct activation of the representation of the emotional behavior or the mental simulation of this behavior without the need for overt mimicry (Barsalou, 1999; Neumann & Strack, 2000; Niedenthal et al., 2010; see Parkinson & Manstead, 2015).

9.3 Stimulus Evaluation Theories (SETs)

9.3.1 Personal Version

Within the family of SETs, biological appraisal theories share the view of evolutionary theories that the soft flesh of the avocado pear can be infused by socio-cultural information but not the hard core, that is, the fixed link between appraisal patterns and action tendencies or emotions. This link is supposed to be universal (see Mesquita & Ellsworth, 2001). On the input side, members of different cultures may appraise stimuli in different ways because they differ, say, in the goals they value, the control they have, and the attributions they make (Roseman et al., 1995; Scherer, 1997b). If individualist cultures attach a higher value to the goal for autonomy and collectivist cultures a higher value to the goal for connectedness, members of the former culture should experience stronger emotions when their freedom is restricted and those of the latter culture when they detect signs of social disapproval (Kitayama et al., 2006). If certain cultures attribute goal-incongruent stimuli such as illness to their own actions and other cultures more to external forces, guilt should be more common in the former culture and sadness more in the latter (Mesquita &

https://doi.org/10.1017/9781107588882.013 Published online by Cambridge University Press

Ellsworth, 2001). On the output side, culture exerts its influence via emotion regulation, by approving some emotions and disapproving others, but again, not by changing the shape of the emotion itself (Scherer & Brosch, 2009). Non-biological appraisal theories, on the other hand, have a harder time maintaining the universality hypothesis because different thresholds may exist for various components (see Chapter 6). This puts the avocado-pear plasticity model under pressure.

9.3.2 Social Version

Social extensions of SETs, in particular appraisal theories, have taken at least four different forms. A first extension focuses on goals that are more stable, abstract, and shared across the members of a society or culture, called socio-cultural values or norms.[92] The goals for autonomy vs. connectedness mentioned above already qualify as examples. Other examples are honor in honor cultures and happiness in American society (Mesquita & Ellsworth, 2001). More than values, norms have a compulsory or "ought" character. Moreover, the norms at issue here must be norms that the person has internalized, in other words, norms that have become the content of a goal and are highly valued. Norms that one knows but does not endorse (even implicitly) should remain emotionally inert.

A second extension of appraisal theory (Frijda & Mesquita, 1994; Mesquita & Ellsworth, 2001) explores the options that (a) not all appraisal variables have the same weight on the resulting emotions in all cultures and (b) not all cultures share the same set of appraisal variables. The first, weaker option, is that all appraisal variables are universal but that some are more relevant in some cultures than others. For instance, agency appraisals may be more focal in cultures that put a strong focus on autonomy than on connectedness (Matsumoto et al., 1988; see Frijda & Mesquita, 1994). It may be difficult, however, to disambiguate between the weight of appraisal variables and the value of certain goals (see Manstead & Fischer, 2001; Mesquita & Ellsworth, 2001). If members of collectivist cultures do not get angry when someone blocks their goal, this may be because they don't care so much about who is to blame (i.e., the agency appraisal has less weight) or because they don't care so much about their goals being blocked (i.e., the goal for autonomy has less value).

[92] Note that "value" is used here in a different sense than in the context of the goal-directed theory. There, "value" is used as a predicate of an outcome or goal. This is also different from "value" as a position on a variable.

https://doi.org/10.1017/9781107588882.013 Published online by Cambridge University Press

The second, stronger option, is that not all appraisal variables are universal and some cultures have unique appraisal variables. For instance, Markus and Kitayama (1991) suggested adding an appraisal variable of interpersonal engagement in Japanese culture because it allowed them to differentiate between self-focused (e.g., anger, pride) and other-focused emotions (e.g., feeling connected, shame, guilt). The existence of appraisal variables that are unique to a culture could also create unique emotions in that culture (e.g., feeling connected). The idea of culture-specific appraisal variables also aligns with the social extension of the judgmental theory (Solomon, 1984), which holds that the formal objects of emotions reflect cultural challenges (e.g., social threats) rather than natural challenges (e.g., natural threats).[93] In all of this, it is important to note that both weak and strong sources of cultural variation in appraisals, although still based on appraisal theory's terminology, mark a clear departure from the theory's long-standing ideal of a universal relationship between appraisals and action tendencies or emotions.

A third extension of appraisal theory puts forward "social appraisal" as a novel mechanism for emotion causation that can occur during social interactions (e.g., Manstead & Fischer, 2001). Social appraisal is an umbrella term for various mechanisms that allow a receiver to infer information from a sender's emotional behavior (e.g., facial behavior). According to one such mechanism, a receiver recognizes the emotion of a sender and reverse-engineers the appraisals that the sender must have made (Hareli & Hess, 2010). These appraisals help the receiver to make her own appraisals about a stimulus, especially if the stimulus is ambiguous. An alternative mechanism is that a receiver directly infers the appraisals of the sender without having to first identify the emotion that the sender had (Parkinson, 2011; Parkinson & Manstead, 2015).

Parkinson and Simons (2012) also pointed out that in addition to object-centered social appraisal, in which the appraisal of a sender helps the receiver to figure out how to appraise an object, there can also be sender-centered social appraisal, in which the appraisal of a sender helps the receiver to figure out how to appraise the sender. Both object-centered and sender-centered forms of social appraisal may trigger an emotion in the receiver. For instance, if other surfers do not seem threatened by the high waves in the ocean, Sam may feel less scared, but he may also

[93] Griffiths (1997) sorted the social version of SETs under the broader class of social constructionist theories. He called this version the social concept version, which he distinguished from the social role version (which I discuss in the section below). The social concept version focuses more on cultural influences on the input side (evaluation process) whereas the social role version focuses more on cultural influences on the output side (action tendencies and behavior).

conclude that the others are professionals and feel embarrassed to surf in their vicinity. If colleagues seem offended by Sunny's off-color joke, she may start feeling guilty, but she may also conclude that they are softies and feel contempt.

Interpersonal emotion transfer based on social appraisal is contrasted with emotional contagion (discussed above) on the grounds that the former is based on an inferential process whereas the latter is not. Van Kleef (2009) proposed two factors that should determine which of both mechanisms – social appraisal or emotional contagion – is the more dominant mechanism for interpersonal emotion transfer. In addition to a set of contextual moderators, the most important moderator put forward is the presence of ample vs. poor operating conditions, such as attentional resources and the motivation to engage in inferential processing. If operating conditions are ample, social appraisal is more likely; if they are poor, emotional contagion should take over. This model relies heavily on classic dual-system assumptions, however, and seems to disregard the possibility that social appraisal can proceed in an automatic way (see also Parkinson & Manstead, 2015). The existence of implicit social appraisal processes has been confirmed in research showing gaze-cueing effects (e.g., Bayliss et al., 2007; Mumenthaler & Sander, 2012, 2015). Parkinson (2019a, 2019b) raised the possibility, however, that gaze-cueing may do little more than orient or nudge the onlooker' attention, without conveying full-blown appraisal information.

Parkinson (2017a) introduced the notion of relation alignment as a container concept for all kinds of processes by which a receiver can get influenced by a sender and that contributes to the receiver's emotion or emotion orientation. In addition to processes of indirect relation alignment such as social appraisal and emotional contagion, he also points to processes of direct relation alignment. The latter include (a) the direct cueing of the receiver's attention by the sender's behavior (e.g., the sender's gaze orientation) and (b) the receiver's direct adjustment to relational dynamics, already discussed in Chapter 7 (e.g., an infant struggling from a tight embrace). As argued there, direct adjustment may still involve rudimentary personal appraisals of goal relevance and goal congruence.

A fourth extension of appraisal theory proposes a distributed social appraisal process, which is spread out over several interaction units (Parkinson, 1996, 2001b, 2019b, pp. 132–133). If appraisal is understood as a sequence of molecular appraisals (see Scherer et al., 2018), it is possible that the different appraisals are distributed across interaction units, and that reciprocal calibration among the interaction partners leads to a shared appraisal pattern.

Social appraisal, emotional contagion, and direct relation alignment can also be expanded beyond dyads to groups. The impact of these processes on interpersonal emotion transfer may be further facilitated by joint activities (especially when supported by rhythmic music; Parkinson, 2020). Together with social sharing (i.e., talking about emotions), these mechanisms are considered important ingredients of collective emotions (i.e., when different group members experience the same emotion). Several of these processes may also contribute to differentiated emotions (i.e., when different group members experience different emotions depending on their momentary or stable social roles) (Garcia & Rimé, 2019; Keltner & Haidt, 1999; Niedenthal & Brauer, 2012; Parkinson, 2019a, 2020; Rimé, 2009; Salmela & Nagatsu, 2016). DS principles may be invoked to explain why collective emotions sometimes spiral out of control and stabilize again over time (Butler, 2011).

9.4 Psychological Constructionist Theories (PCTs)

9.4.1 Personal Version

In PCTs, the capacity to experience core affect is supposed to be innate while the construction process is based on learned knowledge about how to link sensory input to specific emotions. This process is constrained and shaped by socio-cultural information (e.g., Barrett, 2006b; Russell, 2003).

9.4.2 Social Version

The natural social counterparts of PCTs are social constructionist theories (SCTs). While PCTs focus on the mechanisms inside the individual, SCTs work out the role of culture in the construction process. Averill (2012) declared affinity with Russell's (2003) categorization process understood as the recognition of a pattern of components that matches the prototype of an emotion. The source of this prototype, he insisted, are the social norms that are prevalent in a particular culture. If emotions are like the stellar constellations we single out in the sky, then different cultures may single out different constellations. Mesquita's (in Mesquita & Parkinson, 2022) version of SCT tunes in more with Barrett's (2017b) version of PCT. It may be noted, however, that not all theories that fall under the broad label of SCTs are loyal to the mechanisms proposed in existing PCTs. Several SCTs adopt a more liberal understanding of "emotions as constructions" as meaning that emotions are made up of smaller parts that are not emotional in themselves. Such an understanding is not only compatible with certain PCTs (Russell, 2003, 2009), but also with elemental appraisal theories (e.g., Scherer, 2009b). Parkinson (1996, 2007, 2012, 2019b), for instance, does not seem squarely opposed to a causal role for molecular appraisals in emotion causation, nor does he seem to assume

https://doi.org/10.1017/9781107588882.013 Published online by Cambridge University Press

that without categorization or labeling there is nothing that already qualifies as emotional according to common standards (Parkinson, 2012, p. 296). Likewise, at an earlier time, Mesquita and Boiger (2014) declared being agnostic about the mechanisms in the individual, suggesting that their theory could be compatible with PCTs and elemental appraisal theories alike.

Like PCTs, SCTs are hostile to the idea of evolutionary theories that emotions stem from innate affect programs and therefore come as prepackaged entities. Ekman's (1992a) claim that affect programs are solutions to recurrent social problems does not help to draw both theories closer. This is because evolutionary theories provide *innate* solutions to these problems, whereas SCTs provide *culture-sensitive* solutions (Averill, 1980; Harré, 1986; Hochschild, 1979; see M. D. Lewis & Liu, 2011). In evolutionary theories, the (contours of the) solutions are supposed to be baked in during phylogenesis (i.e., the history of the species), whereas in SCTs, the solutions develop during ontogenesis (i.e., the life course of the individual; see Ekman, 1992a). The key evidence for SCTs are cultural differences in emotions, especially the existence of emotions that are unique to some cultures, such as *amae* in Japan (indulgent dependency, e.g., when a child acts over-dependent and the adult indulges; Morsbach & Tyler, 1986) and *tęsknota* in Poland (reversed homesickness, e.g., when a child leaves the family home; Wierzbicka, 1986). For SCTs, it is not enough to say that the same emotions exist in different cultures but that they are elicited by different stimuli or that they are regulated in different ways (Armon-Jones, 1986; see Griffiths, 1997, p. 137). In SCTs, culture does not merely influence the input side (stimulus evaluation) or the output side (regulation) but also the heart of the emotion, that is, the link between stimulus evaluation and action tendencies (Griffiths, 1997, p. 143). This corresponds to a "wax" model of plasticity (Faucher & Tappolet, 2008, p. 112). Wax is plastic to the core. It may lose its plasticity temporally after cooling down but it can become plastic again when heated. Experience may shape and reshape emotions provided that some energy is invested.

Because of their focus on ontogenesis, SCTs address the question of how children come to acquire and apply the socio-cultural knowledge that shapes their emotions. Evolutionary theories hold that children come in the world with basic emotions and that via socialization they learn to refine and regulate these emotions according to cultural display and feeling rules. Parkinson (1995, p. 270) proposed turning around the logic that emotions are first generated in the head and are next expressed to a real audience in the world. He proposed instead that during ontogenetic development, children start out from overt interactions with caregivers in the world, and only after maturation, develop the capacity to represent and simulate these interaction in their head.[94] It is these simulations that we tend to call emotions. According to Parkinson (1995) then, discrepancies that people report between internal emotions and external expressions can be reframed as discrepancies between internal and external social behaviors.

Proponents of recent SCTs (Parkinson, 2012) have criticized older SCTs (Averill, 1980; Harré, 1986) for drawing too sharp an opposition between nature and nurture, and for overemphasizing the role of top-down processes in generating cultural meaning (see also Griffiths, 1997, p. 132). The strict separation of nature and nurture is untenable given the mutual constraining influence of genes and environment (Parkinson, 2012). Genes constrain the environmental factors an organism is sensitive to, and epigenetic mechanisms allow enduring environments to impact on the genes (Crivelli & Fridlund, 2018, 2019; Fridlund, 2017; Griffiths, 1997, p. 159; Mason & Capitanio, 2012). Nevertheless, recent SCTs continue to reject a strong nativism in which emotions are generated by preformed affect programs. According to Parkinson (2012), for instance, humans come into the world with a sensitivity or preparedness for participating in social interactions, but learning still determines the precise shape of these social interactions. Most of this learning, moreover, happens in a bottom-up fashion via operant conditioning and is culturally infused from the very start of life. Top-down application of explicit cultural rules and norms is only the icing on the cake baked earlier by this bottom-up learning.

In addition to detailing the role of socio-cultural information in emotion causation, SCTs have extended the scope of PCTs from a single interaction unit to several interaction units over time. This opens up the possibility that the categorization process of PCTs (Barrett, 2006b) happens in a distributed fashion (Mesquita & Parkinson, 2022). Social interaction allows people to mutually negotiate, and after several iterations, perhaps settle on a shared emotion category (e.g., anger). Here too, the DS approach turns out to be a useful tool for describing the consolidation and dissolution of emotion categories as the moving in and out of attractor states. In line with this distributed picture, several authors have argued that emotions are not in the head, but should be located in the interactions between people. It is not people who are angry, it is the interaction that has angry properties and this is negotiated among the interactants. Some authors put up this "socially extended emotion thesis" as a radical ontological claim about emotions (e.g., Colombetti & Krueger, 2015) while others have argued that shared emotions need not

[94] A remotely similar idea had already been raised by Tomkins (1962), who argued that facial movements are first external after which feedback of them creates affect. In addition, he argued that simulation of this feedback in the absence of external facial movements can also produce affect.

https://doi.org/10.1017/9781107588882.013 Published online by Cambridge University Press

involve a "breakdown of individual boundaries" (León et al., 2019, p. 4856; see also Griffiths & Scarantino, 2005).

A recurrent idea in SCTs is that emotions are strategic (see also Chapter 7). This idea flows naturally from the view that emotions are constructed attempts to reach social goals and to live up to cultural norms (Averill, 1980). Empirical support for the strategic nature of emotions is growing (see reviews by Griffiths, 2004b; Mesquita & Parkinson, 2022; Mesquita et al., 2017; van Kleef et al., 2011). In a study on road rage, for instance, Parkinson (2001a) showed that car drivers reported more anger on than off the road and that this was not predicted by the negativity of the anger-eliciting stimulus nor by the difficulty to regulate anger, but rather by the greater desire and difficulty to communicate. This finding suggests that emotions are strategically used to reach social goals. Another illustration is the finding that emotions that are approved in a culture are more likely to be experienced in that culture (see Mesquita et al., 2017, for a review). For instance, Kitayama et al. (2006) found that in North American culture, which places a higher value on autonomy, self-focused emotions (e.g., pride, anger) were more intense whereas in Japanese culture, which attaches more value to connectedness, other-focused emotions (e.g., friendly feelings, guilt) were more intense. This finding suggests that emotions are strategically used to reach a cultural ideal. In sum, social theories tend to claim that emotions are strategically used to reach goals, whether this is the goal to influence others (Parkinson, 2001a; van Kleef et al., 2011) or the goal to follow cultural norms as a means to fit in (Mesquita et al., 2017).

It is important to reiterate the point made in Chapter 7 that the strategic use of emotions is uncontroversial in itself. Evolutionary theories and SETs already accepted that emotions can be faked as a form of antecedent-focused emotion regulation. The bolder claim made by SCTs is that the emotions thus created are equally emotional as those that are less strategic (Averill, 1980). The boldest claim, however, is made by the goal-directed theory (Moors, 2017a). This theory holds that so-called emotions are *nothing but* strategic behavior episodes. There is no strategic use of otherwise non-strategic emotions. There are simply strategic behavior episodes, full stop. And some of these episodes are singled out and labeled as emotions because they are conspicuous.

9.5 Response Evaluation Theories (RETs)

9.5.1 Personal Version

In RETs such as the goal-directed theory (Moors, 2017a), socio-cultural factors influence the values of goals and the magnitude of the response-outcome expectancies. The values of goals are acquired early in life via the reward structure of the physical and cultural environment. This enables the goal-directed theory to naturally account for cultural differences in alleged emotions. If the goal for autonomy is valued more in individualist cultures and the goal for connectedness more in collectivist cultures, a restriction of freedom creates a larger discrepancy in individualist cultures while social disapproval creates a larger discrepancy in collectivist cultures. Larger discrepancies, in turn, may translate in more intense, control-precedent urges to reduce this discrepancy.

So far, the predictions of the goal-directed theory converge with those of the first extension of SETs discussed above. In addition to the value of goals, however, the goal-directed theory also emphasizes the expectancies of action options for reaching these goals. Here too, cultural infusion is possible. The action options selected in different cultures to reduce discrepancies may differ because they hold different social costs and benefits. Assertive behavior may be accepted more in individualist cultures and conciliatory behavior more in collectivist cultures. Finally, the goal hierarchy central to the goal-directed theory sheds light on the coercive force that seems to emanate from cultural norms (e.g., Parkinson, 2020). Like SETs, the goal-directed theory predicts that cultural norms that are not endorsed should remain inert. Nevertheless, not endorsing a cultural norm may prove to be no mean feat because noncompliance is punished by social disapproval and even exclusion. To the extent that people value being part of a group, it may be hard to disobey the group's norms. In goal-directed lingo, a cultural norm has a high value if it is a subordinate goal with a high expectancy for reaching the valued superordinate goal of social approval or inclusion.

9.5.2 Social Version

A social theory that shows similarity with the goal-directed theory is the behavioral-ecological theory of facial behavior proposed by Fridlund (1994, 2017; Crivelli & Fridlund, 2018, 2019; Fridlund & Russell, 2022). Although this theory files as an evolutionary theory of *facial behavior*, it has nothing in common with the evolutionary theories of *emotion* discussed so far. Fridlund (1994) argued, contra these evolutionary emotion theories, that facial behaviors do not communicate or express emotions (e.g., anger), but are declarations of social intentions (e.g., the tendency to attack) or of requests for behavior change from the part of the receiver (e.g., the request to concede). Stated in this way, however, there is a risk of misunderstanding Fridlund (1994) as saying that facial behaviors *communicate* social intentions or social requests. In more recent writings, however, this ambiguity has been resolved (e.g., Crivelli & Fridlund, 2018, 2019; Fridlund, 2017; Fridlund & Russell, 2022). Facial behaviors do not communicate social intentions or requests; they do not communicate anything. They are mere social tools or strategies to reach a particular social goal. Returning to the taxonomy of behavior presented at the start of this chapter, Fridlund's (2017) facial behaviors must be situated in the category of *pure indirect* social behaviors that are caused by the tendency to influence the receiver in a particular way (see Figure 9.1). So the difference between Fridlund (2017) and evolutionary emotion theories is not merely a difference about whether facial behaviors are *emotional* or not, but also whether facial behaviors are *communicative* or not. That being said, Fridlund (2017) does not exclude the possibility that some facial behaviors become used as tools for communication, but these are specific cases and they can never communicate emotions. The latter rejection is rooted in a skeptical view about the scientific status of emotions, which is in strong agreement with the goal-directed theory (Moors, 2017a) and Russell's (2003) PCT. There is one caveat, however. The trouble with communication is that anything can be communicated. The content of a message does not have to be something that really exists. Sam can communicate that he is angry just as he can communicate that he has supernatural powers or that he is a unicorn. So if a sender has the intention to express what she believes is anger, then she is expressing anger, even if she is mistaken about being angry. Likewise, a receiver may observe anger in a face if she possesses this concept. Again, there are no limits to what a receiver can see in a sender's face, just as there are no limits to the stellar constellations that one can see in the sky.

Fridlund (2017) also proposed the following ontogenetic explanation for social facial behaviors. Infants do not come in the world with facial behaviors that have an innate social function. The raw material consists of facial behaviors that are (a) non-social protective reflexes such as the startle reflex, (b) non-social behaviors that prepare for certain social behaviors, such as pursing the lips before lashing out or averting gaze before hiding, or (c) non-social goal-directed behaviors such as scrunching the noise to reduce air intake and widening the eyes to increase information intake. Once these facial behaviors start having a consistent effect on a receiver (she gives support, engages in play, or concedes) that is rewarding to the sender, the sender may start using these facial behaviors to achieve these social effects. At that point, facial behaviors have become social behaviors that are goal-directed. To make the effects in the receiver more likely, the sender may exaggerate them (i.e., ritualization). Finally, in humans at least, these goal-directed social behaviors may acquire a meaning and this meaning may be stored and used occasionally in communication. Examples are symbolized facial behaviors such as tongue protrusions and certain smiles. The trajectory described here, however, can be marked as an evolution from the outside (behavior) to the inside (meaning), and not the other way around. Children learn to engage in goal-directed social behavior before they learn to attach a meaning to it (see also Parkinson, 1995).

Empirical support for the idea that facial behaviors reflect social intentions comes from audience effects (Fridlund, 1994; see reviews in Crivelli & Fridlund, 2018, 2019; Griffiths & Scarantino, 2005; Parkinson, 2005). Fernández-Dols and Ruiz-Belda (1997), for instance, found that winning athletes smiled less when they were waiting behind the podium for the awards ceremony to start than when they interacted with the audience. Assuming that these athletes were equally happy about their performance throughout, but that the need to seek affiliation was lower behind the podium than when interacting, these findings support the idea that facial behaviors serve social goals rather than expressing emotions. In a laboratory study by Schneider and Josephs (1991), it was found that children in a competitive game smiled more in interactive than non-interactive conditions and that smiling was even higher among losers than winners, thus undermining any direct dependence on happiness.

Two objections have been raised by proponents of the evolutionary theory of emotion. The first is that evidence that facial behaviors can convey social intentions is not yet evidence against the idea that they can also convey emotions. Audience effects can easily be accommodated in the evolutionary theory where they would be explained by the intervention of display rules (i.e., consequent emotion regulation) or plain faking (i.e., antecedent emotion regulation). Evidence consistent with the role of display rules came from Ekman's (1971) American-Japanese study described above. The second objection is that people still present facial expressions when they are alone, and that even when they try to suppress or mask their expressions in the presence of others, expressions of true emotions may leak out pointing at the tension between emotions and regulation (Ekman & Friesen, 1969).

In response to these objections, Fridlund (1994) argued that being alone does not rule out audience effects. This is because when people are alone, they may still communicate to the audience "in their head" (see also Parkinson, 1995). Evidence for the audience-in-the head hypothesis (i.e., implicit sociality) was obtained by Fridlund et al. (1990, 1992), who showed that participants who were alone but were asked to imagine an audience showed more facial behaviors than participants in a control group. Ekman (1997) later integrated this finding by arguing that when people are alone they may also follow private display rules. As noted by Fridlund (1994), however, such a turn would undermine the conclusions previously drawn from the American-Japanese study. It would mean that the facial behaviors registered in the alone condition no longer reflect pure emotions.

https://doi.org/10.1017/9781107588882.013 Published online by Cambridge University Press

Crucially for our purpose, audience effects can be interpreted as evidence for the goal-directed origin of facial behaviors if it can be shown that the presence or absence of an audience alters the expected utilities of these facial behaviors. Indeed, it seems plausible to assume that certain facial behaviors (e.g., smiles) have a higher expected utility when a real audience is present than when such an audience is absent because smiling to a real audience may solicit real social approval. But even if a real audience is absent, the expected utility may not be zero. I see two options. The first option is that the person practices facial behaviors with an eye on future encounters with a real audience. The second option is that when people are alone, stored expected utilities still get activated and/or simulated, similar to what happens in the case of emotions elicited by fiction (see Chapter 7).

Another piece of evidence for the goal-directed origin of facial behaviors comes from studies showing that emotional contagion may not be as encapsulated as generally assumed but may instead depend on a cost-benefit analysis. For instance, people imitate smiles when imitation fulfills the goal to affiliate and when the cost of smiling remains low (Bourgeois & Hess, 2008; Hess & Fischer, 2013; Lakin et al., 2008).

By applying the goal-directed theory to indirect social behaviors, whether communicative or pure, the scope of the theory is significantly expanded. Most notably, the theory is now capable of explaining a particular class of practically irrational behaviors known as arational behaviors (Hursthouse, 1991; Scarantino & Nielsen, 2015; see Chapter 2).[95] Classic examples of arational behaviors are ones for which an emotion can easily be invoked as their cause, such as (a) stroking someone's hair out of love, (b) jumping up and down out of joy, (c) covering one's face in the dark out of shame, (d) slamming the door out of anger, and (e) kissing someone's picture out of grief (examples adapted from Hursthouse, 1991). According to Scarantino and Nielsen (2015), the first three examples can be explained by emotions alone, whereas the last two examples are the result of the interplay between emotions and regulation and/or planning. The first three examples (a, b, c) are behaviors that fulfill the action tendencies characteristic for their respective emotions. Stroking someone's hair fulfills the action tendency of love, which is to seek proximity. Jumping up and down fulfills the action tendency of joy, which is to broaden and build. Covering one's face in the dark fulfills the action tendency of shame, which is to disappear from sight. The last two examples (d, e) are behaviors that fulfill the action tendencies characteristic of their emotions, but which need to be

[95] It may be kept in mind that the task that theories face is to explain the apparent irrationality of behavior. They must not necessarily conclude that these behaviors are genuinely irrational.

displaced to a different target either because the costs of picking the appropriate target would be too high so that regulation is required or because opportunities for execution are not available. Slamming the door out of anger partly fulfills the action tendency of anger, which is to fight, but seeks out an inanimate target because of emotion regulation (the door cannot retaliate). Kissing someone's picture out of grief partly fulfills the action tendency of getting closer to the lost person, but for lack of opportunity, seeks out an object that comes close to this person.

The goal-directed theory can explain all these cases with a single, goal-directed process, without the need to invoke emotions. Stroking someone's hair fulfills the tendency to seek proximity, but the goal to seek proximity need not be an emotion. Jumping up and down when a goal is achieved may be a way to open up and explore new opportunities for goal fulfillment, or it may be a way to show to an audience that one has achieved a goal. Even when one is alone, the audience may still be in the head. Likewise, covering one's face in the dark after norm transgression may be done for an audience in the head. That there is no need to invoke an emotion becomes clear if we compare these cases with the case in which people gesticulate when talking on the phone about a trivial matter. Slamming the door instead of slapping the harm-doer may likewise be an act of goal-directed communication. Finally, kissing the picture of a lost love may be a form of goal-directed communication, perhaps with the lost love taking the role of the audience in the head. Alternatively, it may be a way to get mentally close to the lost love. In conclusion, I believe that when the only mechanism in our toolbox is the goal-directed process, we get far enough to explain alleged emotional phenomena, without having to invoke the existence of emotions with their own dedicated mechanisms.

A final way to socially extend the goal-directed theory is to consider the possibility that the goal-directed process gets distributed over several iterations in the social interaction. Instead of demanding that the system relies on complex stored and computed values and expectancies, natural social interactions often provide ample opportunity for trial-and-error. Thus, instead of having to rely on a one-shot weighing up of a large number of benefits and costs of an aggressive outburst versus a calm conversation, the sender may gradually raise her voice and calibrate the expected utilities of both behaviors bit by bit depending on the receiver's reactions.

In summary, personal theories have developed their mechanisms for emotion causation while accepting the predominantly social nature of eliciting stimuli and the demands of social environments in terms of emotion regulation. The inadequacy of these mechanisms to account for

https://doi.org/10.1017/9781107588882.013 Published online by Cambridge University Press

the complexity that is inherent in social interactions with multiple interaction partners and across time has inspired the birth of social extensions of personal theories, called social theories. Not all social theories (e.g., SCTs) grew historically from personal theories, though, even if they can still be seen as affiliative with one or more of them (e.g., PCT, SET). Social theories have eventually been able to do more than merely add an eagle eye view to social interactions. They have also been able to point at genuinely social mechanisms – emotional contagion, social appraisal, direct relation alignment, and distributed forms of social appraisal and goal-directed processes – that have enriched the mechanistic toolbox.

PART III

Conclusion

https://doi.org/10.1017/9781107588882.014 Published online by Cambridge University Press

CHAPTER 10

Conclusion

The aim of this book was twofold: to work out a new typology for comparing various emotion theories from psychology and philosophy and to propose and defend my own goal-directed theory of emotions that can be situated within this typology. I engaged in an in-depth exercise in which I compared a number of theories from psychology and philosophy making use of a framework inspired by lessons from philosophy of science. In this concluding chapter, I highlight the main insights gained from this exercise. Ideally, attempts towards analyzing a domain into smaller parts should be followed by attempts towards synthesizing the parts in a coherent whole. Therefore, I close by briefly exploring the question of whether there are ways to reconnect the various theories and whether the creation of a unified theory of emotion, or at least a general framework for conducting emotion research, might be within reach.

10.1 Typology of Emotion Theories

As the basis for the typology of emotion theories, I started from an idealized path towards theory development, called the demarcation-explanation cycle, which ultimately attempts to answer the questions of what emotions are in terms of how they should be demarcated from other phenomena and of how their variety can best be organized. Stages in the cycle are (a) the provisional demarcation of the set of emotions (explanandum) from other phenomena in a working definition, (b) the search for explanations (linking explananda to explanantia), (c) the validation of these explanations in empirical research, and (d) the proposal of a scientific definition in which instances of the set of emotions are held together by the explanantia. Each of these stages raises questions to which different theories have given similar or different answers. The questions form the axes for a typology, the answers provided by the theories form the positions occupied by the theories on these axes (see Table 2.4).

Theories of emotion show a substantial degree of overlap in the working definitions that they propose in Stage 1. They agree that an

https://doi.org/10.1017/9781107588882.015 Published online by Cambridge University Press

adequate theory of emotion should explain phenomena known as fear, anger, sadness, joy, and so forth, as well as their typical and apparent properties. These include mental and bodily changes that occur in individuals (across the ontogenetic and phylogenetic ladder) after they encountered events of major importance. These changes, moreover, are characterized by heat (in terms of valence and intensity), automaticity, control precedence, and irrationality (see Table 2.1).

It is when moving to Stage 2, in which constitutive and causal-mechanistic explanations of the phenomena called emotions are in order, that theories begin drifting apart. The constitutive explanations range from narrow (including a minimal number of components) to broad (including a maximal number of components). Yet if the explanandum is reconstituted from "emotion" to "emotional episodes," some differences between theories melt away. This is because theories roughly agree on the components to include in emotional episodes, whether these are sampled from the traditional list (cognitive, motivational, somatic, motor, and feeling; see Table 2.2) or from the novel list proposed in Chapter 2 (external and internal stimuli, afferent and efferent representations that can be unconscious or conscious, somatic responses, and motor responses; see Table 2.3).

The most pronounced differences between theories reside in the causal-mechanistic explanations that they propose, and by implication, in the empirical research set up to validate these explanations in Stage 3 and the scientific definitions flowing from them in Stage 4. This is why I took the causal-mechanistic axis as the primary basis for differentiating theories and for labeling them. The five families of emotion theories thus differentiated were (a) evolutionary theories with affect programs (Chapter 4), (b) network theories with emotion networks in memory (Chapter 5), (c) SETs with stimulus evaluation (Chapter 6), (d) RETs with response evaluation (Chapter 7), and (e) PCTs with diffuse bodily feelings and a construction process (Chapter 8) (see Figure 10.1). Several members of these theoretical families are hybrids. Examples are evolutionary-evaluation hybrids with an emphasis on the evolutionary part and evaluation-evolutionary hybrids with an emphasis on the evaluation part. Also, many network theories are in fact network-evaluation hybrids. Each of these five "personal" theories (discussed in Chapters 4–8) has given rise to "social" versions (discussed in Chapter 9) in which increased attention to the social embeddedness of individuals yielded specifically social mechanisms such as emotional contagion, social appraisal, direct relation alignment, and distributed forms of social appraisal and goal-directed processes. Prior to my discussion of these theories, I zoomed in on Darwin (1872) and James (1890b) (Chapter 3), who are often, and with good reason, presented as general precursors for subsequent theories.

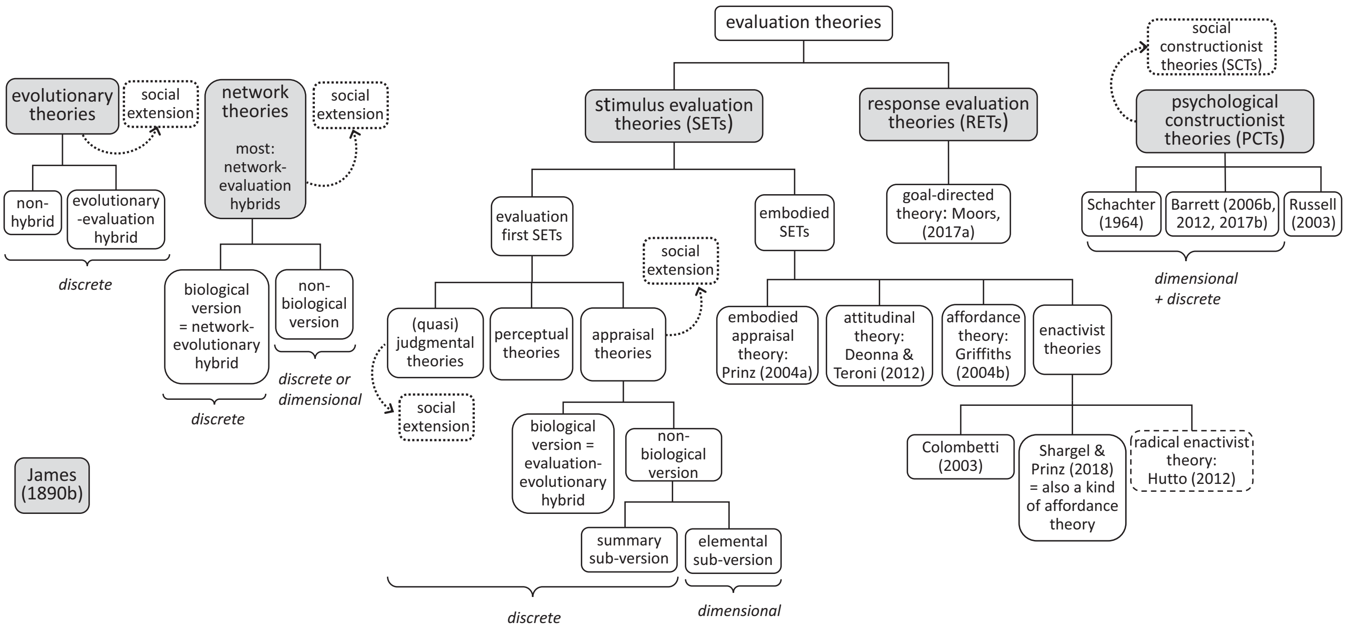


Figure 10.1 Typology of emotion theories based on the causal-mechanistic axis

If we stick to the causal-mechanistic axis, five fault lines between theories are worth reiterating. A first fault line has to do with whether bodily components are explained by a stimulus-driven or goal-directed mechanism (**Axis 6b2, FL**). Darwin (1872) kicked off with instincts and habits, although these were still in the form of [mental-state–R] links. James (1890b) embroidered further on Darwin's (1872) mechanisms, modifying his instincts and habits into [S–R] links. James's (1890b) theory can be considered as the center of a patchwork of theories that has been woven ever since. Theories have claimed to adjust or overthrow James's (1890b) theory, but many subsequent theories have remained loyal to the idea that [S–R] links, whether innate or learned, form the backbone of emotions, thereby explaining the automaticity, control precedence, and irrationality of emotions. Within these "[S–R]" theories, I distinguished between (a) evolutionary theories, which focus on innate [S–R] links, (b) network theories, which put the center of gravity on learned [S–R] links, and (c) SETs, which propose sophisticated [S–R] links, augmented by a stimulus evaluation process.

Evolutionary theories count only the innate [S–R] links (i.e., primary processes) as pure emotions, the hard liquor that can get mixed with learning and computation (i.e., secondary and tertiary processes). Learning and computation precede or follow the emotional core, but do not completely wash it away. Network theories, on the other hand, allow learning processes to overwrite the innate [S–R] links to a greater or lesser extent but without thereby sacrificing emotionality. So considered, evolutionary and network theories can be placed on a continuum for which the end points represent a larger role for innate vs. learned processes, respectively. If we were to identify a tipping point, however, it would be that evolutionary theories take innateness to be the mark of the emotional whereas network theories expand this to learning, as long as this learning still yields automatic, control precedent, and rigid action tendencies.

The main raison d'être for SETs is to provide the Intentional object of emotions that the previous theories could not account for in a satisfactory way. They do this by adding a stimulus evaluation process. I distinguished between two brands of SETs. Evaluation-first SETs plant a stimulus evaluation process prior to activation of the abstract [S–R] link. This evaluation process produces abstract stimulus features for the [S] part. Embodied SETs were designed to alleviate the worry that a purely mental evaluation process would be too abstract and too cold to produce hot emotions. Some embodied SETs hold that individuals infer evaluative stimulus properties from their bodily reactions and/or action tendencies (Deonna & Teroni, 2012; J. J. Prinz, 2004a); others try to close the gap between stimulus evaluation ([S]) and action tendencies ([R]) by gluing them together (Colombetti, 2003; Griffiths, 2003; Shargel & Prinz, 2018).

Turning to the PCTs of Schachter (1964) and Barrett (2012), it may be recalled that their primary explanandum are feelings and not bodily components. Labeled feelings result when a construction process binds diffuse bodily feelings to external stimuli based on prior knowledge. Nevertheless, both theories have suggested that after labeled feelings are constructed, they may in turn dictate action tendencies and behavior. In Barrett's (2012, p. 419) words, "[a] change in heart rate (X) can function as a feeling of offense (Y1) or a feeling of threat (Y2), depending on the situated conceptualization that is constructed, and each will dictate a different action tendency." Or to use a more situated example, an adrenaline rush categorized as an instance of anger-after-being-cut-in-line-in-the-supermarket may activate the tendency to nonverbally communicate disapproval whereas the same adrenaline rush categorized as an instance of anger-after-being-falsely-accused may activate the tendency to protest. With a little imagination, this [emotion–R] link can be considered as another sophisticated [S–R] link in which the [S] part corresponds to the abstract representation of sensory input in terms of a specific emotion ([aS]). The crucial point is this: The action tendency is ultimately dictated by (external and internal) stimulus information rather than by weighing up the costs and benefits of available action options.

The only RET discussed in this book, the goal-directed theory (Moors, 2017a), breaks with the idea that emotional action tendencies are part of innate or learned [S–R] links, whether simple or sophisticated. Instead, this theory proposes that action tendencies, including those that are called emotional, are caused by goal-directed processes that weigh up the expected utilities of the available action options ([S:R–O^v]). Although the theory does not deny the existence of [S–R] links, it considers them to be weak enough to mostly lose out against goal-directed processes. While SETs focus on stimulus evaluation, the goal-directed theory shifts focus to response evaluation. That said, stimulus evaluation still has a place in this theory in that it establishes the need for behavior (i.e., detection of stimulus-goal discrepancy) and serves as a source of information to determine the expected utilities of behavior options (i.e., in action selection). For instance, a rabbit meeting a wolf first has to detect an imminent discrepancy with its goal for safety, and in order to choose to flee, it needs to evaluate whether a rabbit hole or other escape route is in the vicinity. The goal-directed theory can account for the desiderata of heat, automaticity, control-precedence, and (apparent as well as real) irrationality. In addition, it can account for a rich form of Intentionality that combines world-directedness (mind-to-world) and self-directedness (world-to-mind).

https://doi.org/10.1017/9781107588882.015 Published online by Cambridge University Press

A second fault line between theories centers on whether they locate the origin of feelings peripherally or centrally (**Axis 6a2*, FL**). In some theories (Damasio, 2004; Deonna & Teroni, 2012; James, 1890b; Levenson, 2014; Prinz, 2004b), feelings stem from peripheral bodily responses that are fed back to the cognitive apparatus. In other theories, feelings are produced centrally. In evaluation-first SETs, such as perceptual theories, feelings are generated by the cognitive component (i.e., stimulus evaluation). The mere perception of danger, for instance, determines the feeling of fear. Some evolutionary theories have followed Cannon (1927) and McDougall (1908) in arguing that the activation of instincts or affect programs in the brain (i.e., link between cognitive and motivational component) generates labeled feelings. Some of these theories, moreover, advocate that the qualia of these centrally generated feelings are not reducible to the cognitive and motivational components that make up these instincts or affect programs. In still other theories, feelings are as much a central as a peripheral affair. In non-biological appraisal theories, for instance, the cognitive and motivational components account for a large part of the variance but they are accompanied by conscious traces of somatic and motor components. For instance, the feeling of fear reflects an appraisal of goal-incongruence and uncontrollability combined with the urge to flee, but this feeling typically also includes traces from an adrenaline rush, a fearful face, and actual fleeing behavior. The feelings from appraisals and action tendencies are centrally generated; those of somatic and motor responses count as peripherally generated.

A third and fourth fault line between theories has to do with whether the mechanisms they propose are dedicated to emotions or not (**Axis 6f, FL**) and flowing from this, whether this gives them a ground for setting emotions apart as a scientific set (**Axis 8, FL**). Several theories deliver a special-purpose mechanism for emotions. Evolutionary theories put forward affect programs, and so do biological network theories and biological SETs in their wake. The stimulus evaluation process proposed by SETs and the construction process proposed by Schachter (1964) and Barrett (2006b) qualify as general-purpose in terms of the operations involved, but as special-purpose in terms of the content of their output representations. Trusting that they have found the mark of the emotional, these vindicator theories have come to accept emotions as a scientific set, or at least as a collection of scientific sets. By contrast, James's (1890b) theory, Moors's (2017a) goal-directed theory, and Russell's (2003) PCT propose entirely general-purpose mechanisms, thereby dissolving the boundary between emotions and non-emotional phenomena. These theorists all count as skeptics turned eliminativists because they doubt that a mark of the emotional will ever be found. The biggest virtue of the goal-directed theory is that it combines a broad scope with parsimony. Indeed,

https://doi.org/10.1017/9781107588882.015 Published online by Cambridge University Press

it fulfills all the desiderata that other emotion theories cover but manages to do so without postulating a special-purpose mechanism.

A fifth fault line, which can only be meaningfully considered for vindicator theories, revolves around how to best organize the variety within the set of emotions (**Axis 9, FL**). Theories with a discrete view propose a mechanism for emotion causation that allows partitioning the set of emotions into discrete subsets, corresponding more or less to the prototypical emotions we know from natural language. Examples are evolutionary theories and biological versions of network theories and SETs. In theories with a dimensional view, by contrast, the mechanisms put forward create variety that is more sparingly described with the help of dimensions. Elemental appraisal theories, which organize the set of emotions with the help of appraisal dimensions, fit into this picture. A final category of theories combines a discrete with a dimensional view. Here, two-factor PCTs (e.g., Barrett, 2006b; Schachter, 1964) count as examples.

10.2 Unifying Framework?

Now that I have delineated the most important differences between emotion theories, the question arises of whether there is any hope for a synthesis of several of these theories into a unified theory of emotion, or at least a general framework for conducting emotion research. If the ideal is to maximize the contribution of all theories reviewed, an additive approach is recommended that hoists on board as many mechanisms as possible. Such an approach would require relaxing strong claims about the necessity of mechanisms or about the exclusivity of some mechanisms to the exclusion of others. It may also require that some mechanisms be allocated a different role than what authors originally had in mind. Moreover, to give this enterprise any chance of success, the mechanisms would also have to be spelled out in very general terms, abstracting from idiosyncratic details. A unified theory should map out all the logically coherent possibilities and leave them open for empirical testing – to the extent that they are indeed testable.

A complete hybrid theory might look somewhat like this: A stimulus is first perceived and/or evaluated, using mechanisms that range from simple to complex. The information extracted from the stimulus can subsequently lead to action tendencies, either via an innate or learned stimulus-driven process or a goal-directed process. The action tendency recruits peripheral somatic activity that prepares and supports overt behavior. All these components find their way into the feeling component (provided that they are intense enough). The mental components (cognitive, motivational) do so directly because they already involve

https://doi.org/10.1017/9781107588882.015 Published online by Cambridge University Press

representations. The bodily components (somatic, motor) do so indirectly, after their immediate outcomes are perceived. This comes down to accepting that feelings can be generated both centrally and peripherally. Together, the mental components and perceived bodily components provide the raw feeling quality. All components can also be entered into a construction process that categorizes the raw feelings in terms of specific emotions resulting in labeled feelings. If the stimulus evaluation at the start of the chain is commensurate with the quality and intensity of the raw experience, this construction occurs naturally. Otherwise, an active search process is initiated which may or may not end in a satisfactory "solution." All components that occur in close temporal contiguity may wire together in a network, after which activation of one component can spread to the other components. This opens up the possibility that components not only get activated from the stimulus side but also from the response side.

The additive approach has the potential to maximize the scope, that is, the set of empirical phenomena that can be explained, but it comes at an obvious cost of parsimony. The main argument for starting from a broad set of mechanisms is that the current state of empirical research is not advanced enough yet to allow definitively shutting the door on any of the mechanisms proposed by contemporary theories. The task for future research is two-fold.

A first task is to (continue to) examine the existence of mechanisms. This will eventually allow us to prune the broad set of potential mechanisms into a more limited set, thereby increasing parsimony. Let me speculate about possible ways in which this pruning might take shape. In the ongoing stream of events that bombard us every second of our waking and sleeping life, stimulus evaluation seems a no-brainer. SETs are right in emphasizing this, but the question remains of which types of information are processed. From the perspective of the goal-directed theory defended in this book, stimuli are evaluated in terms of whether and to what extent they present a discrepancy with goals (i.e., goal congruence and goal relevance). Other stimulus features are processed insofar as they help select a suitable (overt or covert) behavior. The latter stimulus features need not correspond to the (molar and/or molecular) appraisals proposed in SETs. Crucially, the goal-directed theory insists that stimulus evaluation is not (or only rarely) sufficient to determine the action tendency that should be activated. Organisms inevitably and effortlessly process the expected utilities of the available behavior options. If operating conditions are poor, fewer outcomes may be considered for fewer behavior options than when operating conditions are ample, but organisms will rarely engage in a behavior without foreshadowing the immediate impact of this behavior for at least one valued outcome. A similar exercise can be conducted for the mechanisms involved in the emergence of feelings. Here, I see no reason, however, to abandon the earlier idea that all components can leave their trace in consciousness and hence can contribute to the eventual experience.

A second task for future research is to examine the interplay between the mechanisms that survived the pruning exercise of the first task. Demonstrating the existence of a mechanism is one thing. Demonstrating its explanatory power when in competition with other mechanisms is another. Thus, demonstrating the existence of associations between representations with various contents is not the same as demonstrating that these associations play a substantial role in behavior and feelings. For instance, certain instincts (innate [S–R] links) may well exist at the start of an organism's life, after which they are modified and/or have to compete with goal-directed processes. Regardless of whether the activation of innate or learned [S–R] links happens via the stimulus side (via stimulus presentation) or via the response side (via artificial induction of responses), this process may eventually turn out to be so weak that it is unlikely to survive the competition from goal-directed processes. Turning again to feelings, I would speculate that all traces from all representations coagulate in some kind of Gestalt in which the dominant flavor is determined by the representations that were activated most strongly.

But no amount of research aimed at dissecting the emotional "elephant" (see Russell & Barrett, 1999) can keep us ignoring the elephant in the room, which is that the unifying theory of emotions thus construed does not provide a principled way to distinguish between the things called emotion and other things, and hence does not qualify as a theory of emotions properly speaking. I am not the first to draw this skeptical conclusion (Duffy, 1941a; Fridlund, 1994; Russell, 2003) and probably will not be the last, but the struggle was and continues to be an uphill one. People are reluctant to give up on folk terms that serve them well in one way or the other (Averill, 1980). If emotions have been compared to elephants, emotion theories have been compared to whales in a quote from Meyer (1933, p. 292) previously cited in Fridlund (1994, p. 185) and crispy enough to repeat: "The whale has a twofold distinction among the fishes: first, when seen from a distance, it looms large among them; and second, on close examination, it is found to be no fish at all. Something like that I predict for the theory of emotions among the theories in psychological textbooks and periodicals."

References

Aarts, H., & Dijksterhuis, A. P. (2000). The automatic activation of goal-directed behaviour: The case of travel habit. *Journal of Environmental Psychology*, *20*(1), 75–82. https://doi.org/10.1006/jevp.1999.0156

Adams, C. D. (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. *The Quarterly Journal of Experimental Psychology Section B*, *34*(2b), 77–98. https://doi.org/10.1080/14640748208400878

Adams, C. D., & Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *The Quarterly Journal of Experimental Psychology Section B*, *33*(2b), 109–121. https://doi.org/10.1080/14640748108400816

Adelman, P. K., & Zajonc, R. (1989). Facial efference and the experience of emotion. *Annual Review of Psychology*, *40*, 249–280. https://doi.org/10.1146/annurev.ps.40.020189.001341

Ainslie, G. (2001). *Breakdown of will*. Cambridge University Press. https://doi.org/10.1017/cbo9781139164191

Ainslie, G., & Monterosso, J. (2005). Why not emotions as motivated behaviors? *Behavioral and Brain Sciences*, *28*(2), 194–195. https://doi.org/10.1017/s0140525x05220046

Allen, J. B., Kenrick, D. T., Linder, D. E., & McCall, M. A. (1989). Arousal and attraction: A response-facilitation alternative to misattribution and negative-reinforcement models. *Journal of Personality and Social Psychology*, *57*(2), 261–270. https://doi.org/10.1037/0022-3514.57.2.261

Allport, F. H. (1924). *Social psychology*. Houghton Mifflin.

Armon-Jones, C. (1986). The thesis of constructionism. In R. Harré (Ed.), *The social construction of emotions* (pp. 32–56). Blackwell.

Armon-Jones, C. (1991). *Varieties of affect*. Harvester Wheatsheaf. https://doi.org/10.3138/9781487571887

Arnold, M. B. (1960). *Emotion and personality*. Columbia University Press.

Arntz, A., Rauner, M., & Van den Hout, M. (1995). "If I feel anxious, there must be danger": Ex-consequentia reasoning in inferring danger in anxiety disorders. *Behaviour Research and Therapy*, *33*(8), 917–925. https://doi.org/10.1016/0005-7967(95)00032-s

Averill, J. R. (1969). Autonomic response patterns during sadness and mirth. *Psychophysiology*, *5*(4), 399–414. https://doi.org/10.1111/j.1469-8986.1969.tb02840.x

Averill, J. R. (1975). A semantic atlas of emotion concepts. *JSAS Catalog of Selected Documents in Psychology*, *5*, 330 (Manuscript # 421).

Averill, J. R. (1980). A constructivist view of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research, and experience* (Vol. 1, pp. 305–339). Academic Press. https://doi.org/10.1016/b978–0-12-558701-3.50018-1

Averill, J. R. (1982). *Anger and aggression: An essay on emotion*. Springer-Verlag. https://doi.org/10.1007/978-1-4612-5743-1

Averill, J. R. (2012). The future of social constructionism: Introduction to a special section of *Emotion Review*. *Emotion Review*, *4*(3), 215–220. https://doi.org/10.1177/1754073912439811

Ax, A. F. (1953). The physiological differentiation between fear and anger in humans. *Psychosomatic Medicine*, *15*(5), 433–442. https://doi.org/10.1097/00006842-195309000-00007

Bagozzi, R. P., & Pieters, R. (1998). Goal-directed emotions. *Cognition and Emotion*, *12*(1), 1–26. https://doi.org/10.1080/026999398379754

Bain, A. (1864). *The senses and the intellect*. Longman, Green, Longman, Roberts, and Green. https://doi.org/10.1037/12243-000

Bakker, B. (2005). The concept of circular causality should be discarded. *Behavioral and Brain Sciences*, *28*(2), 195–196. https://doi.org/10.1017/s0140525x05230042

Balleine, B. W., Dezfouli, A., Ito, M., & Doya, K. (2015). Hierarchical control of goal-directed action in the cortical–basal ganglia network. *Current Opinion in Behavioral Sciences*, *5*, 1–7. https://doi.org/10.1016/j.cobeha.2015.06.001

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology*, *37*(4–5), 407–419. https://doi.org/10.1016/s0028–3908(98)00033-1

Barbalet, J. M. (1999). William James' theory of emotions: Filling in the picture. *Journal for the Theory of Social Behaviour*, *29*(3), 251–266. https://doi.org/10.1111/1468-5914.00101

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 1, pp. 1–40). Erlbaum. https://doi.org/10.4324/9781315807102

Bargh, J. A., & Barndollar, K. (1996). Automaticity in action: The unconscious as repository of chronic goals and motives. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 457–481). Guilford.

Bargh, J. A., Gollwitzer, P. M., & Oettingen, R. (2010). Motivation. In S. Fiske, D. Gilbert, & R. Lindzey (Eds.), *Handbook of social psychology* (5th ed., pp. 268–316). Wiley. https://doi.org/10.1002/9780470561119.socpsy001008

Bargh, J. A., & Williams, L. E. (2007). The nonconscious regulation of emotion. In J. J. Gross (Ed.), *Handbook of emotion regulation* (pp. 429–445). Guilford.

Barlassina, L., & Newen, A. (2014). The role of bodily perception in emotion: In defense of an impure somatic theory. *Philosophy and Phenomenological Research*, *89*(3), 637–678. https://doi.org/10.1111/phpr.12041

Barnes-Holmes, D., & Hughes, S. (2013). A functional approach to the study of human emotion. In D. Hermans, B. Rimé, & B. Mesquita (Eds.), *Changing* *emotions* (pp. 202–209). Psychology Press. https://doi.org/10.4324/9780203075630-37

Barrett, L. F. (2006a). Are emotions natural kinds? *Perspectives on Psychological Science*, *1*(1), 28–58. https://doi.org/10.1111/j.1745-6916.2006.00003.x

Barrett, L. F. (2006b). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review*, *10*(1), 20–46. https://doi.org/10.1207/s15327957pspr1001_2

Barrett, L. F. (2011). Was Darwin wrong about emotional expressions? *Current Directions in Psychological Science*, *20*(6), 400–406. https://doi.org/10.1177/0963721411429125

Barrett, L. F. (2012). Emotions are real. *Emotion*, *12*(3), 413–429. https://doi.org/10.1037/a0027555

Barrett, L. F. (2013). Psychological construction: The Darwinian approach to the science of emotion. *Emotion Review*, *5*(4), 379–389. https://doi.org/10.1177/1754073913489753

Barrett, L. F. (2014). The conceptual act theory: A précis. *Emotion Review*, *6*(4), 292–297. https://doi.org/10.1177/1754073914534479

Barrett, L. F. (2017a). *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt.

Barrett, L. F. (2017b). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1–23. https://doi.org/10.1093/scan/nsx060

Barrett, L. F. (2022). Constructionist accounts of emotion. In A. Scarantino (Ed.), *Routledge handbook of emotion theory*. Routledge.

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, *20*(1), 1–68. https://doi.org/10.1177/1529100619832930

Barrett, L. F., & Finlay, B. L. (2018). Concepts, goals and the control of survival-related behaviors. *Current Opinion in Behavioral Sciences*, *24*, 172–179. https://doi.org/10.1016/j.cobeha.2018.10.001

Barrett, L. F., Gendron, M., & Huang, Y. M. (2009). Do discrete emotions exist? *Philosophical Psychology*, *22*(4), 427–437. https://doi.org/10.1080/09515080903153634

Barrett, L. F., Gross, J., Christensen, T. C., & Benvenuto, M. (2001). Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition and Emotion*, *15*(6), 713–724. https://doi.org/10.1080/02699930143000239

Barrett, L. F., Lindquist, K. A., Bliss-Moreau, E., Duncan, S., Gendron, M., Mize, J., & Brennan, L. (2007). Of mice and men: Natural kinds of emotions in the mammalian brain? A response to Panksepp and Izard. *Perspectives on Psychological Science*, *2*(3), 297–312. https://doi.org/10.1111/j.1745-6916.2007.00046.x

Barrett, L. F., Lindquist, K. A., & Gendron, M. (2007). Language as context for the perception of emotion. *Trends in Cognitive Sciences*, *11*(8), 327–332. https://doi.org/10.1016/j.tics.2007.06.003

Barrett, L. F., Ochsner, K. N., & Gross, J. J. (2007). On the automaticity of emotion. In J. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 173–217). Psychology Press.

Barrett, L. F., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology*, *74*(4), 967–984. https://doi.org/10.1037/0022-3514.74.4.967

Barrett, L. F., & Satpute, A. B. (2019). Historical pitfalls and new directions in the neuroscience of emotion. *Neuroscience Letters*, *693*, 9–18.

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, *16*(7), 419–429. https://doi.org/10.1038/nrn3950

Barrett, L. F., & Wager, T. (2006). The structure of emotion: Evidence from the neuroimaging studies. *Current Directions in Psychological Science*, *15*, 79–85. https://doi.org/10.1111/j.0963-7214.2006.00411.x

Barrett, L. F., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). The conceptual act theory: A roadmap. In L. F. Barrett & J. A. Russell (Eds.), *The psychological construction of emotion* (pp. 83–110). Guilford.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 629–654. https://doi.org/10.1037/0278-7393.11.1-4.629

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660. https://doi.org/10.1017/s0140525x99002149

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*(2), 84–91. https://doi.org/10.1016/s1364–6613(02)00029-3

Bauer, R. M. (1998). Physiologic measures of emotion. *Journal of Clinical Neurophysiology*, *15*(5), 388–396. https://doi.org/10.1097/00004691-199809000-00003

Baumgartner, M., & Gebharter, A. (2016). Constitutive relevance, mutual manipulability, and fat-handedness. *British Journal for the Philosophy of Science*, *67*(3), 731–756. https://doi.org/10.1093/bjps/axv003

Bayliss, A. P., Frischen, A., Fenske, M. J., & Tipper, S. P. (2007). Affective evaluations of objects are influenced by observed gaze direction and emotional expression. *Cognition*, *104*(3), 644–653. https://doi.org/10.1016/j.cognition.2006.07.012

Bayne, T., & Hattiangadi, A. (2013). Belief and its bedfellows. In N. Nottelman (Ed.), *New essays on belief: Constitution, content and structure* (pp. 124–144). Palgrave Macmillan. https://doi.org/10.1057/9781137026521_7

Bechara, A., Damasio, H., Damasio, A. R., & Lee, G. P. (1999). Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *Journal of Neuroscience*, *19*(13), 5473–5481. https://doi.org/10.1523/jneurosci.19-13-05473.1999

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*(5304), 1293–1295. https://doi.org/10.1126/science.275.5304.1293

Bechtel, W. (1998a). Representations and cognitive explanations: Assessing the dynamicist's challenge in cognitive science. *Cognitive Science*, *22*(3), 295–318. https://doi.org/10.1207/s15516709cog2203_2

Bechtel, W. (1998b). Dynamicists versus computationalists: Whither mechanists? *Behavioral and Brain Sciences*, *21*(5), 629. https://doi.org/10.1017/s0140525x9822173x

Bechtel, W. (2001). Representations: From neural systems to cognitive systems. In W. Bechtel, P. Mandik, J. Mundale, & R. S. Stufflebeam (Eds.), *Philosophy and the neurosciences: A reader* (pp. 332–348). Blackwell.

Bechtel, W. (2005). The challenge of characterizing operations in the mechanisms underlying behavior. *Journal of the Experimental Analysis of Behavior*, *84*(3), 313–325. https://doi.org/10.1901/jeab.2005.103-04

Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. Routledge.

Bechtel, W., & Richardson, R. C. (2010). *Discovering complexity: Decomposition and localization as strategies in scientific research*. MIT Press. https://doi.org/10.7551/mitpress/8328.001.0001

Bechtel, W., & Shagrir, O. (2015). The non-redundant contributions of Marr's three levels of analysis for explaining information-processing mechanisms. *Topics in Cognitive Science*, *7*(2), 312–322. https://doi.org/10.1111/tops.12141

Beckers, T., Krypotos, A. M., Boddez, Y., Effting, M., & Kindt, M. (2013). What's wrong with fear conditioning? *Biological Psychology*, *92*(1), 90–96. https://doi.org/10.1016/j.biopsycho.2011.12.015

Beer, J. S. (2017). Current emotion research in social neuroscience: How does emotion influence social cognition? *Emotion Review*, *9*(2), 172–180. https://doi.org/10.1177/1754073916650492

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, *74*(3), 183–200. https://doi.org/10.1037/h0024835

Bem, S., & Looren de Jong, H. (2013). *Theoretical issues in psychology: An introduction*. Sage. https://doi.org/10.4135/9781526402097

Bennett, D. S., Bendersky, M., & Lewis, M. (2002). Facial expressivity at 4 months: A context by expression analysis. *Infancy*, *3*(1), 97–113. https://doi.org/10.1207/S15327078IN0301_5

Ben-Ze'ev, A. (2000). *The subtlety of emotions*. MIT Press. https://doi.org/10.7551/mitpress/6548.001.0001

Berkman, E. T., Hutcherson, C. A., Livingston, J. L., Kahn, L. E., & Inzlicht, M. (2017). Self-control as value-based choice. *Current Directions in Psychological Science*, *26*(5), 422–428. https://doi.org/10.1177/0963721417704394

Berkowitz, L., & Harmon-Jones, E. (2004). Toward an understanding of the determinants of anger. *Emotion*, *4*(2), 107. https://doi.org/10.1037/1528-3542.4.2.107

Bermúdez, J. L. (1995). Nonconceptual content: From perceptual experience to subpersonal computational states. *Mind and Language*, *10*(4), 333–369. https://doi.org/10.1111/j.1468-0017.1995.tb00019.x

Berns, G. S., Cohen, J. D., & Mintun, M. A. (1997). Brain regions responsive to novelty in the absence of awareness. *Science*, *276*(5316), 1272–1275. https://doi.org/10.1126/science.276.5316.1272

Bindra, D. (1969). A unified interpretation of emotion and motivation. *Annals of the New York Academy of Sciences*, *159*(3), 1071–1083. https://doi.org/10.1111/j.1749-6632.1969.tb12998.x

Blackford, J. U., Buckholtz, J. W., Avery, S. N., & Zald, D. H. (2010). A unique role for the human amygdala in novelty detection. *Neuroimage*, *50*(3), 1188–1193. https://doi.org/10.1016/j.neuroimage.2009.12.083

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, *6*(3), 242–261. https://doi.org/10.1207/S15327957PSPR0603_8

Blanchard, D. C., & Blanchard, R. J. (2003). What can animal aggression research tell us about human aggression? *Hormones and Behavior*, *44*(3), 171–177. https://doi.org/10.1016/S0018–506X(03)00133-8

Blascovich, J., & Mendes, W. B. (2000). Challenge and threat appraisals: The role of affective cues. In J. P. Forgas (Ed.), *Feeling and thinking: The role of affect in social cognition* (pp. 59–82). Cambridge University Press.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*(2), 227–247. https://doi.org/10.1017/S0140525X00038188

Boddez, Y., Moors, A., Mertens, G., & De Houwer, J. (2020). Tackling fear: Beyond associative memory activation as the only determinant of fear responding. *Neuroscience & Biobehavioral Reviews*, *112*, 410–419. https://doi.org/10.1016/j.neubiorev.2020.02.009

Boiger, M., & Mesquita, B. (2012). The construction of emotion in interactions, relationships, and cultures. *Emotion Review*, *4*(3), 221–229. https://doi.org/10.1177/1754073912439765

Bolles, R. C., & Fanselow, M. S. (1980). A perceptual-defensive-recuperative model of fear and pain. *Behavioral and Brain Sciences*, *3*(2), 291–301. https://doi.org/10.1017/s0140525x0000491x

Bonanno, G., & Keltner, D. (2004). Brief Report. The coherence of emotion systems: Comparing “on-line” measures of appraisal and facial expressions, and self-report. *Cognition and Emotion*, *18*(3), 431–444. https://doi.org/10.1080/02699930341000149

Bonnie, K. E., & de Waal, F. B. M. (2004). Primate social reciprocity and the origin of gratitude. In R. A. Emmons & M. E. McCullough (Eds.), *The psychology of gratitude* (pp. 213–229). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195150100.003.0011

Bossuyt, E., Moors, A., & De Houwer, J. (2014a). Exploring the relations between regret, self-agency, and the tendency to repair using experimental methods and structural equation modeling. *Psychological Record*, *64*(4), 841–857. https://doi.org/10.1007/s40732–014-0065-4

Bossuyt, E., Moors, A., & De Houwer, J. (2014b). Unexpected and just missed: The separate influence of the appraisals of expectancy and proximity on negative emotions. *Emotion*, *14*(2), 284–300. https://doi.org/10.1037/a0034675

Bourgeois, P., & Hess, U. (2008). The impact of social context on mimicry. *Biological Psychology*, *77*(3), 343–352. https://doi.org/10.1016/j.biopsycho.2007.11.008

Bower, G. H. (1981). Mood and memory. *American Psychologist*, *36*(2), 129–148. https://doi.org/10.1037/0003-066X.36.2.129

Bower, G. H. (1991). Mood congruity of social judgments. In J. P. Forgas (Ed.), *Emotion and social judgments* (pp. 31–53). Pergamon Press. https://doi.org/10.4324/9781003058731-3

Boyd, R. (1999). Homeostasis, species, and higher taxa. In R. A. Wilson (Ed.), *Species: New interdisciplinary essays* (pp. 141–185). MIT Press. https://doi.org/10.1086/656551

Boyd, R. N. (2010). Homeostasis, higher taxa, and monophyly. *Philosophy of Science*, *77*(5), 686–701. https://doi.org/10.1086/656551

Bradley, M. M., & Lang, P. J. (2000). Measuring emotion: Behavior, feeling, and physiology. In R. D. Lane & L. Nadel (Eds.), *Cognitive neuroscience of emotion* (pp. 49–59). Oxford University Press.

Brady, M. S. (2007). Recalcitrant emotions and visual illusions. *American Philosophical Quarterly*, *44*(3), 273–284. www.jstor.org/stable/20464376

Brandstädter, J., & Greve, W. (1994). The aging self: Stabilizing and protective processes. *Developmental Review*, *14*(1), 52–80. https://doi.org/10.1006/drev.1994.1003

Brandstädter, J., & Rothermund, K. (2002). The life-course dynamics of goal pursuit and goal adjustment: A two-process framework. *Developmental Review*, *22*(1), 117–150. https://doi.org/10.1006/drev.2001.0539

Brentano, F. C. (1874). *Psychologie vom empirischen Standpunkt* [Psychology from an empirical standpoint]. Duncker & Humblot. https://doi.org/10.28937/978-3-7873-2608-2

Brigandt, I. (2003). Species pluralism does not imply species eliminativism. *Philosophy of Science*, *70*(5), 1305–1316. https://doi.org/10.1086/377409

Broad, C. D. (1954). Emotion and sentiment. *Journal of Aesthetics and Art Criticism*, *13*(2), 203–214. https://doi.org/10.2307/425913

Broekens, J., Jacobs, E., & Jonker, C. M. (2015). A reinforcement learning model of joy, distress, hope and fear. *Connection Science*, *27*(3), 215–233. https://doi.org/10.1080/09540091.2015.1031081

Brosch, T., & Sander, D. (2013). Comment: The appraising brain: Towards a neuro-cognitive model of appraisal processes in emotion. *Emotion Review*, *5*(2), 163–168. https://doi.org/10.1177/1754073912468298

Brosch, T., Scherer, K. R., Grandjean, D. M., & Sander, D. (2013). The impact of emotion on perception, attention, memory, and decision-making. *Swiss Medical Weekly*, *143*, w13786. https://doi.org/10.4414/smw.2013.13786

Buabang, E. K., Boddez, Y., Wolf, O. T., & Moors, A. (2021). The role of goal-directed and habitual processes in food consumption under stress after outcome devaluation with taste aversion. *Behavioral Neuroscience*. Accepted Stage-1 preregistered report.

Buabang, E. K., Köster, M., Boddez, Y., Van Dessel, P., De Houwer, J., & Moors, A. (2021). *A goal-directed account of action slips: The reliance on old contingencies*. PsyArXiv. https://doi.org/10.31234/osf.io/y6vbg

Buck, R. (1999). The biological affects: A typology. *Psychological Review*, *106*(2), 301–336. https://doi.org/10.1037/0033-295x.106.2.301

Bunge, M. (2004). How does it work? The search for explanatory mechanisms. *Philosophy of the Social Sciences*, *34*(2), 182–210. https://doi.org/10.1177/0048393103262550

Bushman, B. J. (2002). Does venting anger feed or extinguish the flame? Catharsis, rumination, distraction, anger, and aggressive responding. *Personality and Social Psychology Bulletin*, *28*(6), 724–731. https://doi.org/10.1177/0146167202289002

Butler, E. A. (2011). Temporal interpersonal emotion systems: The "TIES" that form relationships. *Personality and Social Psychology Review*, *15*(4), 367–393. https://doi.org/10.1177/1088868311411164

Byrne, A. (2001). Intentionalism defended. *Philosophical Review*, *110*(2), 199–240. https://doi.org/10.1215/00318108-110-2-199

Cabanac, M. (2002). What is emotion? *Behavioural Processes*, *60*(2), 69–83. https://doi.org/10.1016/s0376–6357(02)00078-5

Cacioppo, J. T., Berntson, G. G., & Klein, D. J. (1992). What is an emotion? The role of somatovisceral afference, with special emphasis on somatovisceral "illusions." In M. S. Clark (Ed.), *Emotion and social behavior* (pp. 63–98). Sage.

Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., & Ito, T. A. (2000). The psychophysiology of emotion. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 173–191). Guilford.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology*, *76*(5), 839–855. https://doi.org/10.1037/0022-3514.76.5.839

Cacioppo, J. T., Priester, J. R., & Berntson, G. G. (1993). Rudimentary determinants of attitudes: II. Arm flexion and extension have differential effects on attitudes. *Journal of Personality and Social Psychology*, *65*(1), 5–17. https://doi.org/10.1037/0022-3514.65.1.5

Calder, A. J., Lawrence, A. D., & Young, A. W. (2001). Neuropsychology of fear and loathing. *Nature Reviews Neuroscience*, *2*, 352–363. https://doi.org/10.1038/35072584

Camerer, C. F., & Loewenstein, G. (2004). Behavioral economics: Past, present, future. In C. F. Camerer, G. Loewenstein, & M. Rabin (Eds.), *Advances in behavioral economics* (pp. 3–51). Princeton University Press. https://doi.org/10.1515/9781400829118-004

Campos, J. J., Frankel, C. B., & Camras, L. (2004). On the nature of emotion regulation. *Child Development*, *75*(2), 377–394. https://doi.org/10.1111/j.1467-8624.2004.00681.x

Camras, L. A. (1992). Expressive development and basic emotions. *Cognition and Emotion*, *6*(3–4), 269–283. https://doi-org/10.1080/02699939208411072

Camras, L. A., & Fatani, S. S. (2008). The development of facial expressions: Current perspectives on infant emotions. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (pp. 291–304). Guilford.

Cannon, W. B. (1927). The James-Lange theory of emotions: A critical examination and an alternative theory. *American Journal of Psychology*, *39*(1), 106–124. https://doi.org/10.2307/1415404

Cannon, W. B., Lewis, J. T., & Britton, S. W. (1927). The dispensability of the sympathetic division of the autonomic system. *The New England Journal of Medicine*, *197*(13), 514–515. https://doi.org/10.1056/nejm192709291971303

Carnap, R. (1950). *Logical foundations of probability*. University of Chicago Press.

Carver, C. S. (2003). Pleasure as a sign you can attend to something else: Placing positive feelings within a general model of affect. *Cognition and Emotion*, *17*(2), 241–261. https://doi.org/10.1080/02699930302294

Carver, C. S. (2004). Self-regulation of action and affect. In R. F. Baumeister & K. D. Vohs (Eds.), *Handbook of self-regulation: Research, theory, and applications* (pp. 13–39). Guilford.

Carver, C. S. (2006). Approach, avoidance, and the self-regulation of affect and action. *Motivation and Emotion*, *30*(2), 105–110. https://doi.org/10.1007/s11031–006-9044-7

Carver, C. S., & Scheier, M. F. (1985). A control-systems approach to the self-regulation of action. In J. Kuhl & J. Beckmann (Eds.), *Action control: From cognition to behavior* (pp. 237–265). Springer-Verlag. https://doi.org/10.1007/978-3-642-69746-3_11

Carver, C. S., & Scheier, M. F. (1990). Origins and functions of positive and negative affect: A control-process view. *Psychological Review*, *97*(1), 19–35. https://doi.org/10.1037/0033-295X.97.1.19

Carver, C. S., & Scheier, M. F. (2000). Autonomy and self-regulation. *Psychological Inquiry*, *11*(4), 284–291. www.jstor.org/stable/1449622

Carver, C. S., & Scheier, M. F. (2002). Control processes and self-organization as complementary principles underlying behaviour. *Personality and Social Psychology Review*, *6*(4), 304–315. https://doi.org/10.1207/s15327957pspr0604_05

Carver, C. S., & Scheier, M. F. (2008). Feedback processes in the simultaneous regulation of action and affect. In J. Y. Shah & W. L. Gardner (Eds.), *Handbook of motivation science* (pp. 308–324). Guilford.

Carver, C. S., & Scheier, M. F. (2013). Goals and emotion. In E. R. Watkins, M. D. Robinson, & E. Harmon-Jones (Eds.), *Handbook of cognition and emotion* (pp. 176–194). Guilford.

Castelfranchi, C., & Miceli, M. (2009). The cognitive-motivational compound of emotional experience. *Emotion Review*, *1*(3), 223–231. https://doi.org/10.1177/1754073909103590

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212–252). Guilford.

Charland, L. C. (1995). Feeling and representing: Computational theory and the modularity of affect. *Synthese*, *105*(3), 273–301. https://doi.org/10.1007/bf01063560

Charland, L. C. (1997). Reconciling cognitive and perceptual theories of emotion: A representational proposal. *Philosophy of Science*, *64*(4), 555–579. https://doi.org/10.1086/392572

Charland, L. C. (2005a). Emotion experience and the indeterminacy of valence. In L. F. Barrett, P. M. Niedenthal, & P. Winkielman (Eds.), *Emotion and consciousness* (pp. 231–254). Guilford.

Charland, L. C. (2005b). The heat of emotion: Valence and the demarcation problem. *Journal of Consciousness Studies*, *12*(8–9), 82–102.

Charland, L. C. (2010). Reinstating the passions: Arguments from the history of psychopathology. In P. Goldie (Ed.), *The Oxford handbook of philosophy of emotion* (pp. 237–259). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199235018.003.0011

Chater, N. (2003). How much can we learn from double dissociations? *Cortex*, *39*(1), 167–169. https://doi.org/10.1016/s0010–9452(08)70093-5

Cherek, D. (1981). Effects of smoking different doses of nicotine on human aggressive behavior. *Psychopharmacology*, *75*, 339–349. https://doi.org/10.1007/bf00435849

Christie, I. C., & Friedman, B. H. (2004). Autonomic specificity of discrete emotion and dimensions of affective space: A multivariate approach. *International Journal of Psychophysiology*, *51*(2), 143–153. https://doi.org/10.1016/j.ijpsycho.2003.08.002

Churchland, P. M., & Churchland, P. S. (1992). Intertheoretic reduction: A neuroscientist's field guide. In Y. Christen & P. S. Churchland (Eds.), *Neurophilosophy and Alzheimer's disease* (pp. 18–29). Springer-Verlag. https://doi.org/10.1007/978-3-642-46759-2_3

Chwalisz, K., Diener, E., & Gallagher, D. (1988). Autonomic arousal feedback and emotional experience: Evidence from the spinal cord injured. *Journal of Personality and Social Psychology*, *54*(5), 820–828. https://doi.org/10.1037/0022-3514.54.5.820

Civai, C., Corradi-Dell'Acqua, C., Gamer, M., & Rumiati, R. I. (2010). Are irrational reactions to unfairness truly emotionally-driven? Dissociated behavioural and emotional responses in the Ultimatum Game task. *Cognition*, *114*(1), 89–95. https://doi.org/10.1016/j.cognition.2009.09.001

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/s0140525x12000477

Clark, M. S., & Isen, A. M. (1982). Towards understanding the relationship between feeling states and social behavior. In A. H. Hastorf & A. M. Isen (Eds.), *Cognitive social psychology* (pp. 73–108). Elsevier. https://doi.org/10.2307/3033676

Clark-Polner, E., Johnson, T. D., & Barrett, L. F. (2017). Multivoxel pattern analysis does not provide evidence to support the existence of basic emotions. *Cerebral Cortex*, *27*(3), 1944–1948. https://doi.org/10.1093/cercor/bhw028

Clore, G. L. (1994). Why emotions are never unconscious. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotions: Fundamental questions* (pp. 285–290). Oxford University Press.
Clore, G. L., & Centerbar, D. B. (2004). Analyzing anger: How to make people mad. *Emotion*, *4*(2), 139–144. https://doi.org/10.1037/1528-3542.4.2.139
Clore, G., & Ketelaar, T. (1997). Minding our emotions: On the role of automatic, unconscious affect. In R. S. Wyer, Jr. (Ed.), *Advances in social cognition* (Vol. 10, pp. 105–120). Erlbaum.
Clore, G. L., & Ortony, A. (2000). Cognition in emotion: Always, sometimes, or never? In R. D. Lane & L. Nadel (Eds.), *Cognitive neuroscience of emotion* (pp. 24–61). Oxford University Press.
Clore, G. L., & Ortony, A. (2008). Appraisal theories: How cognition shapes affect into emotion. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 628–642). Guilford.
Clore, G. L., & Ortony, A. (2013). Psychological construction in the OCC model of emotion. *Emotion Review*, *5*(4), 335–343. https://doi.org/10.1177/1754073913489751
Clore, G. L., Ortony, A., & Foss, M. A. (1987). The psychological foundations of the affective lexicon. *Journal of Personality and Social Psychology*, *53*(4), 751–766. https://doi.org/10.1037/0022-3514.53.4.751
Clore, G. L., Storbeck, J., Robinson, M. D., & Centerbar, D. (2005). Seven sins of research on unconscious affect. In L. F. Barrett, P. Niedenthal, & P. Winkielman (Eds.), *Emotion: Conscious and unconscious* (pp. 384–408). Guilford.
Coan, J. A. (2010). Emergent ghosts of the emotion machine. *Emotion Review*, *2* (3), 274–285. https://doi.org/10.1177/1754073910361978
Coan, J. A., & Gonzalez, M. Z. (2015). Emotions as emergent variables. In L. F. Barrett & J. A. Russell (Eds.), *The psychological construction of emotion* (pp. 209–225). Guilford.
Cochrane, T. (2019). *The emotional mind: A control theory of affective states*. Cambridge University Press. https://doi.org/10.1017/9781108579056
Coles, N. A., Larsen, J. T., & Lench, H. C. (2019). A meta-analysis of the facial feedback literature: Effects of facial feedback on emotional experience are small and variable. *Psychological Bulletin*, *145*(6), 610–651. https://doi.org/10.1037/bul0000194
Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295x.82.6.407
Colombetti, G. (2003). Complexity as a new framework for emotion theories. *Logic and Philosophy of Science*, *1*(1), 1–16.
Colombetti, G. (2005). Appraising valence. *Journal of Consciousness Studies*, *12*(8–9), 103–126.
Colombetti, G. (2007). Enactive appraisal. *Phenomenology and the Cognitive Sciences*, *6*(4), 527–546. https://doi.org/10.1007/s11097–007-9077-8
Colombetti, G. (2009). From affect programs to dynamical discrete emotions. *Philosophical Psychology*, *22*(4), 407–425. https://doi.org/10.1080/09515080903153600

Colombetti, G. (2014). *The feeling body: Affective science meets the enactive mind*. MIT Press. https://doi.org/10.7551/mitpress/9780262019958.001.0001

Colombetti, G., & Krueger, J. (2015). Scaffoldings of the affective mind. *Philosophical Psychology*, *28*(8), 1157–1176. https://doi.org/10.1080/09515089.2014.976334

Colombetti, G., & Thompson, E. (2007). The feeling body: Towards an enactive approach to emotion. In W. F. Overton, U. Müller, & J. L. Newman (Eds.), *Developmental perspectives on embodiment and consciousness* (pp. 61–84). Psychology Press. https://doi.org/10.4324/9780203809778

Colwill, R. M., & Rescorla, R. A. (1985). Instrumental responding remains sensitive to reinforcer devaluation after extensive training. *Journal of Experimental Psychology: Animal Behavior Processes*, *11*(4), 520–536. https://doi.org/10.1037/0097-7403.11.4.520

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The quad model of implicit task performance. *Journal of Personality and Social Psychology*, *89*(4), 469–487. https://doi.org/10.1037/0022-3514.89.4.469

Cordaro, D. T., Keltner, D., Tshering, S., Wangchuk, D., & Flynn, L. (2016). The voice conveys emotion in ten globalized cultures and one remote village in Bhutan. *Emotion*, *16*(1), 117–128. https://doi.org/10.1037/emo0000100

Cova, F., & Deonna, J. A. (2014). Being moved. *Philosophical Studies*, *169*(3), 447–466. https://doi.org/10.1007/s11098–013-0192-9

Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Clarendon Press. https://doi.org/10.1017/s0033291708002973

Craver, C. F. (2013). Functions and mechanisms: A perspectivalist view. In P. Humean (Ed.), *Functions: Selection and mechanisms* (pp. 133–158). Springer. https://doi.org/10.1007/978-94-007-5304-4_8

Craver, C. F. (2015). Levels. In T. Metzinger & J. M. Windt (Eds.), *Open MIND* (8th ed., pp. 1–26). MIND Group. https://doi.org/10.15502/9783958570498

Craver, C. F., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology & Philosophy*, *22*(4), 547–563. https://doi.org/10.1007/s10539–006-9028-8

Craver, C. F., & Bechtel, W. (2013). Interlevel causation. In D. Werner, O. Wolkenhauer, K. H. Cho, & H. Yokota (Eds.), *Encyclopedia of systems biology* (pp. 1044–1047). Springer. https://doi.org/10.1007/978-1-4419-9863-7_69

Craver, C. F., & Kaplan, D. M. (2020). Are more details better? On the norms of completeness for mechanistic explanations. *British Journal for the Philosophy of Science*, *71*(1), 287–319. https://doi.org/10.1093/bjps/axy015

Craver, C. F., & Tabery, J. (2019, Summer). Mechanisms in science. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. https://plato.stanford.edu/archives/sum2019/entries/science-mechanisms/

Crisp, T. M., & Warfield, T. A. (2001). Kim's master argument. *Noûs*, *35*(2), 304–316. www.jstor.org/stable/2672009

Crivelli, C., & Fridlund, A. J. (2018). Facial displays are tools for social influence. *Trends in Cognitive Sciences*, *22*(5), 388–399. https://doi.org/10.1016/j.tics.2018.02.006

Crivelli, C., & Fridlund, A. J. (2019). Inside-out: From basic emotions theory to the behavioral ecology view. *Journal of Nonverbal Behavior*, *43*(2), 161–194. https://doi.org/10.1007/s10919–019-00294-2

Crivelli, C., & Gendron, M. (2017). Facial expressions and emotions in indigenous societies. In J. M. Fernández-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp. 497–515). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190613501.003.0026

Crivelli, C., Jarillo, S., & Fridlund, A. J. (2016). A multidisciplinary approach to research in small-scale societies: Studying emotions and facial expressions in the field. *Frontiers in Psychology*, *7*, 1073. https://doi.org/10.3389/fpsyg.2016.01073

Crivelli, C., Jarillo, S., Russell, J. A., & Fernández-Dols, J. M. (2016). Reading emotions from faces in two indigenous societies. *Journal of Experimental Psychology: General*, *145*(7), 830–843. https://doi.org/10.1037/xge0000172

Crivelli, C., Russell, J. A., Jarillo, S., & Fernández-Dols, J. M. (2016). The fear gasping face as a threat display in a Melanesian society. *Proceedings of the National Academy of Sciences*, *113*(4), 12403–12407. https://doi.org/10.1073/pnas.1611622113

Crivelli, C., Russell, J. A., Jarillo, S., & Fernández-Dols, J. M. (2017). Recognizing spontaneous facial expressions of emotion in a small-scale society of Papua New Guinea. *Emotion*, *17*(2), 337–347. https://doi.org/10.1037/emo0000236

Csikszentmihalyi, M., & Larson, R. (2014). Validity and reliability of the experience-sampling method. In M. Csikszentmihalyi (Ed.), *Flow and the foundations of positive psychology* (pp. 35–54). Springer. https://doi.org/10.1007/978-94-017-9088-8_3

Cunningham, W. A., & Brosch, T. (2012). Motivational salience: Amygdala tuning from traits, needs, values, and goals. *Current Directions in Psychological Science*, *21*(1), 54–59. https://doi.org/10.1177/0963721411430832

Cunningham, W. A., Dunfield, K. A., & Stillman, P. E. (2013). Emotional states from affective dynamics. *Emotion Review*, *5*(4), 344–355. https://doi.org/10.1177/1754073913489749

Cunningham, W. A., & Zelazo, P. D. (2007). Attitudes and evaluations: A social cognitive neuroscience perspective. *Trends in Cognitive Sciences*, *11*(3), 97–104. https://doi.org/10.1016/j.tics.2006.12.005

Dael, N., Goudbeek, M., & Scherer K. R. (2013). Perceived gesture dynamics in nonverbal expression of emotion. *Perception*, *42*(6), 642–657. https://doi.org/10.1068/p7364

Dael, N., Mortillaro, M., & Scherer, K. R. (2012a). Emotion expression in body action and posture. *Emotion*, *12*(5), 1085–1101. https://doi.org/10.1037/a0025737

Dael, N., Mortillaro, M., & Scherer, K. R. (2012b). The body action and posture coding system (BAP): Development and reliability. *Journal of Nonverbal Behavior*, *36*(2), 97–121. https://doi.org/10.1007/s10919–012-0130-0

Damasio, A. R. (1994). *Descartes' error: Emotion, rationality and the human brain*. Putnam.

Damasio, A. R. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. Houghton Mifflin Harcourt.

Damasio, A. R. (2004). Emotions and feelings. In A. S. R. Manstead, N. H. Frijda, & A. H. Fischer (Eds.), *Feelings and emotions: The Amsterdam symposium* (pp. 49–57). Cambridge University Press. https://doi.org/10.1017/cbo9780511806582.004

Dan-Glauser, E. S., & Gross, J. J. (2013). Emotion regulation and emotion coherence: Evidence for strategy-specific effects. *Emotion*, *13*(5), 832–842. https://doi.org/10.1037/a0032672

Darlow, A. L., & Sloman, S. A. (2010). Two systems of reasoning: Architecture and relation to emotion. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(3), 382–392. https://doi.org/10.1002/wcs.34

D'Arms, J., & Jacobson, D. (2000). The moralistic fallacy: On the 'appropriateness' of emotions. *Philosophy and Phenomenological Research*, *61*(1), 65–90. https://doi.org/10.2307/2653403

Darrow, S. M., & Follette, W. C. (2014). A behavior analytic interpretation of alexithymia. *Journal of Contextual Behavioral Science*, *3*(2), 98–108. https://doi.org/10.1016/j.jcbs.2014.03.002

Darwin, C. (1872). *The expression of the emotions in man and animals by Charles Darwin*. John Murray. https://doi.org/10.1037/10001-000

Dashiell, J. F. (1928). Are there any native emotions? *Psychological Review*, *35*, 319–327. https://doi.org/10.1037/h0075969

Davidson, D. (1985a). A new basis for decision theory. *Theory and Decision*, *18*(1), 87–98. https://doi.org/10.1007/bf00134079

Davidson, D. (1985b). Incoherence and irrationality. *Dialectica*, *39*(4), 345–354. https://doi.org/10.1111/j.1746-8361.1985.tb01603.x

Davis, J. I., Senghas, A., Brandt, F., & Ochsner, K. N. (2010). The effects of BOTOX injections on emotional experience. *Emotion*, *10*(3), 433–440. https://doi.org/10.1037/a0018690

Davis, M. (1992). The role of the amygdala in fear and anxiety. *Annual Review of Neuroscience*, *15*, 353–375. https://doi.org/10.1146/annurev.ne.15.030192.002033

De Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience*, *7*(3), 242–249. https://doi.org/10.1038/nrn1872

De Houwer, J. (2014). Why a propositional single-process model of associative learning deserves to be defended. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 530–541). Guilford.

De Houwer, J., Barnes-Holmes, D., & Moors, A. (2013). What is learning? On the nature and merits of a functional definition of learning. *Psychonomic Bulletin & Review*, *20*(4), 631–642. https://doi.org/10.3758/s13423–013-0386-3

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2013). A functional-cognitive framework for attitude research. *European Review of Social Psychology*, *24*(1), 252–287. https://doi.org/10.1080/10463283.2014.892320

De Houwer, J., & Moors, A. (2010). Implicit measures: Similarities and differences. In B. Gawronski & K. B. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 176–193). Guilford.

Delancey, C. (2005). Lewis's DS approach is a tool, not a theory. *Behavioral and Brain Sciences*, *28*(2), 201. https://doi.org/10.1017/s0140525x05290040

De Leersnyder, J., Mesquita, B., Kim, H., Eom, K., & Choi, H. (2014). Emotional fit with culture: A predictor of individual differences in relational well-being. *Emotion*, *14*(2), 241–245. https://doi.org/10.1037/a0035296

de Montigny, C. (1989). Cholecystokinin tetrapeptide induces panic-like attacks in healthy volunteers: Preliminary findings. *Archives of General Psychiatry*, *46*(6), 511–517. https://doi.org/10.1001/archpsyc.1989.01810060031006

Deonna, J. A. (2006). Emotion, perception and perspective. *Dialectica*, *60*(1), 29–46. https://doi.org/10.1111/j.1746-8361.2005.01031.x

Deonna, J. A., & Scherer, K. R. (2010). The case of the disappearing intentional object: Constraints on a definition of emotion. *Emotion Review*, *2*(1), 44–52. https://doi.org/10.1177/1754073909345544

Deonna, J. A., Tappolet, C., & Teroni, F. (2015). Emotion, philosophical issues about. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(3), 193–207. https://doi.org/10.1002/wcs.1342

Deonna, J. A., & Teroni, F. (2012). *The emotions: A philosophical introduction*. Routledge. https://doi.org/10.4324/9780203721742

Deonna, J. A., & Teroni, F. (2014). In what sense are emotions evaluations? In S. Roeser & C. Todd (Eds.), *Emotion and value* (pp. 15–31). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199686094.003.0002

Deonna, J. A., & Teroni, F. (2015). Emotions as attitudes. *Dialectica*, *69*(3), 293–311. https://doi.org/10.1111/1746-8361.12116

Deonna, J. A., & Teroni, F. (2017). Getting bodily feelings into emotional experience in the right way. *Emotion Review*, *9*(1), 55–63. https://doi.org/10.1177/1754073916639666

Deonna, J. A., & Teroni, F. (2020). Emotional experience: Affective consciousness and its role in emotion theory. In U. Kriegel (Ed.), *Oxford handbook of the philosophy of consciousness* (pp. 102–123). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198749677.013.5

De Rivera, J., & Grinkis, C. (1986). Emotions as social relationships. *Motivation and Emotion*, *10*, 351–369. https://doi.org/10.1007/bf00992109

Descartes, R. (1989). *The passions of the soul* (S. Voss, Trans.). Hackett. (Original work published 1649)

de Sousa, R. (1987). *The rationality of emotion*. MIT Press. https://doi.org/10.7551/mitpress/5760.001.0001

de Sousa, R. (2003, Spring). Emotion. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. https://plato.stanford.edu/archives/win2017/entries/emotion/

de Sousa, R. (2018). Les émotions contemplatives et l'objectivité des valeurs [Contemplative emotion and the objectivity of values]. *Philosophiques*, *45*(2), 499–505. https://doi.org/10.7202/1055276ar

Dewey, J. (1894). The theory of emotion. I: Emotional attitudes. *Psychological Review*, *1*(6), 553–569. https://doi.org/10.1037/h0069054

Dewey, J. (1895). The theory of emotion. II: The significance of emotions. *Psychological Review*, *2*(1), 13–32. https://doi.org/10.1037/h0070927 de Wit, S., Kindt, M., Knot, S. L., Verhoeven, A. A. C., Robbins, T. W., Gasull-Camos, J., Evans, M., Mirza, H., & Gillan, C. M. (2018). Shifting the balance between goals and habits: Five failures in experimental habit induction. *Journal of Experimental Psychology: General*, *147*(7), 1043–1065. https://doi.org/10.1037/xge0000402

Dickinson, A. (1985). Actions and habits: The development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *308*(1135), 67–78. https://doi.org/10.1098/rstb.1985.0010

Dickinson, A., & Balleine, B. (1994). Motivational control of goal-directed action. *Animal Learning & Behavior*, *22*(1), 1–18. https://doi.org/10.3758/bf03199951

Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin*, *125*(2), 276–302. https://doi.org/10.1037/0033-2909.125.2.276

Dik, G., & Aarts, H. (2007). Behavioral cues to others' motivation and goal pursuits: The perception of effort facilitates goal inference and contagion. *Journal of Experimental Social Psychology*, *43*(5), 727–737. https://doi.org/10.1016/j.jesp.2006.09.002

Dixon, T. (2003). *From passions to emotions: The creation of a secular psychological category*. Cambridge University Press. https://doi.org/10.1017/cbo9780511490514

Dixon, T. (2012). "Emotion": The history of a keyword in crisis. *Emotion Review*, *4*(4), 338–344. http://doi.org/10.1177/1754073912445814

Dokic, J., & Lemaire, S. (2013). Are emotions perceptions of value? *Canadian Journal of Philosophy*, *43*(2), 227–247. https://doi.org/10.1080/00455091.2013.826057

Dokic, J., & Lemaire, S. (2015). Are emotions evaluative modes? *Dialectica*, *69*(3), 271–292. https://doi.org/10.1111/1746-8361.12117

Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science*, *298*(5596), 1191–1194. https://doi.org/10.1126/science.1076358

Donnerstein, E., & Wilson, D. W. (1976). Effects of noise and perceived control on ongoing and subsequent aggressive behavior. *Journal of Personality and Social Psychology*, *34*(5), 774–781. https://doi.org/10.1037/0022-3514.34.5.774

Döring, S. A. (2007). Seeing what to do: Affective perception and rational motivation. *Dialectica*, *61*(3), 363–394. https://doi.org/10.1111/j.1746-8361.2007.01105.x

Döring, S. A., & Lutz, A. (2015). Beyond perceptualism: Introduction to the special issue. *Dialectica*, *69*(3), 259–270. https://doi.org/10.1111/1746-8361.12106

Douven, I. (2017, Summer). Abduction. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. https://plato.stanford.edu/archives/sum2017/entries/abduction/

Dretske, F. I. (1995). *Naturalizing the mind*. MIT Press. https://doi.org/10.7551/mitpress/4872.001.0001

Dub, R. (2022). Emotional experience: Functional or dispensable? In A. Scarantino (Ed.), *Routledge handbook of emotion theory*. Routledge.

Duchenne, B. (1990). *The mechanism of human facial expression or an electro-physiological analysis of the expression of the emotions* (A. Cuthbertson, Trans.).

Cambridge University Press. (Original work published 1862). https://doi.org/10.1017/cbo9780511752841.024

Duckworth, K. L., Bargh, J. A., Garcia, M., & Chaiken, S. (2002). The automatic evaluation of novel stimuli. *Psychological Science*, *13*(6), 513–519. https://doi.org/10.1111/1467-9280.00490

Duclos, S., Laird, J., Schneider, E., Sexter, M., Stern, L., & Van Lighten, O. (1989). Emotion-specific effects of facial expressions and postures on emotional experience. *Journal of Personality and Social Psychology*, *57*(1), 100–108. https://doi.org/10.1037/0022-3514.57.1.100

Duffy, E. (1934). Emotion: An example of the need for reorientation in psychology. *Psychological Review*, *41*(2), 184–198. https://doi.org/10.1037/h0074603

Duffy, E. (1941a). An explanation of "emotional" phenomena without the use of the concept "emotion". *Journal of General Psychology*, *25*(2), 283–293. https://doi.org/10.1080/00221309.1941.10544400

Duffy, E. (1941b). The conceptual categories of psychology: A suggestion for revision. *Psychological Review*, *48*(3), 177–203. https://doi.org/10.1037/h0059035

Dunlap, K. (1932). Are emotions teleological constructs? *American Journal of Psychology*, *44*(3), 572–576. https://doi.org/10.2307/1415359

Durán, J. I., Reisenzein, R., & Fernández-Dols, J. M. (2017). Coherence between emotions and facial expressions. In J. M. Fernández-Dols & J. Russell (Eds.), *The science of facial expression* (pp. 107–129). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190613501.003.0007

Dutton, D. G., & Aron, A. P. (1974). Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of Personality and Social Psychology*, *30*(4), 510–517. https://doi.org/10.1037/h0037031

Eder, A. B. (2011). Control of impulsive emotional behaviour through implementation intentions. *Cognition and Emotion*, *25*(3), 478–489. https://doi.org/10.1080/02699931.2010.527493

Eder, A. B. (2017). From boxology to scientific theories: On the emerging field of emotional action sciences. *Emotion Review*, *9*(4), 343–345. https://doi.org/10.1177/1754073916687827

Eder, A. B., Leuthold, H., Rothermund, K., & Schweinberger, S. R. (2012). Automatic response activation in sequential affective priming: An ERP study. *Social Cognitive and Affective Neuroscience*, *7*(4), 436–445. https://doi.org/10.1093/scan/nsr033

Eickers, G., Loaiza, J. R., & Prinz, J. J. (2017). Embodiment, context-sensitivity, and discrete emotions: A response to Moors. *Psychological Inquiry*, *28*(1), 31–38. https://doi.org/10.1080/1047840x.2017.1255492

Eilan, N. (1992). The a priori and the empirical in theories of emotion: Smedslund's "conceptual analysis" of emotion. *Cognition and Emotion*, *6*(6), 457–466. https://doi.org/10.1080/02699939208409698

Eilan, N. (1998). Perceptual intentionality. Attention and consciousness. In A. O'Hear (Ed.), *Current issues in philosophy of mind* (pp. 181–202). Cambridge University Press. https://doi.org/10.1017/s1358246100004367

Ekman, P. (1971). Universals and cultural differences in facial expressions of emotion. In J. K. Cole (Ed.), *Nebraska symposium on motivation* (4th ed., pp. 207–283). University of Nebraska Press.

Ekman, P. (1972). *Emotions in the human face*. Pergamon Press. https://doi.org/10.1016/c2013–0-02458-9

Ekman, P. (1973). Cross-cultural studies of facial expression. In P. Ekman (Ed.), *Darwin and facial expression: A century of research in review* (pp. 169–222). Academic Press.

Ekman, P. (1984). Expression and the nature of emotion. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 319–344). Erlbaum. https://doi.org/10.4324/9781315798806

Ekman, P. (1992a). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Ekman, P. (1992b). Are there basic emotions? *Psychological Review*, *99*(3), 550–553. https://doi.org/10.1037/0033-295X.99.3.550

Ekman, P. (1992c). Facial expressions of emotion: New findings, new questions. *Psychological Science*, *3*(1), 34–38. https://doi.org/10.1111/j.1467-9280.1992.tb00253.x

Ekman, P. (1994). All emotions are basic. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 15–19). Oxford University Press.

Ekman, P. (1997). Should we call it expression or communication? *Innovation: The European Journal of Social Science Research*, *10*(4), 333–344. https://doi.org/10.1080/13511610.1997.9968538

Ekman, P. (1999a). Basic emotions. In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 45–60). Wiley. https://doi.org/10.1002/0470013494.ch3

Ekman, P. (1999b). Facial expressions. In T. Dalgleish & T. Power (Eds.), *Handbook of cognition and emotion* (pp. 301–320). Wiley. https://doi.org/10.1002/0470013494.ch16

Ekman, P. (2003). *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. Henry Holt.

Ekman, P. (2006). Darwin, deception, and facial expression. *Annals of the New York Academy of Sciences*, *1000*(1), 205–221. https://doi.org/10.1196/annals.1280.010

Ekman, P. (2016). What scientists who study emotion agree about. *Perspectives on Psychological Science*, *11*(1), 31–34. https://doi.org/10.1177/1745691615596992

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic? *Emotion Review*, *3*(4), 364–370. https://doi.org/10.1177/1754073911410740

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry*, *32*(1), 88–106. https://doi.org/10.1080/00332747.1969.11023575

Ekman, P., & Friesen, W. V. (1978). *The Facial Action Coding System (FACS): A technique for the measurement of facial movement*. Consulting Psychologists' Press. https://doi.org/10.1080/00332747.1969.11023575

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior*, *6*(4), 238–252. https://doi.org/10.1007/BF00987191

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, *221*(4616), 1208–1210. https://doi.org/10.1126/science.6612338

Ekman, P., Roper, G., & Hager, J. C. (1980). Deliberate facial movement. *Child Development*, *51*(3), 886–891. https://doi.org/10.2307/1129478

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, *164*(3875), 86–88. https://doi.org/10.1126/science.164.3875.86

Elfenbein, H. A. (2014). The many faces of emotional contagion: An affective process theory of affective linkage. *Organizational Psychology Review*, *4*(4), 326–362. https://doi.org/10.1177/2041386614542889

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*(2), 203–235. https://doi.org/10.1037/0033-2909.128.2.203

Elfenbein, H. A., Beaupré, M., Lévesque, M., & Hess, U. (2007). Toward a dialect theory: Cultural differences in the expression and recognition of posed facial expressions. *Emotion*, *7*(1), 131–146. https://doi.org/10.1037/1528-3542.7.1.131

Ellis, R., & Tucker, M. (2000). Micro-affordance: The potentiation of components of action by seen objects. *British Journal of Psychology*, *91*(4), 451–471. https://doi.org/10.1348/000712600161934

Ellsworth, P. C. (1991). Some implications of cognitive appraisal theories of emotion. In K. T. Strongman (Ed.), *International review of studies on emotion* (pp. 143–161). Wiley.

Ellsworth, P. C. (1994). William James and emotion: Is a century of fame worth a century of misunderstanding? *Psychological Review*, *101*(2), 222–229. https://doi.org/10.1037/0033-295X.101.2.222

Ellsworth, P. C. (2006, August). *Are appraisals causes or constituents of emotion?* [Conference presentation]. International Society for Research on Emotion (ISRE) Conference, Atlanta, GA, USA.

Ellsworth, P. C. (2013). Appraisal theory: Old and new questions. *Emotion Review*, *5*(2), 125–131. https://doi.org/10.1177/1754073912463617

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 572–595). Oxford University Press.

Elqayam, S., & Evans, J. S. B. T. (2011). Subtracting "ought" from "is": Descriptivism versus normativism in the study of human thinking. *Behavioral and Brain Sciences*, *34*(5), 233–290. https://doi.org/10.1017/s0140525x1100001x

Elster, J. (2010). Emotional choice and rational choice. In P. Goldie (Ed.), *Oxford handbook of the philosophy of emotion* (pp. 263–281). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199235018.003.0012

Evans, D. (2008). Not so new: Overblown claims for "new" approaches to emotion. In M. Ghallab, C.D. Spyropoulos, N. Fakotakis, & N., Avouris (Eds.), *Frontiers in Artifical Intelligence and Applications: Vol. 178. ECAI 2008* (pp. 813–814). IOS Press. https://doi.org/10.3233/978-1-58603-891-5-813

Evans, J. S. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), 454–459. https://doi.org/10.1016/j.tics.2003.08.012

Evers, C., Hopp, H., Gross, J. J., Fischer, A. H., Manstead, A. S., & Mauss, I. B. (2014). Emotion response coherence: A dual-process perspective. *Biological Psychology*, *98*, 43–49. https://doi.org/10.1016/j.biopsycho.2013.11.003

Fadiga, L., Fogassi, L., Gallese, G., & Rizzolatti, G. (2000) Visuomotor neurons: Ambiguity of the discharge or "motor" perception? *International Journal of Psychophysiology*, *35*(2–3), 165–177. https://doi.org/10.1016/s0167–8760(99)00051-3

Fang, F. C., & Casadevall, A. (2015). Competitive science: Is competition ruining science? *Infection and Immunity*, *83*(4), 1229–1233. https://doi.org/10.1128/iai.02939-14

Fanselow, M. S. (1994). Neural organization of the defensive behavior system responsible for fear. *Psychonomic Bulletin & Review*, *1*(4), 429–438. https://doi.org/10.3758/bf03210947

Fantino, E. (1973). Emotion. In J. A. Nevin (Ed.), *The study of behavior: Learning, motivation, emotion, and instinct* (pp. 281–320). Scott, Foresman.

Fast, N. J., & Chen, S. (2009). When the boss feels inadequate: Power, incompetence, and aggression. *Psychological Science*, *20*(11), 1406–1413. https://doi.org/10.1111/j.1467-9280.2009.02452.x

Faucher, L., & Tappolet, C. (2008). Facts and values in emotional plasticity. In L. C. Charland & P. Zachar (Eds.), *Facts and value in emotion* (pp. 101–137). John Benjamins. https://doi.org/10.1075/ceb.4.07fau

Fazekas, P., & Kertész, G. (2011). Causation at different levels: Tracking the commitments of mechanistic explanations. *Biology & Philosophy*, *26*(3), 365–383. https://doi.org/10.1007/s10539–011-9247-5

Fazio, R. H., & Olson, M. A. (2003). Implicit attitude measures in social cognition: Their meaning and use. *Annual Review of Psychology*, *54*, 297–327. https://doi.org/10.1146/annurev.psych.54.101601.145225

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*(2), 229–238. https://doi.org/10.1037//0022-3514.50.2.229

Fehr, B., & Russell, J. A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General*, *113*(3), 464–486. https://doi.org/10.1037/0096-3445.113.3.464

Fernández-Dols, J. M. (2017). Natural facial expression: A view from psychological constructionism and pragmatics. In J. M. Fernández-Dols & J. Russell (Eds.), *The science of facial expression* (pp. 457–475). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190613501.003.0024

Fernández-Dols, J. M., & Crivelli, C. (2013). Emotion and expression: Naturalistic studies. *Emotion Review*, *5*(1), 24–29. https://doi.org/10.1177/1754073912457229

Fernández-Dols, J. M., & Ruiz-Belda, M.-A. (1997). Spontaneous facial behavior during intense emotional episodes: Artistic truth and optical truth. In J. A. Russell & J. M. Fernández-Dols (Eds.), *The psychology of facial expression* (pp. 255–294). Cambridge University Press. https://doi.org/10.1017/cbo9780511659911.013

Fini, C., Fischer, M., Bardi, L., Brass, M., & Moors, A. (2020). Support from a TMS/MEP study for a direct link between positive/negative stimuli and approach/avoidance tendencies. *Neuropsychologia*, *143*, 107496. https://doi.org/10.1016/j.neuropsychologia.2020.107496

Fiorillo, C. D., Tobler, P. N., & Schwartz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, *299*(5614), 1898–1902. https://doi.org/10.1126/science.1077349

Fischer, M., Bossuyt, E., Köster, M., Buabang, E., & Moors (2022). On the interplay between stimulus-driven and goal-directed processes in the decision to fight or flee [Manuscript submitted for publication]. Department of Psychology, KU Leuven.

Fischer, M., Fini, C., Brass, M., & Moors, A. (2020). Early approach and avoidance tendencies can be goal-directed: Support from a transcranial magnetic stimulation study. *Cognitive, Affective, and Behavioral Neuroscience*, *20*, 648–657. https://doi.org/10.3758/s13415–020-00793-6

Fischer, M., Kuppens, P., & Moors, A. (2020). When socially excluded people prefer moralizing to anti- and prosocial behavior: Support for a goal-directed account. *Motivation and Emotion*, *44*, 508–524. https://doi.org/10.1007/s11031–020-09826-6

Fodor, J. A. (1981). *Representations*. MIT Press.

Forgas, J. P. (1999). Network theories and beyond. In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 591–611). Wiley. https://doi.org/10.1002/0470013494.ch28

Forgas, J. P. (2017). Can sadness be good for you? On the cognitive, motivational, and interpersonal benefits of mild negative affect. *Australian Psychologist*, *52*(1), 3–13. https://doi.org/10.1111/ap.12232

Fossum, T. A., & Barrett, L. F. (2000). Distinguishing evaluation from description in the personality-emotion relationship. *Personality and Social Psychology Bulletin*, *26*(6), 669–678. https://doi.org/10.1177/0146167200268003

Frank, R. H. (1988). *Passions within reason: The strategic role of the emotions*. Norton.

Fredrickson, B. L. (2004). The broaden-and-build theory of positive emotions. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *359* (1449), 1367–1377. https://doi.org/10.1098/rstb.2004.1512

Fridlund, A. J. (1994). *Human facial expression: An evolutionary view*. Academic Press. https://doi.org/10.1016/c2009–0-02293-4

Fridlund, A. J. (2017). The behavioral ecology view of facial displays, 25 years later. In J. M. Fernández-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp. 77–92). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190613501.003.0005

Fridlund, A. J., Kenworthy, K., & Jaffey, A. K. (1992). Audience effects in affective imagery: Replication and extension to dysphoric imagery. *Journal of Nonverbal Behavior*, *16*, 191–212. https://doi.org/10.1007/bf00988034

Fridlund, A. J., & Russell, J. A. (2022). Evolution, emotion, and facial behavior: A 21st-century view. In L. Al-Shawaf & T. Shackelford (Eds.), *The Oxford handbook of evolution and the emotions*. Oxford University Press.

Fridlund, A. J., Sabini, J. P., Hedlund, L. E., Schaut, J. A., Shenker, J. I., & Knauer, M. J. (1990). Audience effects on solitary faces during imagery: Displaying to the people in your head. *Journal of Nonverbal Behavior*, *14*(2), 113–137. https://doi.org/10.1007/bf01670438

Frijda, N. H. (1986). *The emotions*. Cambridge University Press.

Frijda, N. H. (1988). The laws of emotion. *American Psychologist*, *43*(5), 349. https://doi.org/10.1037/0003-066x.43.5.349

Frijda, N. H. (2005). Emotion experience. *Cognition and Emotion*, *19*(4), 473–497. https://doi.org/10.1080/02699930441000346

Frijda, N. H. (2007a). Klaus Scherer's article on "What are emotions?" Comments. *Social Science Information*, *46*(3), 381–383. https://doi.org/10.1177/0539018407079694

Frijda, N. H. (2007b). *The laws of emotion*. Erlbaum. https://doi.org/10.4324/9781315086071

Frijda, N. H. (2007c). What might emotions be? Comments on the comments. *Social Science Information*, *46*(3), 433–443. https://doi.org/10.1177/05390184070460030112

Frijda, N. H. (2009). Emotion experience and its varieties. *Emotion Review*, *1*(3), 264–271. https://doi.org/10.1177/1754073909103595

Frijda, N. H., Kuipers, P., & ter Schure, L. (1989). Relations between emotion, appraisal and emotional action readiness. *Journal of Personality and Social Psychology*, *57*(2), 212–228. https://doi.org/10.1037/0022-3514.57.2.212

Frijda, N. H., & Mesquita, B. (1994). The social roles and functions of emotions. In S. Kitayama & H. R. Markus (Eds.), *Emotion and culture: Empirical studies of mutual influence* (pp. 51–87). American Psychological Association. https://doi.org/10.1037/10152-002

Frijda, N. H., & Mesquita, B. (2007, October). *The multifareous processes of emotion regulation.* [Conference presentation]. International Conference on the (Non)Expression of Emotions in Health and Disease, Tilburg, The Netherlands.

Frijda, N. H., Mesquita, B., Sonnemans, J., & Van Goozen, S. (1991). The duration of affective phenomena or emotions, sentiments and passions. In K. T. Strongman (Ed.), *International review of studies on emotion* (Vol. 1, pp. 187–225). Wiley.

Frijda, N. H., & Parrott, W. G. (2011). Basic emotions or ur-emotions? *Emotion Review*, *3*(4), 406–415. https://doi.org/10.1177/1754073911410742

Frijda, N. H., & Scherer, K. R. (2009). Emotion definition (psychological perspectives). In D. Sander & K. R. Scherer (Eds.), *Oxford companion to emotion and the affective sciences* (pp. 142–144). Oxford University Press.

Frijda, N. H., & Tcherkassof, A. (1997). Facial expressions as modes of action readiness. In J. A. Russell & J. M. Fernández-Dols (Eds.), *The psychology of facial expression* (pp. 78–102). Cambridge University Press. https://doi.org/10.1017/cbo9780511659911.006

Frijda, N. H., & Zeelenberg, M. (2001). What is the dependent? In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 141–155). Oxford University Press.

Friman, P. C., Hayes, S. C., & Wilson, K. G. (1998). Why behavior analysts should study emotion: The example of anxiety. *Journal of Applied Behavior Analysis*, *31*(1), 137–156. https://doi.org/10.1901/jaba.1998.31-137

Friston, K. J. (2003). Learning and inference in the brain. *Neural Networks*, *16*(9), 1325–1352. https://doi.org/10.1016/j.neunet.2003.06.005

Fuchs, T., & Koch, S. C. (2014). Embodied affectivity: On moving and being moved. *Frontiers in Psychology*, *5*, 508. https://doi.org/10.3389/fpsyg.2014.00508

Funkenstein, D. H., King, S. H., & Drolette, M. (1954). The direction of anger during a laboratory stress-inducing situation. *Psychosomatic Medicine*, *16*(5), 404–413. https://doi.org/10.1097/00006842-195409000-00006

Galati, D., Scherer, K. R., & Ricci-Bitti, P. E. (1997). Voluntary facial expression of emotion: Comparing congenitally blind with normally sighted encoders. *Journal of Personality and Social Psychology*, *73*(6), 1363–1379. https://doi.org/10.1037/0022-3514.73.6.1363

Galinsky, A. D., Gruenfeld, D. H., & Magee, J. C. (2003). From power to action. *Journal of Personality and Social Psychology*, *85*(3), 453–466. https://doi.org/10.1037/0022–3514.85.3.453

Garcia, D., & Rimé, B. (2019). Collective emotions and social resilience in the digital traces after a terrorist attack. *Psychological Science*, *30*(4), 617–628. https://doi.org/10.1177/0956797619831964

Garcia-Marques, L., & Ferreira, M. B. (2011). Friends and foes of theory construction in psychological science: Vague dichotomies, unified theories of cognition, and the new experimentalism. *Perspectives on Psychological Science*, *6*(2), 192–201. https://doi.org/10.1177/1745691611400239

Gati, I., & Ben-Shakhar, G. (1990). Novelty and significance in orientation and habituation: A feature-matching approach. *Journal of Experimental Psychology: General*, *119*(3), 251–263. https://doi.org/10.1037/0096-3445.119.3.251

Gendolla, G. H., & Richter, M. (2010). Effort mobilization when the self is involved: Some lessons from the cardiovascular system. *Review of General Psychology*, *14*(3), 212–226. https://doi.org/10.1037/a0019742

Gendron, M., & Barrett, L. F. (2009). Reconstructing the past: A century of ideas about emotion in psychology. *Emotion Review*, *1*(4), 316–339. https://doi.org/10.1177/1754073909338877

Gendron, M., Crivelli, C., & Barrett, L. F. (2018). Universality reconsidered: Diversity in making meaning of facial expressions. *Current Directions in Psychological Science*, *27*(4), 211–219. https://doi.org/10.1177/0963721417746794

Gendron, M., Hoemann, K., Crittenden, A. N., Mangola, S. M., Ruark, G. A., & Barrett, L. F. (2020). Emotion perception in Hadza hunter-gatherers. *Scientific Reports*, *10*(1), 1–17. https://doi.org/10.1038/s41598–020-60257-2

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014a). Cultural relativity in perceiving emotion from vocalizations. *Psychological Science*, *25*(4), 911–920. https://doi.org/10.1177/0956797613517239

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014b). Perceptions of emotion from facial expressions are not culturally universal:

Evidence from a remote culture. *Emotion*, *14*(2), 251–262. https://doi.org/10.1037/a0036052

George, M. S., Wassermann, E. M., Williams, W. A., Steppel, J., Pascual-Leone, A., Basser, P., Hallet, M., & Post, R. M. (1996). Changes in mood and hormone levels after rapid-rate transcranial magnetic stimulation (rTMS) of the prefrontal cortex. *Journal of Neuropsychiatry and Clinical Neurosciences*, *8*(2), 172–180. https://doi.org/10.1176/jnp.8.2.172

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin. https://doi.org/10.4324/9780203767764

Girard, J. M., Cohn, J. F., Yin, L., & Morency, L.-P. (2021). Reconsidering the Duchenne smile: Formalizing and testing hypotheses about eye constriction and positive emotion. *Affective Science*, *2*, 32–47. https://doi.org/10.1007/s42761–020-00030-w

Glauer, R. D. (2012). *Emergent mechanisms: Reductive explanation for limited beings*. Mentis. https://doi.org/10.30965/9783897859524

Goffin, K. (2021). Better scared than sorry: The pragmatic account of emotional representation. *Erkentniss*. https://doi.org/10.1007/s10670-021-00470-1.

Goldie, P. (2000). *The emotions: A philosophical exploration*. Oxford University Press. https://doi.org/10.1093/0199253048.001.0001

Goldie, P. (2011). Grief: A narrative account. *Ratio*, *24*(2), 119–137. https://doi.org/10.1111/j.1467-9329.2011.00488.x

Goldie, P. (2012). *The mess inside: Narrative, emotion, and the mind*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199230730.001.0001

Goldstein, D., Fink, D. , & Mettee, D. M. (1972). Cognition of arousal and actual arousal as determinants of emotion. *Journal of Personality and Social Psychology*, *21*(1), 41–51. https://doi.org/10.1037/h0031873

Goldstein, I. (2002). Are emotions feelings? A further look at hedonic theories of emotions. *Consciousness and Emotion*, *3*(1), 21–33. https://doi.org/10.1075/ce.3.1.04gol

Gonzaga, G. C., Turner, R. A., Keltner, D., Campos, B., & Altemus, M. (2006). Romantic love and sexual desire in close relationships. *Emotion*, *6*(2), 163–179. https://doi.org/10.1037/1528-3542.6.2.163

Gordon, R. M. (1974). The aboutness of emotions. *American Philosophical Quarterly*, *11*(1), 27–36.

Gordon, R. M. (1978). Emotion labeling and cognition. *Journal for the Theory of Social Behaviour*, *8*(2), 125–135. https://doi.org/10.1111/j.1468-5914.1978.tb00395.x

Gordon, R. M. (1987). *The structure of emotions: Investigations in cognitive philosophy*. Cambridge University Press.

Goudbeek, M., & Scherer, K. R. (2010). Beyond arousal: Valence and potency/control cues in the vocal expression of emotion. *The Journal of the Acoustical Society of America, 128*(3), 1322–1336. https://doi.org/10.1121/1.3466853

Gould, S. J. (1978). Sociobiology: The art of storytelling. *New Scientist*, *80*(1129), 530–533.

Grabenhorst, F., & Rolls, E. T. (2011). Value, pleasure, and choice in the ventral prefrontal cortex. *Trends in Cognitive Sciences*, *15*(2), 56–67. https://doi.org/10.1016/j.tics.2010.12.004

Gray, J. A. (1994). Three fundamental emotion systems. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion* (pp. 243–247). Oxford University Press.

Gray, M. A., Harrison, N. A., Wiens, S., & Critchley, H. D. (2007). Modulation of emotional appraisal by false physiological feedback during fMRI. *PLoS ONE*, *2*(6), e546. https://doi.org/10.1371/journal.pone.0000546

Green, O. H. (1992). *The emotions: A philosophical theory*. Kluwer.

Greenspan, P. S. (1988). *Emotions and reasons: An inquiry into emotional justification*. Routledge. https://doi.org/10.4324/9781315831848

Greenspan, P. S. (2004). Practical reasoning and emotion. In A. R. Mele & P. Rawling (Eds.), *The Oxford handbook of rationality* (pp. 206–221). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195145397.003.0011

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, *109*(1), 3–25. https://doi.org/10.1037/0033-295X.109.1.3

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Grèzes, J., Tucker, M., Armony, J., Ellis, R., & Passingham, R. E. (2003). Objects automatically potentiate action: An fMRI study of implicit processing. *European Journal of Neuroscience*, *17*(12), 2735–2740. https://doi.org/10.1046/j.1460-9568.2003.02695.x

Griffiths, P. E. (1997). *What emotions really are: The problem of psychological categories*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226308760.001.0001

Griffiths, P. E. (2003). Basic emotions, complex emotions, Machiavellian emotions. In A. Hatzimoysis (Ed.), *Philosophy and the emotions* (pp. 39–67). Cambridge University Press. https://doi.org/10.1017/cbo9780511550270.004

Griffiths, P. E. (2004a). Emotions as natural and normative kinds. *Philosophy of Science*, *71*(5), 901–911. https://doi.org/10.1086/425944

Griffiths, P. E. (2004b). Towards a Machiavellian theory of emotional appraisal. In P. Cruse & D. Evans (Eds.), *Emotion, evolution and rationality* (pp. 89–105). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198528975.003.0005

Griffiths, P. E., & Scarantino, A. (2005). Emotions in the wild: The situated perspective on emotion. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 347–353). Cambridge University Press. https://doi.org/10.1017/cbo9780511816826.023

Gross, J. J. (1998a). Antecedent- and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology. *Journal* *of Personality and Social Psychology*, *74*(1), 224–237. https://doi.org/10.1037/0022-3514.74.1.224

Gross, J. J. (1998b). The emerging field of emotion regulation: An integrative review. *Review of General Psychology*, *2*(3), 271–299. https://doi.org/10.1037/1089-2680.2.3.271

Gross, J. J. (1999). Emotion regulation: Past, present, future. *Cognition and Emotion*, *13*(5), 551–573. https://doi.org/10.1080/026999399379186

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry*, *26*(1), 1–26. https://doi.org/10.1080/1047840x.2014.940781

Gross, J. J., & Barrett, L. F. (2011). Emotion generation and emotion regulation: One or two depends on your point of view. *Emotion Review*, *3*(1), 8–16. https://doi.org/10.1177/1754073910380974

Gross, J. J., & Thompson, R. A. (2007). Emotion regulation: Conceptual foundations. In J. J. Gross (Ed.), *Handbook of emotion regulation* (pp. 3–24). Guilford.

Gyurak, A., Gross, J. J., & Etkin, A. (2011). Explicit and implicit emotion regulation: A dual-process framework. *Cognition and Emotion*, *25*(3), 400–412. https://doi.org/10.1080/02699931.2010.544160

Hafri, A., & Firestone, C. (2021). The perception of relations. *Trends in Cognitive Sciences*, *25*(6), 475–492. https://doi.org/10.1016/j.tics.2021.01.006

Hahn, U., & Chater, N. (1998). Similarity and rules: Distinct? Exhaustive? Empirically distinguishable? *Cognition*, *65*(2–3), 197–230. https://doi.org/10.1016/S0010–0277(97)00044-9

Hardwick, R. M., Forrence, A. D., Krakauer, J. W., & Haith, A. M. (2019). Time-dependent competition between goal-directed and habitual response preparation. *Nature Human Behaviour*, *3*(12), 1252–1262. https://doi.org/10.1038/s41562–019-0725-0

Hareli, S., & Hess, U. (2010). What emotional reactions can tell us about the nature of others: An appraisal perspective on person perception. *Cognition and Emotion*, *24*(1), 128–140. https://doi.org/10.1080/02699930802613828

Harlow, H. F., & Stagner, R. (1932). Psychology of feelings and emotions. I. Theory of feelings. *Psychological Review*, *39*(6), 570–589. https://doi.org/10.1037/h0072961

Harré, R. (Ed.). (1986) *The social construction of emotions*. Blackwell.

Hartsuiker, R. J., & Moors, A. (2017). On the automaticity of language processing. In H.-J. Schmid (Ed.), *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge* (pp. 201–225). De Gruyter Mouton. https://doi.org/10.1037/15969-010

Hassin, R. R., Aarts, H., & Ferguson, M. J. (2005). Automatic goal inferences. *Journal of Experimental Social Psychology*, *41*(2), 129–140. https://doi.org/10.1016/j.jesp.2004.06.008

Hastings, J., Ceusters, W., Smith, B., & Mulligan, K. (2011). The emotion ontology: Enabling interdisciplinary research in the affective sciences. In M. Beigl, H. Christiansens, T. R. Roth-Berghofer, A. Kofod-Petersen, K. R. Coventry, & H. R. Schmidtke (Eds.), *Modeling and using context. CONTEXT 2011. Lecture* *notes in computer science* (Vol. 6967, pp. 119–123). Springer. https://doi.org/10.1007/978-3-642-24279-3_14

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). *Emotional contagion*. Cambridge University Press. https://doi.org/10.1017/cbo9781139174138

Hatfield, E., Hsee, C. K., Costello, J., Weisman, M. S., & Denney, C. (1995). The impact of vocal feedback on emotional experience and expression. *Journal of Social Behavior and Personality*, *10*(2), 293–312.

Hatzimoysis, A. (2007). The case against unconscious emotions. *Analysis*, *67*(4), 292–299. https://doi.org/10.1111/j.1467-8284.2007.00693.x

Hatzimoysis, A. (2010). Emotions in Heidegger and Sartre. In P. Goldie (Ed.), *The Oxford handbook of philosophy of emotion* (pp. 215–235). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199235018.003.0010

Heider, F. (1959). Thing and medium. *Psychological Issues*, Monograph 3 (Original work published 1926).

Hélie, S., Roeder, J. L., & Ash, F. G. (2010). Evidence for cortical automaticity in rule-based categorization. *Journal of Neuroscience*, *30*(42), 14225–14234. https://doi.org/10.1523/jneurosci.2393-10.2010

Hélie, S., Waldschmidt, J. G., & Ash, F. G. (2010). Automaticity in rule-based and information-integration categorization. *Attention, Perception, & Psychophysics*, *72*, 1013–1031. https://doi.org/10.3758/app.72.4.1013

Helm, B. W. (2001). Emotions and practical reason: Rethinking evaluation and motivation. *Noûs*, *35*(2), 190–213. https://doi.org/10.1111/0029-4624.00293

Helm, B. W. (2009). Emotions as evaluative feelings. *Emotion Review*, *1*(3), 248–255. https://doi.org/10.1177/1754073909103593

Helm, B. W. (2010). Emotions and motivation: Reconsidering neo-Jamesian accounts. In P. Goldie (Ed.), *The Oxford handbook of philosophy of emotion* (pp. 303–323). Oxford University Press. https://doi.org/10.1177/1754073909103593

Helm, B. W. (2015). Emotions and recalcitrance: Reevaluating the perceptual model. *Dialectica*, *69*(3), 417–433. https://doi.org/10.1111/1746-8361.12119

Hermans, D., Spruyt, A., De Houwer, J., & Eelen, P. (2003). Affective priming with subliminally presented pictures. *Canadian Journal of Experimental Psychology*, *57*(2), 97–114. https://doi.org/10.1037/h0087416

Hernández, J. P. (2020). The attitudinal view and the integration of the particular object of emotions. *European Journal of Philosophy*, *28*(2), 478–491. https://doi.org/10.1111/ejop.12483

Herry, C., Bach, D. R., Esposito, F., Di Salle, F., Perrig, W. J., Scheffler, K., Lüthi, A., & Seifritz, E. (2007). Processing of temporal unpredictability in human and animal amygdala. *Journal of Neuroscience*, *27*(22), 5958–5966. https://doi.org/10.1523/jneurosci.5218-06.2007

Hertenstein, M. J., Holmes, R., McCullough, M., & Keltner, D. (2009). The communication of emotion via touch. *Emotion*, *9*(4), 566–573. https://doi.org/10.1037/a0016108

Hess, U., & Fischer, A. H. (2013). Emotional mimicry as social regulation. *Personality and Social Psychology Review*, *17*(2), 142–157. https://doi.org/10.1177/1088868312472607

Heyes, C. (2008). Beast machines? Questions of animal consciousness. In M. Davies & L. Weiskrantz (Eds.), *Frontiers of consciousness* (pp. 259–274). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199233151.003.0009

Heyes, C., & Dickinson, A. (1990). The intentionality of animal action. *Mind and Language*, *5*(1), 87–103. https://doi.org/10.1111/j.1468-0017.1990.tb00154.x

Higgins, E. T. (2018). What distinguishes promotion and prevention? Attaining "+ 1" from "0" as non-gain versus maintaining "0" as non-loss. *Polish Psychological Bulletin*, *49*(1), 40–49.

Hinde, R. A. (1985a). Expression and negotiation. In G. Zivin (Ed.), *The development of expressive behavior: Biology-environment interactions* (pp. 103–116). Academic Press. https://doi.org/10.1016/b978–0-12-781780-4.50011-2

Hinde, R. A. (1985b). Was "the expression of emotions" a misleading phrase? *Animal Behaviour*, *33*(3), 985–992. https://doi.org/10.1016/s0003–3472(85)80032-4

Hochschild, A. R. (1979). Emotion work, feeling rules, and social structure. *American Journal of Sociology*, *85*(3), 551–575. https://doi.org/10.1086/227049

Hoemann, K., Xu, F., & Barrett, L. F. (2019). Emotion words, emotion concepts, and emotional development in children: A constructionist hypothesis. *Developmental Psychology*, *55*(9), 1830–1849. https://doi.org/10.1037/dev0000686

Hofmann, W., Friese, M., & Strack, F. (2009). Impulse and self-control from a dual-systems perspective. *Perspectives on Psychological Science*, *4*(2), 162–176. https://doi.org/10.1111/j.1745-6924.2009.01116.x

Hohmann, G. W. (1966). Some effects of spinal cord lesions on experienced emotional feelings. *Psychophysiology*, *3*(2), 143–156. https://doi.org/10.1111/j.1469-8986.1966.tb02690.x

Holcomb, H. R. (1996). Just so stories and inference to the best explanation in evolutionary psychology. *Minds and Machines*, *6*(4), 525–540. https://doi.org/10.1007/BF00389657

Holland, J. D. (1975). Features and semantic memory: Set-theoretic or network model? *Psychological Review*, *82*(2), 154–155. https://doi.org/10.1037/h0076777

Hollenstein, T., & Lanteigne, D. (2014). Models and methods of emotional concordance. *Biological Psychology*, *98*, 1–5. https://doi.org/10.1016/j.biopsycho.2013.12.012

Hommel, B. (2009). Action control according to TEC (theory of event coding). *Psychological Research PRPF*, *73*(4), 512–526. https://doi.org/10.1007/s00426–009-0234-2

Hommel, B. (2013). Dancing in the dark: No role for consciousness in action control. *Frontiers in Psychology*, *4*, 380. https://doi.org/10.3389/fpsyg.2013.00380

Hommel, B. (2017). Consciousness and action control. In T. Egner (Ed.) *Handbook of cognitive control* (pp. 111–123). Wiley. https://doi.org/10.1002/9781118920497.ch7

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning.

Behavioral and Brain Sciences, *24*(5), 849–878. https://doi.org/10.1017/s0140525x01000103

Hopp, H., Troy, A. S., & Mauss, I. B. (2011). The unconscious pursuit of emotion regulation: Implications for psychological health. *Cognition and Emotion*, *25* (3), 532–545. https://doi.org/10.1080/02699931.2010.532606

Howle, T. C., & Eklund, R. C. (2013). The effect of induced self-presentation concerns on cognitive appraisal and affect. *Anxiety, Stress & Coping*, *26*(6), 700–710. https://doi.org/10.1080/10615806.2013.763934

Hufendiek, R. (2016), William James and John Dewey on embodied action-oriented emotions. In M. Jung & R. Madzia (Eds.), *Pragmatism and embodied cognitive science: From bodily intersubjectivity to symbolic articulation* (pp. 269–288). De Gruyter. https://doi.org/10.1515/9783110480238-014

Hull, C. L. (1943). *Principles of behavior: An introduction to behavior theory*. Appleton Century Crofts.

Hume, D. (1739). *A treatise on human nature*. Clarendon Press. https://doi.org/10.1093/oseo/instance.00032872

Hunt, W. A. (1941). Recent developments in the field of emotion. *Psychological Bulletin*, *38*(5), 249–276. https://doi.org/10.1037/h0054615

Hurley, S. L. (1998). *Consciousness in action*. Harvard University Press.

Hursthouse, R. (1991). Arational actions. *The Journal of Philosophy*, *88*(2), 57–68. https://doi.org/10.2307/2026906

Hutchinson, J. B., & Barrett, L. F. (2019). The power of predictions: An emerging paradigm for psychological research. *Current Directions in Psychological Science*, *28*(3), 280–291. https://doi.org/10.1177/0963721419831992

Hutto, D. D. (2012). Truly enactive emotion. *Emotion Review*, *4*(2), 176–181. https://doi.org/10.1177/1754073911430134

Hutto, D. D., & Myin, E. (2017). *Evolving enactivism: Basic minds meet content*. MIT Press. https://doi.org/10.7551/mitpress/9780262036115.001.0001

Hutto, D. D., Robertson, I., & Kirchhoff, M. D. (2018). A new, better BET: Rescuing and revising basic emotion theory. *Frontiers in Psychology*, *9*, 1217. https://doi.org/10.3389/fpsyg.2018.01217

Ichino, A. (2019). Imagination and belief in action. *Philosophia*, *47*(5), 1517–1534. https://doi.org/10.1007/s11406–019-00067-7

Ierodiakonou, K. (1993). The Stoic division of philosophy. *Phronesis*, *38*(1), 57–74. https://doi.org/10.1163/156852893321052451

Irons, D. (1897a). The nature of emotion. *Philosophical Review*, *6*(3), 242–256. https://doi.org/10.2307/2175881

Irons, D. (1897b). The primary emotions. *Philosophical Review*, *6*(6), 626–645. https://doi.org/10.2307/2175866

Isen, A. M. (1984). Toward understanding the role of affect in cognition. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. *3*, pp. 179–236). Erlbaum.

Isen, A. M., Shalker, T. E., Clark, M. S., & Karp, L. (1978). Affect, accessibility of material in memory, and behavior: A cognitive loop? *Journal of Personality and Social Psychology*, *36*(1), 1–12. https://doi.org/10.1037/0022-3514.36.1.1

Izard, C. E. (1969). The emotions and emotion constructs in personality and culture research. In R. B. Cattell (Ed.), *Handbook of modern personality theory* (pp. 496–510). Aldine.

Izard, C. E. (1972). Anxiety: A variable combination of interacting fundamental emotions. In C. D. Spielberger (Ed.), *Anxiety: Current trends in theory and research* (Vol. 1, pp. 55–106). Academic Press. https://doi.org/10.1016/b978–0-12-657401-2.50010-1

Izard, C. E. (1977). *Human emotions*. Plenum Press. https://doi.org/10.1007/978-1-4899-2209-0

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives in Psychological Science*, *2*(3), 260–280. https://doi.org/10.1111/j.1745-6916.2007.00044.x

Izard, C. E. (2011). Forms and functions of emotions: Matters of emotion–cognition interactions. *Emotion Review*, *3*(4), 371–378. https://doi.org/10.1177/1754073911410737

Izard, C. E., Ackerman, B. P., Schoff, K. M., & Fine, S. E. (2000). Self-organization of discrete emotions, emotion patterns, and emotion-cognition relations. In M. D. Lewis & I. Granic (Eds.), *Emotion, development, and self-organization: Dynamic systems approaches to emotional development* (pp. 15–36). Cambridge University Press. https://doi.org/10.1017/cbo9780511527883.003

Izard, C. E., & Malatesta, C. Z. (1987). Perspectives on emotional development: Differential emotions theory of early emotional development. In J. D. Osofsky (Ed.), *Handbook of infant development* (2nd ed., pp. 494–554). Wiley Interscience.

Jacob, P. (2019, Winter). Intentionality. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. https://plato.stanford.edu/entries/intentionality/

James, W. (1884). What is an emotion? *Mind*, *9*, 188–205. https://doi.org/10.1093/mind/os-ix.34.188

James, W. (1890a). *The principles of psychology*. Vol. 1. Holt. https://doi.org/10.1037/10538-000

James, W. (1890b). *The principles of psychology*. Vol. 2. Holt.

James, W. (1894). The physical basis of emotion. *Psychological Review*, *1*(5), 516–529. https://doi.org/10.1037/h0065078

James, W. (1897). *The will to believe and other essays in popular philosophy*. Longmans, Green. https://doi.org/10.2307/2175898

Jaworski, W. (2018). Problems for an ontology of emotions. In H. Naar & F. Teroni (Eds.), *The ontology of emotions* (pp. 14–36). Cambridge University Press. https://doi.org/10.1017/9781316275221.009

Johansen, J. P., Cain, C. K., Ostroff, L. E., & LeDoux, J. E. (2011). Molecular mechanisms of fear learning and memory. *Cell*, *147*(3), 509–524. https://doi.org/10.1016/j.cell.2011.10.009

Johnson, M. K., & Multhaup, K. S. (1992). Emotion and MEM. In S.-A. Christianson (Ed.), *The handbook of emotion and memory: Research and theory* (pp. 33–66). Psychology Press.

Johnson-Laird, P. N., & Oatley, K. (1992). Basic emotions, rationality, and folk theory. *Cognition and Emotion*, *6*(3–4), 201–223. https://doi.org/10.1080/02699939208411069

Johnston, M. (2001). The authority of affect. *Philosophy and Phenomenological Research*, *53*(1), 181–214. https://doi.org/10.1111/j.1933-1592.2001.tb00097.x

Johnstone, T., van Reekum, C. M., Hird, K., Kirsner, K., & Scherer, K. R. (2005). Affective speech elicited with a computer game. *Emotion*, *5*(4), 513–518. https://doi.org/10.1037/1528-3542.5.4.513

Johnstone, T., van Reekum, C. M., & Scherer, K. R. (2001). Vocal expression correlates of appraisal processes. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 271–284). Oxford University Press. https://doi.org/10.1080/02699930341000167

Jones, C. R., Vilensky, M. R., Vasey, M. W., & Fazio, R. H. (2013). Approach behavior can mitigate predominately univalent negative attitudes: Evidence regarding insects and spiders. *Emotion*, *13*(5), 989–996. https://doi.org/10.1037/a0033164

Jostmann, N. B., Koole, S. L., Van Der Wulp, N. Y., & Fockenberg, D. A. (2005). Subliminal affect regulation: The moderating role of action vs. state orientation. *European Psychologist*, *10*(3), 209–217. https://doi.org/10.1027/1016-9040.10.3.209

Juslin, P. N., & Scherer, K. R. (2005). Vocal expression of affect. In J. A. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *The new handbook of methods in nonverbal behavior research* (pp. 65–135). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198529620.003.0003

Kagan, J. (2007). *What is emotion? History, measures, and meanings*. Yale University Press.

Kahneman, D., & Frederick, S. (2005). *A model of heuristic judgment*. Cambridge University Press.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, *93*(2), 136–153. https://doi.org/10.1037/0033-295X.93.2.136

Kaiser, M., & Krickel, B. (2017). The metaphysics of constitutive mechanistic phenomena. *British Journal for the Philosophy of Science*, *68*(3), 745–779. https://doi.org/10.1093/bjps/axv058

Kanter, J. W., Holman, G., & Wilson, K. G. (2014). Where is the love? Contextual behavioral science and behavior analysis. *Journal of Contextual Behavioral Science*, *3*(2), 69–73. https://doi.org/10.1016/j.jcbs.2014.02.001

Kappas, A. (2003). What facial activity can and cannot tell us about emotions. In M. Katsikis (Ed.), *The human face: Measurement and meaning* (pp. 215–234). Kluwer Academic Publishers. https://doi.org/10.1007/978-1-4615-1063-5_11

Kappas, A. (2006). Appraisals are direct, immediate, intuitive, and unwitting … and some are reflective … *Cognition and Emotion*, *20*(7), 952–975. https://doi.org/10.1080/02699930600616080

Kappas, A. (2008). Psssst! Dr. Jekyll and Mr. Hyde are actually the same person! A tale of regulation and emotion. In M. Vandekerckhove, C. von Scheve, S.

Ismer, S. Jung, & S. Kronast (Eds.), *Regulating emotions: Culture, social necessity, and biological inheritance* (pp. 15–38). Wiley. https://doi.org/10.1002/9781444301786.ch1

Kappas, A. (2011). Emotion and regulation are one! *Emotion Review*, *3*(1), 17–25. https://doi.org/10.1177/1754073910380971

Kassam, K. S., Markey, A. R., Cherkassky, V. L., Loewenstein, G., & Just, M. A. (2013). Identifying emotions on the basis of neural activation. *PLoS ONE*, *8*(6), e66032. https://doi.org/10.1371/journal.pone.0066032

Kassam, K. S., & Mendes, W. B. (2013). The effects of measuring emotion: Physiological reactions to emotional situations depend on whether someone is asking. *PLoS ONE*, *8*(6), e64959. https://doi.org/10.1371/journal.pone.0064959

Kaup, D. J., & Clarke, T. L. (2005). Where's the example? *Behavioral and Brain Sciences*, *28*(2), 210. https://doi.org/10.1017/s0140525x05380046

Kawakami, K., Phills, C. E., Steele, J. R., & Dovidio, J. F. (2007). (Close) distance makes the heart grow fonder: Improving implicit racial attitudes and interracial interactions through approach behaviors. *Journal of Personality and Social Psychology*, *92*(6), 957–971. https://doi.org/10.1037/0022-3514.92.6.957

Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. MIT Press.

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition and Emotion*, *13*(5), 505–521. https://doi.org/10.1080/026999399379168

Keltner, D., & Haidt, J. (2001). Social functions of emotions. In T. J. Mayne & G. A. Bonanno (Eds.), *Emotions: Current issues and future directions* (pp. 192–213). Guilford. https://doi.org/10.1080/026999399379168

Keltner, D., Haidt, J., & Shiota, M. N. (2006). Social functionalism and the evolution of emotions. In M. Schaller, J. A. Simpson, & D. T. Kenrick (Eds.), *Evolution and social psychology* (pp. 115–142). Psychology Press. https://doi.org/10.4324/9780203782965

Keltner, D., Sauter, D., Tracy, J., & Cowen, A. (2019). Emotional expression: Advances in basic emotion theory. *Journal of Nonverbal Behavior*, *43*, 133–160. https://doi.org/10.1007/s10919–019-00293-3

Keltner, D., Tracy, J. L., Sauter, D. A., Cordaro, D. C., & McNeil, G. D. (2016). Expression of emotion. In L. F. Barrett, M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (4th ed., pp. 467–482). Guilford.

Keltner, D., Tracy, J. L., Sauter, D., & Cowen, A. (2019). What basic emotion theory really says for the twenty-first century study of emotion. *Journal of Nonverbal Behavior*, *43*, 195–201. https://doi.org/10.1007/s10919–019-00298-y

Kenny, A. (1963). *Action, emotion, and will*. Routledge. https://doi.org/10.4324/9780203711460

Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science*, *4*(6), 533–550. https://doi.org/10.1111/j.1745-6924.2009.01164.x

Khoo, G. S., & Adkins, B. (2020). Catharsis. *The International Encyclopedia of Media Psychology*, 1–7. https://doi.org/10.1002/9781119011071.iemp0179

Kitayama, S., Mesquita, B., & Karasawa, M. (2006). Cultural affordances and emotional experience: Socially engaging and disengaging emotions in Japan and the United States. *Journal of Personality and Social Psychology*, *91*(5), 890–903. https://doi.org/10.1037/0022-3514.91.5.890

Kleinginna, P. R., & Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion*, *5*(4), 345–379. https://doi.org/10.1007/bf00992553

Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K., & Wager, T. D. (2008). Functional grouping and cortical–subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *Neuroimage*, *42*(2), 998–1031. https://doi.org/10.1016/j.neuroimage.2008.03.059

Koch, S. C. (2014). Rhythm is it: Effects of dynamic body feedback on affect and attitudes. *Frontiers in Psychology*, *5*, 537. https://doi.org/10.3389/fpsyg.2014.00537

Kolodyazhniy, V., Kreibig, S. D., Gross, J. J., Roth, W. T., & Wilhelm, F. H. (2011). An affective computing approach to physiological emotion specificity: Toward subject-independent and stimulus-independent classification of film-induced emotions. *Psychophysiology*, *48*(7), 908–922. https://doi.org/10.1111/j.1469-8986.2010.01170.x

Koole, S. L. (2009). The psychology of emotion regulation: An integrative review. *Cognition and Emotion*, *23*(1), 4–41. https://doi.org/10.1080/02699930802619031

Köster, M., Moors, A., De Houwer, J., Ross-Hellauer, T., Van Nieuwerburgh, I., & Verbruggen, F. (2021). Behavioral reluctance in adopting open access publishing: Insights from a goal-directed perspective. *Frontiers in Psychology*, *12*, 1111. https://doi.org/10.3389/fpsyg.2021.649915

Kotovsky, L., & Baillargeon, R. (1994). Calibration-based reasoning about collision events in 11-month-old infants. *Cognition*, *51*(2), 107–129. https://doi.org/10.1016/0010-0277(94)90012-4

Kovacs, P., Hélie, S., Tran, A. N., & Ashby, F. G. (2021). A neurocomputational theory of how rule-guided behaviors become automatic. *Psychological Review*, *128*(3), 488–508. https://doi.org/10.1037/rev0000271

Kozlik, J., Neumann, R., & Lozo, L. (2015). Contrasting motivational orientation and evaluative coding accounts: On the need to differentiate the effectors of approach/avoidance responses. *Frontiers in Psychology*, *6*, 563. https://doi.org/10.3389/fpsyg.2015.00563

Kragel, P. A., Koban, L., Barrett, L. F., & Wager, T. D. (2018). Representation, pattern information, and brain signatures: From neurons to neuroimaging. *Neuron*, *99*(2), 257–273. https://doi.org/10.1016/j.neuron.2018.06.009

Kragel, P. A., & LaBar, K. S. (2014). Advancing emotion theory with multivariate pattern classification. *Emotion Review*, *6*(2), 160–174. https://doi.org/10.1177/1754073913512519

Kragel, P. A., & LaBar, K. S. (2015). Multivariate neural biomarkers of emotional states are categorically distinct. *Social Cognitive and Affective Neuroscience*, *10*(11), 1437–1448. https://doi.org/10.1093/scan/nsv032

Kragel, P. A., & LaBar, K. S. (2016). Decoding the nature of emotion in the brain. *Trends in Cognitive Sciences*, *20*(6), 444–455. https://doi.org/10.1016/j.tics.2016.03.011

Kreibig, S. D. (2010). Autonomic nervous activity in emotion: A review. *Biological Psychology*, *84*(3), 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010

Kreibig, S. D. (2012). Emotion, motivation, and cardiovascular response. In R. A. Wright & G. H. E. Gendolla (Eds.), *How motivation affects cardiovascular response: Mechanisms and applications* (pp. 93–117). American Psychological Association. https://doi.org/10.1037/13090-005

Kreibig, S. D., Gendolla, G. H., & Scherer, K. R. (2012). Goal relevance and goal conduciveness appraisals lead to differential autonomic reactivity in emotional responding to performance feedback. *Biological Psychology*, *91*(3), 365–375. https://doi.org/10.1016/j.biopsycho.2012.08.007

Kreibig, S. D., Wilhelm, F. H., Roth, W. T., & Gross, J. J. (2007). Cardiovascular, electrodermal, and respiratory response patterns to fear-and sadness-inducing films. *Psychophysiology*, *44*(5), 787–806. https://doi.org/10.1111/j.1469-8986.2007.00550.x

Krickel, B. (2017). Making sense of interlevel causation in mechanisms from a metaphysical perspective. *Journal for General Philosophy of Science*, *48*(3), 453–468. https://doi.org/10.1007/s10838–017-9373-0

Krieglmeyer, R., De Houwer, J., & Deutsch, R. (2013). On the nature of automatically triggered approach–avoidance behavior. *Emotion Review*, *5*(3), 280–284. https://doi.org/10.1177/1754073913477501

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological Review*, *118*(1), 97–109. https://doi.org/10.1037/a0020762

Kruglanski, A. W., & Szumowska, E. (2020). Habitual behavior is goal-driven. *Perspectives on Psychological Science*, *15*(5), 1256–1271. https://doi.org/10.1177/1745691620917676

Krumhuber, E. G., Kappas, A., & Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: A review. *Emotion Review*, *5*(1), 41–46. https://doi.org/10.1177/1754073912451349

Kumar, A. A. (2020). Semantic memory: A review of methods, models, and current challenges. *Psychonomic Bulletin & Review*, *28*, 40–80. https://doi.org/10.3758/s13423–020-01792-x

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review*, *103*(2), 284–308. https://doi.org/10.1037/0033-295X.103.2.284

Kunst-Wilson, W. R., & Zajonc, R. B. (1980). Affective discrimination of stimuli that cannot be recognized. *Science*, *207*(4430), 557–558. https://doi.org/10.1126/science.7352271

Kuppens, P. (2013). Comment: Appraisal affords flexibility to emotion in more ways than one. *Emotion Review*, *5*(2), 176–179. https://doi.org/10.1177/1754073912468167

Kuppens, P., Van Mechelen, I., Smits, D. J. M., & De Boeck, P. (2003). The appraisal basis of anger: Specificity, necessity and sufficiency of components. *Emotion*, *3*(3), 254–269. https://doi.org/10.1037/1528-3542.3.3.254

Kuppens, P., Van Mechelen, I., Smits, D. J., De Boeck, P., & Ceulemans, E. (2007). Individual differences in patterns of appraisal and anger experience. *Cognition and Emotion*, *21*(4), 689–713. https://doi.org/10.1080/02699930600859219

Lähteenmäki, M., Hyönä, J., Koivisto, M., & Nummenmaa, L. (2015). Affective processing requires awareness. *Journal of Experimental Psychology: General*, *144*(2), 339–365. https://doi-org.kuleuven.e-brognnen.be/10.1037/xge0000040

Laird, J. D. (1974). Self-attribution of emotion: The effects of expressive behavior on the quality of emotional experience. *Journal of Personality and Social Psychology*, *29*(4), 475–486. https://doi.org/10.1037/h0036125

Lakin, J. L., Chartrand, T. L., & Arkin, R. M. (2008). I am too just like you: Nonconscious mimicry as an automatic behavioral response to social exclusion. *Psychological Science*, *19*(8), 816–822. https://doi.org/10.1111/j.1467-9280.2008.02162.x

Lamarck, J.-B. (1809). *Philosophie zoologique* [*Zoological philosophy*]. Dentu et L'Auteur. https://doi.org/10.5962/bhl.title.50194

Lambie, J., & Marcel, A. (2002). Consciousness and emotion experience: A theoretical framework. *Psychological Review*, *109*(2), 219–259. https://doi.org/10.1037/0033-295x.109.2.219

Lammers, J., Galinsky, A. D., Gordijn, E. H., & Otten, S. (2008). Illegitimacy moderates the effects of power on approach. *Psychological Science*, *19*(6), 558–564. https://doi.org/10.1111/j.1467-9280.2008.02123.x

Lang, P. J. (1968). Fear reduction and fear behavior: Problems in treating a construct. In J. M. Shlien (Ed.), *Research in psychotherapy* (Vol. 3, pp. 90–102). American Psychological Association. https://doi.org/10.1037/10546-004

Lang, P. J. (1984). Cognition in emotion: Concept and action. In C. E. Izard, J. Kagan, & R. B. Zajonc (Eds.), *Emotions, cognition, and behavior* (pp. 192–228). Cambridge University Press.

Lang, P. J. (1985). The cognitive psychology of emotion: Fear and anxiety. In A. H. Tuma & J. D. Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 131–170). Erlbaum. https://doi.org/10.4324/9780203728215-10

Lang, P. J. (1993). The network model of emotion: Motivational connections. In R. Wyer & T. K. Srull (Eds.), *Advances in social cognition* (Vol. 6, pp. 109–133). Erlbaum.

Lang, P. J. (1994). The motivational organization of emotion: Affect-reflex connections. In S. H. M. van Goozen, N. E. Van de Poll, & J. A. Sergeant (Eds.), *Emotions: Essays on emotion theory* (pp. 61–93). Erlbaum. https://doi.org/10.4324/9781315806914

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review*, *97*(3), 377–395. https://doi.org/10.1037/0033-295x.97.3.377

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2005). International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center* *for the Study of Emotion and Attention*, *1*, 39–58. https://doi.org/10.1037/t66667–000

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, *30* (3), 261–273. https://doi.org/10.1111/j.1469-8986.1993.tb03352.x

Lange, C. G. (1922). The emotions: A psychophysiological study (I. A. Haupt, Trans.). In K. Dunlap (Ed.), *The emotions* (pp. 33–90). Williams & Wilkins. (Original work published 1885). https://doi.org/10.1037/10735-002

Lange, J., Dalege, J., Borsboom, D., van Kleef, G. A., & Fischer, A. H. (2020). Toward an integrative psychometric model of emotions. *Perspectives on Psychological Science*, *15*(2), 444–468. https://doi.org/10.1177/1745691619895057

Lange, J., & Zickfeld, J. H. (2021). Emotions as overlapping causal networks of emotion components: Implications and methodological approaches. *Emotion Review*, *13*(2), 157–167. https://doi.org/10.1177/1754073920988787

Lanzetta, J., Cartwright-Smith, J., & Kleck, R. (1976). Effects of nonverbal dissimulation on emotional experience and autonomic arousal. *Journal of Personality and Social Psychology*, *33*(3), 354–370. https://doi.org/10.1037/0022-3514.33.3.354

Lapinski, M. K., & Rimal, R. N. (2005). An explication of social norms. *Communication Theory*, *15*(2), 127–147. https://doi.org/10.1111/j.1468-2885.2005.tb00329.x

Larsen, J. T., Berntson, G. G., Poehlmann, K. M., Ito, T. A., & Cacioppo, J. T. (2008). The psychophysiology of emotion. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (pp. 180–195). Guilford.

Larsen, R. J., & Prizmic, Z. (2004). Affect regulation. In R. F. Baumeister & K. D. Vohs (Eds.), *Handbook of self-regulation: Research, theory, and applications* (pp. 40–61). Guilford.

Laukka, P., & Elfenbein, H. A. (2012). Emotion appraisal dimensions can be inferred from vocal expressions. *Social Psychological and Personality Science*, *3* (5), 529–536. https://doi.org/10.1177/1948550611428011

Lauria, F. (2017). The "guise of the ought to be": A deontic view of the intentionality of desire. In F. Lauria & J. Deonna (Eds.), *The nature of desire* (pp. 139–164). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199370962.003.0006

Lazarus, R. S. (1982). Thoughts on the relations between emotion and cognition. *American Psychologist*, *37*(9), 1019–1024. https://doi.org/10.1037/0003-066X.37.9.1019

Lazarus, R. S. (1984). On the primacy of cognition. *American Psychologist*, *39*(2), 124–129. https://doi.org/10.1037/0003–066X.39.2.124

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press.

Lazarus, R. S. (1999). The cognition-emotion debate: A bit of history. In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 3–19). Wiley. https://doi.org/10.1002/0470013494.ch1

Lazarus, R. S. (2001). Relational meaning and discrete emotions. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 37–67). Oxford University Press.

Leary, M. R., Twenge, J. M., & Quinlivan, E. (2006). Interpersonal rejection as a determinant of anger and aggression. *Personality and Social Psychology Review*, *10*(2), 111–132. https://doi.org/10.1207/s15327957pspr1002_2

LeDoux, J. E. (1989). Cognitive-emotional interactions in the brain. *Cognition and Emotion*, *3*(4), 267–289. https://doi.org/10.1080/02699938908412709

LeDoux, J. E. (1991). Emotion and the limbic system concept. *Concepts in Neuroscience*, *2*, 169–199.

LeDoux, J. E. (1996). *The emotional brain: The mysterious underpinnings of emotional life*. Simon & Schuster.

LeDoux, J. E. (2003). *Synaptic self: How our brains become who we are*. Penguin.

LeDoux, J. E. (2012a). Evolution of human emotion: A view through fear. *Progress in Brain Research*, *195*, 431–442. https://doi.org/10.1016/b978–0-444-53860-4.00021-0

LeDoux, J. E. (2012b). Rethinking the emotional brain. *Neuron*, *73*(4), 653–676. https://doi.org/10.1016/j.neuron.2012.02.004

LeDoux, J. E., & Brown, R. (2017). A higher-order theory of emotional consciousness. *Proceedings of the National Academy of Sciences*, *114*(10), E2016–E2025. https://doi.org/10.1073/pnas.1619316114

LeDoux, J. E., & Daw, N. D. (2018). Surviving threats: Neural circuit and computational implications of a new taxonomy of defensive behaviour. *Nature Reviews Neuroscience*, *19*(5), 269–282. https://doi.org/10.1038/nrn.2018.22

LeDoux, J. E., & Hofmann, S. G. (2018). The subjective experience of emotion: A fearful view. *Current Opinion in Behavioral Sciences*, *19*, 67–72. https://doi.org/10.1016/j.cobeha.2017.09.011

LeDoux, J. E., Romanski, L., & Xagoraris, A. (1989). Indelibility of subcortical emotional memories. *Journal of Cognitive Neuroscience*, *1*(3), 238–243. https://doi.org/10.1162/jocn.1989.1.3.238

Lee, D. H., & Anderson, A. K. (2017). Form and function in facial expressive behavior. In J. M. Fernández-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp. 173–196). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190613501.003.0010

Leeper, R. W. (1948). A motivational theory of emotion to replace "emotion as disorganized response." *Psychological Review*, *55*(1), 5–21. https://doi.org/10.1037/h0061922

Lemaire, S. (2012). Values and value judgments: New perspectives. *Dialogue*, *51*(1), 1–6. https://doi.org/10.1017/s0012217312000406

Lemaire, S. (2014). Norms for emotions: Intrinsic or extrinsic ? In J. Dutant, D. Fassio, & A. Meylan (Eds.), *Liber Amicorum Pascal Engel* (pp. 828–843). University of Geneva. https://doi.org/10.1007/s11229–017-1352-9

Lemaire, S. (2021). Subjectivism without idealization and adaptive preferences. *Utilitas*, *33*(1), 85–100. https://doi.org/10.1017/s095382082000031x

LeMoult, J., & Gotlib, I. H. (2019). Depression: A cognitive perspective. *Clinical Psychology Review*, *69*, 51–66. https://doi.org/10.1016/j.cpr.2018.06.008

Lench, H. C., Flores, S. A., & Bench, S. W. (2011). Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology:

A meta-analysis of experimental emotion elicitations. *Psychological Bulletin*, *137*(5), 834–855. https://doi.org/10.1037/a0024244

León, F., Szanto, T., & Zahavi, D. (2019). Emotional sharing and the extended mind. *Synthese*, *196*(12), 4847–4867. https://doi.org/10.1007/s11229–017-1351-x

Lépine, S. (2016). Découvertes émotionelles et motivations affectives [Emotional discoveries and affective motivations]. *Revue Philosophique de la France et de l'Etranger*, *141*(2), 179–196. https://doi.org/10.3917/rphi.162.0179

Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology*, *81*(1), 146–159. https://doi.org/10.1037/0022-3514.81.1.146

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology*, *66*, 799–823. https://doi.org/10.1146/annurev-psych-010213-115043

Lerner, J. S., & Tiedens, L. Z. (2006). Portrait of the angry decision maker: How appraisal tendencies shape anger's influence on cognition. *Journal of Behavioral Decision Making*, *19*(2), 115–137. https://doi.org/10.1002/bdm.515

Leuridan, B. (2012). Three problems for the mutual manipulability account of constitutive relevance in mechanisms. *British Journal for the Philosophy of Science*, *63*(2), 399–427. https://doi.org/10.1093/bjps/axr036

Levenson, R. W. (1992). Autonomic nervous system differences among emotions. *Psychological Science*, *3*(1), 23–27. https://doi.org/10.1111/j.1467-9280.1992.tb00251.x

Levenson, R. W. (1994). Human emotions: A functional view. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 123–126). Oxford University Press.

Levenson, R. W. (2011). Basic emotion questions. *Emotion Review*, *3*(4), 379–386. https://doi.org/10.1177/1754073911410743

Levenson, R. W. (2014). The autonomic nervous system and emotion. *Emotion Review*, *6*(2), 100–112. https://doi.org/10.1177/1754073913512003

Levenson, R. W., Ekman, P., & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology*, *27*(4), 363–384. https://doi.org/10.1111/j.1469-8986.1990.tb02330.x

Levenson, R. W., Ekman, P., Heider, K., & Friesen, W. V. (1992). Emotion and autonomic nervous system activity in the Minangkabau of West Sumatra. *Journal of Personality and Social Psychology*, *62*(6), 972–988. https://doi.org/10.1037/0022-3514.62.6.972

Leventhal, H. (1980). Toward a comprehensive theory of emotion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 13, pp. 139–207). Academic Press. https://doi.org/10.1016/S0065–2601(08)60132-X

Leventhal, H. (1984). A perceptual-motor theory of emotion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 17, pp. 117–182). Academic Press. https://doi.org/10.1016/s0065–2601(08)60119-7

Leventhal, H., & Scherer, K. (1987). The relationship of emotion to cognition: A functional approach to a semantic controversy. *Cognition and Emotion*, *1*(1), 3–28. https://doi.org/10.1080/02699938708408361

Lewin, K. (1935). *A dynamic theory of personality*. McGraw-Hill.

Lewis, M. (2000). The emergence of human emotions. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 265–280). Guilford.

Lewis, M., Alessandri, S. M., & Sullivan, M. W. (1990). Violation of expectancy, loss of control, and anger expressions in young infants. *Developmental Psychology*, *26*(5), 745–751. https://doi.org/10.1037/0012-1649.26.5.745

Lewis, M., & Ramsay, D. (2005). Infant emotional and cortisol responses to goal blockage. *Child Development*, *76*(2), 518–530. https://doi.org/10.1111/j.1467-8624.2005.00860.x

Lewis, M. D. (2005). Bridging emotion theory and neurobiology through dynamic systems modeling. *Behavioral and Brain Sciences*, *28*(2), 169–245. https://doi.org/10.1017/s0140525x0500004x

Lewis, M. D., & Liu, Z. X. (2011). Three time scales of neural self-organization underlying basic and nonbasic emotions. *Emotion Review*, *3*(4), 416–423. https://doi.org/10.1177/1754073911410748

Lhommet, M., & Marsella, S. C. (2014). Expressing emotion through posture and gesture. In R. A. Calvo, S. D'Mello, J. Gratch, & A. Kappas (Eds.), *The Oxford handbook of affective computing* (pp. 273–285). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199942237.013.039

Li, S., & Deng, W. (2020). Deep facial expression recognition: A survey. *IEEE Transactions on Affective Computing*. https://doi.org/10.1109/taffc.2020.2981446

Lieberman, M. D., Jarcho, J. M., & Satpute, A. B. (2004). Evidence-based and intuition-based self-knowledge: An fMRI study. *Journal of Personality and Social Psychology*, *87*(4), 421–435. https://doi.org/10.1037/0022-3514.87.4.421

Lindquist, K. A. (2017). The role of language in emotion: Existing evidence and future directions. *Current Opinion in Psychology*, *17*, 135–139. https://doi.org/10.1016/j.copsyc.2017.07.006

Lindquist, K. A., & Barrett, L. F. (2008). Constructing emotion: The experience of fear as a conceptual act. *Psychological Science*, *19*(9), 898–903. https://doi.org/10.1111/j.1467-9280.2008.02174.x

Lindquist, K. A., & Barrett, L. F. (2012). A functional architecture of the human brain: Emerging insights from the science of emotion. *Trends in Cognitive Sciences*, *16*(11), 533–540. https://doi.org/10.1016/j.tics.2012.09.005

Lindquist, K. A., Barrett, L. F., Bliss-Moreau, E., & Russell, J. A. (2006). Language and the perception of emotion. *Emotion*, *6*(1), 125–138. https://doi.org/10.1037/1528-3542.6.1.125

Lindquist, K. A., & Gendron, M. (2013). What's in a word? Language constructs emotion perception. *Emotion Review*, *5*(1), 66–71. https://doi.org/10.1177/1754073912451351

Lindquist, K. A., MacCormack, J. K., & Shablack, H. (2015). The role of language in emotion: Predictions from psychological constructionism. *Frontiers in Psychology*, *6*, 444. https://doi.org/10.3389/fpsyg.2015.00444

Lindquist, K. A., Satpute, A. B., & Gendron, M. (2015). Does language do more than communicate emotion? *Current Directions in Psychological Science*, *24*(2), 99–108. https://doi.org/10.1177/0963721414553440

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences*, *35*(3), 121–143. https://doi.org/10.1017/s0140525x11000446

Liu, J., Shen, K., & Li, H. (2019). How state anxiety and attentional bias interact with each other: The moderating effect of cognitive appraisal. *Attention, Perception, & Psychophysics*, *81*, 694–706. https://doi.org/10.3758/s13414–018-01650-y

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, *127*(2), 267–286. https://doi.org/10.1037/0033-2909.127.2.267

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*(4), 492–527. https://doi.org/10.1037/0033-295X.95.4.492

Lohr, J. M., Olatunji, B. O., Baumeister, R. F., & Bushman, B. J. (2007). The psychology of anger venting and empirically supported alternatives that do no harm. *Scientific Review of Mental Health Practice*, *5*(1), 53–64.

Longworth, F., & Scarantino, A. (2010). The disjunctive theory of art: The cluster account reformulated. *British Journal of Aesthetics*, *50*(2), 151–167. https://doi.org/10.1093/aesthj/ayq001

Lyons, J. (1977). *Semantics* (Vol. 1). Cambridge University Press. https://doi.org/10.1017/CBO9781139165693

Lyons, W. (1980). *Action, emotion and will*. Routledge & Kegan Paul.

MacCormack, J. K., & Lindquist, K. A. (2017). Bodily contributions to emotion: Schachter's legacy for a psychological constructionist view on emotion. *Emotion Review*, *9*(1), 36–45. https://doi.org/10.1177/1754073916639664

MacKay, D. M. (1951). Mind-like behaviour in artefacts. *Bulletin of the British Society for the History of Science*, *1*(S5), 164–165. https://doi.org/10.1017/s0950563600001901

Mackie, J. L. (1974). *The cement of the universe: A study of causation*. Clarendon Press. https://doi.org/10.1093/0198246420.001.0001

MacLean, P. D. (1949). Psychosomatic disease and the "visceral brain": Recent developments bearing on the Papez theory of emotion. *Psychosomatic Medicine*, *11*(6), 338–353. https://doi.org/10.1097/00006842-194911000-00003

MacLeod, C., Grafton, B., & Notebaert, L. (2019). Anxiety-linked attentional bias: Is it reliable? *Annual Review of Clinical Psychology*, *15*, 529–554. https://doi.org/10.1146/annurev-clinpsy-050718-095505

Macpherson, F. (2011). Taxonomising the senses. *Philosophical Studies*, *153*(1), 123–142. https://doi.org/10.1007/s11098–010-9643-8

Mandler, G. (1975). *Mind and emotion*. Wiley

Mandler, G., Mandler, J. M., Kremen, I., & Sholiton, R. D. (1961). The response to threat: Relations among verbal and physiological indices. *Psychological Monographs: General and Applied*, *75*(9), 1–22. https://doi.org/10.1037/h0093803

Manicas, P. T., & Secord, P. F. (1983). Implications for psychology of the new philosophy of science. *American Psychologist*, *38*(4), 399–413. https://doi.org/10.1037/0003-066X.38.4.399

Manstead, A. S., & Fischer, A. H. (2001). Social appraisal: The social world as object of and influence on appraisal processes. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 221–232). Oxford University Press.

Marañon, G. (1924). Contribution à l'étude de l'action emotive de l'adrénaline [Contribution to the study of the emotive action of adrenalin]. *Revue Française d'Endocrinologie*, *2*, 301–325.

Marks, J. (1982). A theory of emotion. *Philosophical Studies*, *42*(2), 227–242. https://doi.org/10.1007/bf00374036

Marks, J. (1993). Reviewed work: *Varieties of affect* by Claire Armon-Jones. *Mind*, *102*(405), 177–179. https://doi.org/10.1093/mind/102.405.177

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*(2), 224–253. https://doi.org/10.1037/0033-295X.98.2.224

Marr, D. (1982). *Vision: A computational approach*. Freeman. https://doi.org/10.7551/mitpress/9780262514620.001.0001

Marshall, G. D., & Zimbardo, P. G. (1979). Affective consequences of inadequately explained physiological arousal. *Journal of Personality and Social Psychology*, *37*(6), 970–988. https://doi.org/10.1037/0022-3514.37.6.970

Martin, I., & Levey, A. B. (1978). Evaluative conditioning. *Advances in Behaviour Research and Therapy*, *1*(2), 57–101. https://doi.org/10.1016/0146-6402(78)90013-9

Mason, W. A., & Capitanio, J. P. (2012). Basic emotions: A reconstruction. *Emotion Review*, *4*(3), 238–244. https://doi.org/10.1177/1754073912439763

Matarazzo, O., Pizzini, B., & Greco, C. (2020). Influences of a luck game on offers in ultimatum and dictator games: Is there a mediation of emotions? *Frontiers in Psychology*, *11*, 13. https://doi.org/10.3389/fpsyg.2020.00013

Matsumoto, D., & Ekman, P. (2009). Basic emotions. In D. Sander & K. R. Scherer (Eds.), *Oxford companion to affective sciences* (pp. 69–72). Oxford University Press.

Matsumoto, D., Keltner, D., Shiota, M. N., O'Sullivan, M., & Frank, M. (2008). Facial expressions of emotion. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (pp. 211–234). Guilford.

Matsumoto, D., Kudoh, T., Scherer, K., & Wallbott, H. (1988). Antecedents of and reactions to emotions in the United States and Japan. *Journal of Cross-Cultural Psychology*, *19*(3), 267–286. https://doi.org/10.1177/0022022188193001

Matsumoto, D., & Willingham, B. (2009). Spontaneous facial expressions of emotion of congenitally and noncongenitally blind individuals. *Journal of Personality and Social Psychology*, *96*(1), 1–10. https://doi.org/10.1037/a0014037

Matsumoto, D., Yoo, S. H., Nakagawa, S., & Multinational Study of Cultural Display Rules (2008). Culture, emotion regulation, and adjustment. *Journal of Personality and Social Psychology*, *94*(6), 925–937. https://doi.org/10.1037/0022-3514.94.6.925

Mauss, I. B., Bunge, S. A., & Gross, J. J. (2007). Automatic emotion regulation. *Social and Personality Psychology Compass*, *1*(1), 146–167. https://doi.org/10.1111/j.1751-9004.2007.00005.x

Mauss, I. B., Cook, C. L., & Gross, J. J. (2007). Automatic emotion regulation during anger provocation. *Journal of Experimental Social Psychology*, *43*(5), 698–711. https://doi.org/10.1016/j.ijpsycho.2007.03.017

Mauss, I. B., Levenson, R. W., McCarter, L., Wilhelm, F. H., & Gross, J. J. (2005). The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion*, *5*(2), 175–190. https://doi.org/10.1037/1528-3542.5.2.175

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion*, *23*(2), 209–237. https://doi.org/10.1080/02699930802204677

Mayr, E. (1974). Behavior programs and evolutionary strategies: Natural selection sometimes favors a genetically "closed" behavior program, sometimes an "open" one. *American Scientist*, *62*(6), 650–659.

McCauley, R. N. (1996). Explanatory pluralism and the coevolution of theories in science. In R. N. McCauley (Ed.), *The Churchlands and their critics* (pp. 17–47). Blackwell.

McCauley, R. N., & Bechtel, W. (2001). Explanatory pluralism and heuristic identity theory. *Theory & Psychology*, *11*(6), 736–760. https://doi.org/10.1177/0959354301116002

McCloskey, M. S., Berman, M. E., & Coccaro, E. F. (2005). Providing an escape option reduces retaliatory aggression. *Aggressive Behavior*, *31*(3), 228–237. https://doi.org/10.1002/ab.20084

McDougall, W. (1908). *An introduction to social psychology*. John Luce. https://doi.org/10.1037/12261-000

McNally, R. J. (2019). Attentional bias for threat: Crisis or opportunity? *Clinical Psychology Review*, *69*, 4–13. https://doi.org/10.1016/j.cpr.2018.05.005

Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, *1*(2), 108–141. https://doi.org/10.1207/s15327965pli0102_1

Mehu, M., & Scherer, K. R. (2015). Emotion categories and dimensions in the facial communication of affect: An integrated approach. *Emotion*, *15*(6), 798–811. https://doi.org/10.1037/a0039416

Meinong, A. (1894). *Psychologisch-ethische Untersuchungen zur Werttheorie* [*Psychological-ethical investigations concerning the theory of value*]. Leuschner & Lubensky.

Melnikoff, D. E., & Bargh, J. A. (2018a). The insidious number two. *Trends in Cognitive Sciences*, *22*(8), 668–669. https://doi.org/10.1016/j.tics.2018.05.005

Melnikoff, D. E., & Bargh, J. A. (2018b). The mythical number two. *Trends in Cognitive Sciences*, *22*(4), 280–293. https://doi.org/10.1016/j.tics.2018.02.001

Mesquita, B., & Boiger, M. (2014). Emotions in context: A sociodynamic model of emotions. *Emotion Review*, *6*(4), 298–302. https://doi.org/10.1177/1754073914534480

Mesquita, B., Boiger, M., & De Leersnyder, J. (2017). Doing emotions: The role of culture in everyday emotions. *European Review of Social Psychology*, *28*(1), 95–133. https://doi.org/10.1080/10463283.2017.1329107

Mesquita, B., De Leersnyder, J., & Albert, D. (2014). The cultural regulation of emotions. In J. J. Gross (Ed.), *Handbook of emotion regulation* (2nd ed., pp. 284–301). Guilford.

Mesquita, B., & Ellsworth, P. C. (2001). The role of culture in appraisal. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 233–248). Oxford University Press.

Mesquita, B., & Parkinson, B. (2022). Social constructionism. In A. Scarantino (Ed.), *Routledge handbook of emotion theory*. Routledge.

Meyer, M. (1933). That whale among the fishes – The theory of emotions. *Psychological Review*, *40*(3), 292–300. https://doi.org/10.1037/h0071608

Miceli, M., & Castelfranchi, C. (2015). *Expectancy and emotion*. Oxford University Press.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. Holt, Rinehart & Winston. https://doi.org/10.1037/10039-000

Millikan, R. G. (1993). Content and vehicle. In N. Eilan, R. McCarthy, & B. Brewer (Eds.), *Spatial representation* (pp. 256–268). Blackwell.

Millikan, R. G. (1995). Pushmi-pullyu representations. *Philosophical Perspectives*, *9*, 185–200. https://doi.org/10.2307/2214217

Mitchell, J. (2020a). The attitudinal opacity of emotional experience. *Philosophical Quarterly*, *70*(280), 524–546. https://doi.org/10.1093/pq/pqz085

Mitchell, J. (2020b). The bodily-attitudinal theory of emotion. *Philosophical Studies*, *178*, 2635–2663. https://doi.org/10.1007/s11098–020-01567-z

Mobbs, D., Adolphs, R., Fanselow, M. S., Barrett, L. F., LeDoux, J. E., Ressler, K., & Tye, K. M. (2019). Viewpoints: Approaches to defining and investigating fear. *Nature Neuroscience*, *22*(8), 1205–1216. https://doi.org/10.1038/s41593–019-0456-6

Mogg, K., Bradley, B. P., & Williams, R. (1995). Attentional bias in anxiety and depression: The role of awareness. *British Journal of Clinical Psychology*, *34*(1), 17–36. https://doi.org/10.1111/j.2044-8260.1995.tb01434.x

Mohanty, A., Gitelman, D. R., Small, D. M., & Mesulam, M. M. (2008). The spatial attention network interacts with limbic and monoaminergic systems to modulate motivation-induced attention shifts. *Cerebral Cortex*, *18*(11), 2604–2613. https://doi.org/10.1093/cercor/bhn021

Montagrin, A., Brosch, T., & Sander, D. (2013). Goal conduciveness as a key determinant of memory facilitation. *Emotion*, *13*(4), 622–628. https://doi.org/10.1037/a0033066

Montagrin, A., Sterpenich, V., Brosch, T., Grandjean, D., Armony, J., Ceravolo, L., & Sander, D. (2018). Goal-relevant situations facilitate memory of neutral faces. *Cognitive, Affective, & Behavioral Neuroscience*, *18*(6), 1269–1282. https://doi.org/10.3758/s13415–018-0637-x

Moors, A. (2007). Can cognitive methods be used to study the unique aspect of emotion: An appraisal theorist's answer. *Cognition and Emotion*, *21*(6), 1238–1269. https://doi.org/10.1080/02699930701438061

Moors, A. (2009). Theories of emotion causation: A review. *Cognition and Emotion*, *23*(4), 625–662. https://doi.org/10.1080/02699930802645739

Moors, A. (2010a). Automatic constructive appraisal: A reply to the commentaries of Parkinson and Kuppens. *Emotion Review*, *2*(2), 161–162. https://doi.org/10.1177/1754073909355012

Moors, A. (2010b). Automatic constructive appraisal as a candidate cause of emotion. *Emotion Review*, *2*(2), 139–156. https://doi.org/10.1177/1754073909351755

Moors, A. (2012). Comparison of affect program theories, appraisal theories, and psychological construction theories. In P. Zachar & R. D. Ellis (Eds.), *Categorical versus dimensional models of affect: A seminar on the theories of Panksepp and Russell* (pp. 257–278). John Benjamins. https://doi.org/10.1075/ceb.7.13moo

Moors, A. (2013a). Author reply: Appraisal is transactional, not all-inclusive, and cognitive in a broad sense. *Emotion Review*, *5*(2), 185–186. https://doi.org/10.1177/1754073912468300

Moors, A. (2013b). On the causal role of appraisal in emotion. *Emotion Review*, *5*(2), 132–140. https://doi.org/10.1177/1754073912463601

Moors, A. (2013c). Understanding emotion change requires an understanding of emotion causation. In D. Hermans, B. Rimé, & B. Mesquita (Eds.), *Changing emotions* (pp. 144–150). Psychology Press. https://doi.org/10.4324/9780203075630-28

Moors, A. (2014a). Flavors of appraisal theories of emotion. *Emotion Review*, *6*(4), 303–307. https://doi.org/10.1177/1754073914534477

Moors, A. (2014b). Author reply: Toward a multilevel mechanistic explanation of complex regularities between environment and emotional components. *Emotion Review*, *6*(4), 328–330. https://doi.org/10.1177/1754073914534503

Moors, A. (2014c). Examining the mapping problem in dual process models. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind* (pp. 20–34). Guilford.

Moors, A. (2016). Automaticity: Componential, causal, and mechanistic explanations. *Annual Review of Psychology*, *67*, 263–287. https://doi.org/10.1146/annurev-psych-122414-033550

Moors, A. (2017a). Integration of two skeptical emotion theories: Dimensional appraisal theory and Russell's psychological construction theory. *Psychological Inquiry*, *28*(1), 1–19. https://doi.org/10.1080/1047840X.2017.1235900

Moors, A. (2017b). The integrated theory of emotional behavior follows a radically goal-directed approach. *Psychological Inquiry*, *28*(1), 68–75. https://doi.org/10.1080/1047840X.2017.1275207

Moors, A. (2019). *Towards a goal-directed account of weak-willed behavior*. PsychArchives. https://doi.org/10.23668/psycharchives.5690

Moors, A. (2022). An overview of theories of emotions in psychology. In A. Scarantino (Ed.), *Routledge handbook of emotion theory*. Routledge.

Moors, A., & Boddez, Y. (2017). Author reply: Emotional episodes are action episodes. *Emotion Review*, *9*(4), 353–354. https://doi.org/10.1177/1754073917692131

Moors, A., & Boddez, Y. (2021). The goal-directed model as an alternative to reductionist and network approaches of psychopathology. *Current Opinion in Psychology*, *41*, 84–87. https://doi.org/10.1016/j.copsyc.2021.03.014

Moors, A., Boddez, Y., & De Houwer, J. (2017). The power of goal-directed processes in the causation of emotional and other actions. *Emotion Review*, *9*(4), 310–318. https://doi.org/10.1177/1754073916669595

Moors, A., & De Houwer, J. (2001). Automatic appraisal of motivational valence: Motivational affective priming and Simon effects. *Cognition and Emotion*, *15*(6), 749–766. https://doi.org/10.1080/02699930143000293

Moors, A., & De Houwer, J. (2005). Automatic processing of dominance and submissiveness. *Experimental Psychology*, *52*(4), 296–302. https://doi.org/10.1027/1618-3169.52.4.296

Moors, A., & De Houwer, J. (2006a). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297–326. https://doi.org/10.1037/0033-2909.132.2.297

Moors, A., & De Houwer, J. (2006b). Problems with dividing the realm of processes. *Psychological Inquiry*, *17*(3), 199–204. www.jstor.org/stable/20447324

Moors, A., De Houwer, J., & Eelen, P. (2004). Automatic stimulus-goal comparisons: Support from motivational affective priming studies. *Cognition and Emotion*, *18*(1), 29–54. https://doi.org/10.1080/02699930244000462

Moors, A., De Houwer, J., Hermans, D., & Eelen, P. (2005). Unintentional processing of motivational valence. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *58*(6), 1043–1063. https://doi.org/10.1080/02724980443000467

Moors, A., Ellsworth, P., Scherer, K. R., & Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and future development. *Emotion Review*, *5*(2), 119–124. https://doi.org/10.1177/1754073912468165

Moors, A., Fini, C., Everaert, T., Bardi, L., Bossuyt, E., Kuppens, P., & Brass, M. (2019). The role of stimulus-driven versus goal-directed processes in fight and flight tendencies measured with motor evoked potentials induced by transcranial magnetic stimulation. *PLoS ONE*, *14*(5), e0217266, 1–22. https://doi.org/10.1371/journal.pone.0217266

Moors, A., & Fischer, M. (2019). Demystifying the role of emotion in behaviour: Toward a goal-directed account. *Cognition and Emotion*, *33*(1), 94–100. https://doi.org/10.1080/02699931.2018.1510381

Moors, A., & Scherer, K. R. (2013). The role of appraisal in emotion. In M. Robinson, E. Watkins, & E. Harmon-Jones (Eds.), *Handbook of cognition and emotion* (pp. 135–155). Guilford.

Moors, A., Van de Cruys, S., & Pourtois, G. (2021). Comparison of the determinants for positive and negative affect proposed by appraisal theories, goal-directed theories, and predictive processing theories. *Current Opinion in Behavioral Sciences*, *39*, 147–152. https://doi.org/10.1016/j.cobeha.2021.03.015

Morina, N., Ijntema, H., Meyerbröker, K., & Emmelkamp, P. M. (2015). Can virtual reality exposure therapy gains be generalized to real-life? A meta-analysis of studies applying behavioral assessments. *Behaviour Research and Therapy*, *74*, 18–24. https://doi.org/10.1016/j.brat.2015.08.010

Morris, R. W., Dezfouli, A., Griffiths, K. R., & Balleine, B. W. (2014). Action-value comparisons in the dorsolateral prefrontal cortex control choice between goal-directed actions. *Nature Communications*, *5*(1), 1–10. https://doi.org/10.1038/ncomms5390

Morsbach, H., & Tyler, W. J. (1986). A Japanese emotion, Amae. In R. Harré (Ed.), *The social construction of emotions* (pp. 289–307). Blackwell.

Motro, D., Ordóñez, L. D., Pittarello, A., & Welsh, D. T. (2018). Investigating the effects of anger and guilt on unethical behavior: A dual-process approach. *Journal of Business Ethics*, *152*(1), 133–148. https://doi.org/10.1007/s10551–016-3337-x

Mowrer, O. H. (1960). *Learning theory and behavior*. Wiley. https://doi.org/10.1037/10802-000

Müller, J. M. (2018). Emotion as position-taking. *Philosophia*, *46*(3), 525–540. https://doi.org/10.1007/s11406–017-9905-1

Mulligan, K., & Scherer, K. R. (2012). Toward a working definition of emotion. *Emotion Review*, *4*(4), 345–357. https://doi.org/10.1177/1754073912445818

Mumenthaler, C., & Sander, D. (2012). Social appraisal influences recognition of emotions. *Journal of Personality and Social Psychology*, *102*(6), 1118–1135. https://doi.org/10.1037/a0026885

Mumenthaler, C., & Sander, D. (2015). Automatic integration of social information in emotion recognition. *Journal of Experimental Psychology: General*, *144*(2), 392–399. https://doi.org/10.1037/xge0000059

Mundale, J. (1998). Brain mapping. In W. Bechtel & G. Graham (Eds.), *A companion to cognitive science* (pp. 121–128). Blackwell. https://doi.org/10.1002/9781405164535.ch4

Mundale, J., & Bechtel, W. (1996). Integrating neuroscience, psychology, and evolutionary biology through a teleological conception of function. *Minds and Machines*, *6*(4), 481–505. https://doi.org./10.1007/BF00389655

Murphy, F. C., Nimmo-Smith, I. A. N., & Lawrence, A. D. (2003). Functional neuroanatomy of emotions: A meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience*, *3*(3), 207–233. https://doi.org/10.3758/cabn.3.3.207

Murphy, S. T., & Zajonc, R. B. (1993). Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology*, *64*(5), 723–739. https://doi.org/10.1037/0022-3514.64.5.723

Mylopoulos, M., & Pacherie, E. (2017). Intentions and motor representations: The interface challenge. *Review of Philosophy and Psychology*, *8*(2), 317–336. https://doi.org/10.1007/s13164–016-0311-6

Naar, H. (2018). Sentiments. In H. Naar & F. Teroni (Eds.), *The ontology of emotions* (pp. 149–168). Cambridge University Press. https://doi.org/10.1017/9781316275221.009

Nagel, T. (1994). What is it like to be a bat? *Philosophical Review*, *83*(4), 434–450. https://doi.org/10.2307/2183914

Natsoulas, T. (1981). Basic problems of consciousness. *Journal of Personality and Social Psychology*, *41*(1), 132–178. https://doi.org/10.1037/0022-3514.41.

Needham, A., & Baillargeon, R. (1993). Intuitions about support in 4.5-month-old infants. *Cognition*, *47*(2), 121–148. https://doi.org/10.1016/0010-0277(93)90002-d

Nelson, N. L., & Russell, J. A. (2013). Universality revisited. *Emotion Review*, *5*(1), 8–15. https://doi.org/10.1177/1754073912457227

Neu, J. (2000). *A tear is an intellectual thing: The meanings of emotion*. Oxford University Press. https://doi.org/10.1086/339591

Neumann, R. (2000). The causal influences of attributions on emotions: A procedural priming approach. *Psychological Science*, *11*(3), 179–182. https://doi.org/10.1111/1467-9280.00238

Neumann, R., Förster, J., & Strack, F. (2003). Motor compatibility: The bidirectional link between behavior and evaluation. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 371–391). Erlbaum. https://doi.org/10.4324/9781410606853-22

Neumann, R., & Strack, F. (2000). Mood contagion: The automatic transfer of mood between persons. *Journal of Personality and Social Psychology*, *79*(2), 211–223. https://doi.org/10.1037/0022-3514.79.2.211

Newell, B. R., & Shanks, D. R. (2014). Unconscious influences on decision making: A critical review. *Behavioral and Brain Sciences*, *37*(1), 1–19. https://doi.org/10.1017/s0140525x12003214

Newman, I. R., Gibb, M., & Thompson, V. A. (2017). Rule-based reasoning is fast and belief-based reasoning can be slow: Challenging current explanations of belief-bias and base-rate neglect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(7), 1154–1170. https://doi.org/10.1037/xlm0000372

Niedenthal, P. M., & Brauer, M. (2012). Social functionality of human emotion. *Annual Review of Psychology*, *63*, 259–285. https://doi.org/10.1146/annurev.psych.121208.131605

Niedenthal, P. M., Krauth-Gruber, S., & Ric, F. (2006). *Psychology of emotion: Interpersonal, experiential, and cognitive approaches*. Psychology Press.

Niedenthal, P. M., Mermillod, M., Maringer, M., & Hess, U. (2010). The simulation of smiles (SIMS) model: Embodied simulation and the meaning of facial expression. *Behavioral and Brain Sciences*, *33*(6), 417–433. https://doi.org/10.1017/s0140525x10000865

Nisbett, R. E., & Schachter, S. (1966). Cognitive manipulation of pain. *Journal of Experimental Social Psychology*, *2*(3), 227–236. https://doi.org/10.1016/0022-1031(66)90081-3

Nummenmaa, L., Glerean, E., Hari, R., & Hietanen, J. K. (2014). Bodily maps of emotions. *Proceedings of the National Academy of Sciences*, *111*(2), 646–651. https://doi.org/10.1073/pnas.1321664111

Nummenmaa, L., & Niemi, P. (2004). Inducing affective states with success-failure manipulations: A meta-analysis. *Emotion*, *4*(2), 207–214. https://doi.org/10.1037/1528-3542.4.2.207

Nummenmaa, L., & Saarimäki, H. (2019). Emotions as discrete patterns of systemic activity. *Neuroscience Letters*, *693*, 3–8. https://doi.org/10.1016/j.neulet.2017.07.012

Nussbaum, M. C. (2001). *Upheavals of thought: The intelligence of emotions*. Cambridge University Press. https://doi.org/10.1017/CBO9780511840715

Nussbaum, M. C. (2004). Emotions as judgments of value and importance. In R. C. Solomon (Ed.), *Thinking about emotion: Contemporary philosophers on emotion* (pp. 183–199). Oxford University Press.

Oakes, P., & Turner, J. C. (1986). Distinctiveness and the salience of social category memberships: Is there an automatic perceptual bias towards novelty? *European Journal of Social Psychology*, *16*(4), 325–344. https://doi.org/10.1002/ejsp.2420160403

Oatley, K. (1992). *Best laid schemes: The psychology of the emotions*. Cambridge University Press. https://doi.org/10.2307/2075681

Oatley, K., & Johnson-Laird, P. N. (1987). Towards a cognitive theory of emotions. *Cognition and Emotion*, *1*(1), 29–50. https://doi.org/10.1080/02699938708408362

Oatley, K., & Johnson-Laird, P. N. (2011). Basic emotions in social relationships, reasoning, and psychological illnesses. *Emotion Review*, *3*(4), 424–433. https://doi.org/10.1177/1754073911410738

Oddie, G. (2005). *Value, reality and desire*. Oxford University Press. https://doi.org/10.1093/0199273413.001.0001

Öhman, A. (1986). Face the beast and fear the face: Animal and social fears as prototypes for evolutionary analyses of emotion. *Psychophysiology*, *23*(2), 123–145. https://doi.org/10.1111/j.1469-8986.1986.tb00608.x

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, *108*(3), 483–522. https://doi.org/10.1037/0033-295X.108.3.483

Oppenheim, P., & Putnam, H. (1958). Unity of science as a working hypothesis. In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Minnesota studies in the philosophy of science II* (pp. 3–36). University of Minnesota Press.

Orilia F., & Paolini Paoletti, M. (2020, Winter). Properties. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. https://plato.stanford.edu/archives/win2020/entries/properties/

Ortony, A. (2021). Are all "basic emotions" emotions? A problem for the (basic) emotions construct. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691620985415

Ortony, A., & Clore, G. L. (2015). Can appraisal models be compatible with constructionism? In L. F. Barrett & J. R. Russell (Eds.), *The psychological construction of emotion* (pp. 305–333). Guilford.

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge University Press. https://doi.org/10.1017/cbo9780511571299

Ortony, A., Clore, G. L., & Foss, M. A. (1987). The referential structure of the affective lexicon. *Cognitive Science*, *11*(3), 341–364. https://doi.org/10.1207/s15516709cog1103_4

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, *97*(3), 315–331. https://doi.org/10.1037/0033-295X.97.3.315

Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review*, *11*(6), 988–1010. https://doi.org/10.3758/BF03196730

Oudiette, D., Vinckier, F., Bioud, E., & Pessiglione, M. (2019). A Pavlovian account for paradoxical effects of motivation on controlling response vigour. *Scientific Reports*, *9*(1), 1–13. https://doi.org/10.1038/s41598–019-43936-7

Ousdal, O. T., Jensen, J., Server, A., Hariri, A. R., Nakstad, P. H., & Andreassen, O. A. (2008). The human amygdala is involved in general behavioral relevance detection: Evidence from an event-related functional magnetic resonance imaging Go-NoGo task. *Neuroscience*, *156*(3), 450–455. https://doi.org/10.1016/j.neuroscience.2008.07.066

Oyama, S. (2000). *The ontogeny of information: Developmental systems and evolution* (2nd ed.). Duke University Press. https://doi.org/10.2307/j.ctv1220mm5

Panksepp, J. (1982). Toward a general psychobiological theory of emotions. *Behavioral and Brain Sciences*, *5*(3), 407–468. https://doi.org/10.1017/s0140525x00012759

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. Oxford University Press.

Panksepp, J. (2000). The neuro-evolutionary cusp between emotions and cognitions. *Evolution and Emotion*, *7*(2), 141–163. https://doi.org/10.1075/ce.1.1.04pan

Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition*, *14*(1), 30–80. https://doi.org/10.1016/j.concog.2004.10.004

Panksepp, J. (2007). Neurologizing the psychology of affects: How appraisal-based constructivism and basic emotion theory can coexist. *Perspectives in Psychological Science*, *2*(3), 281–296. https://doi.org/10.1111/j.1745-6916.2007.00045.x

Panksepp, J. (2012). In defense of multiple core affects. In P. Zachar & R. D. Ellis (Eds.), *Categorical versus dimensional models of affect: A seminar on the theories of Panksepp and Russell* (Vol. 7, pp. 31–78). John Benjamins. https://doi.org/10.1075/ceb.7.02pan

Panksepp, J., & Watt, D. (2011). What is basic about basic emotions? Lasting lessons from affective neuroscience. *Emotion Review*, *3*(4), 387–396. https://doi.org/10.1177/1754073911410741

Parkinson, B. (1985). Emotional effects of false autonomic feedback. *Psychological Bulletin*, *98*(3), 471–494. https://doi.org./10.1037/0033-2909.98.3.471

Parkinson, B. (1995). *Ideas and realities of emotions*. Routledge. https://doi.org/10.4324/9780203135648

Parkinson, B. (1996). Emotions are social. *British Journal of Psychology*, *87*(4), 663–683. https://doi.org/10.1111/j.2044-8295.1996.tb02615.x

Parkinson, B. (1997). Untangling the appraisal-emotion connection. *Personality and Social Psychology Review*, *1*(1), 62–79. https://doi.org/10.1207/s15327957pspr0101_5

Parkinson, B. (1999). Relations and dissociations between appraisal and emotion ratings of reasonable and unreasonable anger and guilt. *Cognition and Emotion*, *13*(4), 347–385. https://doi.org/10.1080/026999399379221

Parkinson, B. (2001a). Anger on and off the road. *British Journal of Psychology*, *92*(3), 507–526. https://doi.org/10.1348/000712601162310

Parkinson, B. (2001b). Putting appraisal in context. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 173–186). Oxford University Press.

Parkinson, B. (2005). Do facial movements express emotions or communicate motives? *Personality and Social Psychology Review*, *9*(4), 278–311. https://doi.org/10.1207/s15327957pspr0904_1

Parkinson, B. (2007). Getting from situations to emotions: Appraisal and other routes. *Emotion*, *7*(1), 21–25. https://doi.org/10.1037/1528-3542.7.1.21q

Parkinson, B. (2008). Emotions in direct and remote social interaction: Getting through the spaces between us. *Computers in Human Behavior*, *24*, 1510–1529. https://doi.org/10.1016/j.chb.2007.05.006

Parkinson, B. (2011). Interpersonal emotion transfer: Contagion and social appraisal. *Social and Personality Psychology Compass*, *5*(7), 428–439. https://doi.org/10.1207/s15327957pspr0904_1

Parkinson, B. (2012). Piecing together emotion: Sites and time-scales for social construction. *Emotion Review*, *4*(3), 291–298. https://doi.org/10.1177/1754073912439764

Parkinson, B. (2013). Contextualizing facial activity. *Emotion Review*, *5*(1), 97–103. https://doi.org/10.1177/1754073912457230

Parkinson, B. (2017a). Comment: Respecifying emotional influence. *Emotion Review*, *9*(3), 263–265. https://doi.org/10.1177/1754073916669597

Parkinson, B. (2017b). Emotion components and social relations. *Psychological Inquiry*, *28*(1), 45–49. https://doi.org/10.1080/1047840x.2017.1255493

Parkinson, B. (2019a). Calibrating emotional orientations: Social appraisal and other kinds of relation alignment. In D. Dukes & F. Clément (Eds.), *Foundations of affective social learning: Conceptualizing the social transmission of value* (pp. 117–141). Cambridge University Press. https://doi.org/10.1017/9781108661362.006

Parkinson, B. (2019b). Social referencing in adults and children: Extending social appraisal approaches. In U. Hess & S. Hareli (Eds.), *The social nature of emotion expression: What emotions can tell us about the world* (pp. 119–140). Springer. https://doi.org/10.1007/978-3-030-32968-6_7

Parkinson, B. (2020). Intragroup emotion convergence: Beyond contagion and social appraisal. *Personality and Social Psychology Review*, *24*(2), 121–140. https://doi.org/10.1177/1088868319882596

Parkinson, B., & Manstead, A. S. (2015). Current emotion research in social psychology: Thinking about emotions and other people. *Emotion Review*, *7*(4), 371–380. https://doi.org/10.1177/1754073915590624

Parkinson, B., & Simons, G. (2012). Worry spreads: Interpersonal transfer of problem-related anxiety. *Cognition and Emotion*, *26*(3), 462–479. https://doi.org/10.1080/02699931.2011.651101

Parkinson, B., & Totterdell, P. (1999). Classifying affect-regulation strategies. *Cognition and Emotion*, *13*(3), 277–303. https://doi.org/10.1080/026999399379285

Parr, L. A., Waller, B. M., Vick, S. J., & Bard, K. A. (2007). Classifying chimpanzee facial expressions using muscle action. *Emotion*, *7*(1), 172–181. https://doi.org/10.1037/1528-3542.7.1.172

Parrott, G. W. (2007). Components and the definition of emotion. *Social Science Information*, *46*, 419–423. https://doi.org/10.1177/05390184070460030109

Pecchinenda, A. (2001). The psychophysiology of appraisals. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 301–315). Oxford University Press.

Pedersen, W. C., Aviles, F. E., Ito, T. A., Miller, N., & Pollock, V. E. (2002). Psychological experimentation on alcohol-induced human aggression. *Aggression and Violent Behavior*, *7*(3), 293–312. https://doi.org/10.1016/s1359–1789(01)00044-1

Peirce, C. S. (1940). Logic as semiotic: The theory of signs. In J. Buchler (Ed.), *The philosophy of Peirce: Selected writings* (pp. 98–119). Routledge. https://doi.org/10.4324/9781315822723

Pessiglione, M., Petrovic, P., Daunizeau, J., Palminteri, S., Dolan, J. D., & Frith, C. D. (2008). Subliminal instrumental conditioning demonstrated in the human brain. *Neuron*, *59*(4), 561–567. https://doi.org/10.1016/j.neuron.2008.07.005

Pessiglione, M., Schmidt, L., Draganski, B., Kalisch, R., Lau, H., Dolan, R. J., & Frith, C. D. (2007). How the brain translates money into force: A neuroimaging study of subliminal motivation. *Science*, *316*(5826), 904–906. https://doi.org/10.1126/science.1140459

Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: From a "low road" to "many roads" of evaluating biological significance. *Nature Reviews Neuroscience*, *11*(11), 773–782. https://doi.org/10.1038/nrn2920

Pezzulo, G., & Cisek, P. (2016). Navigating the affordance landscape: Feedback control as a process model of behavior and cognition. *Trends in Cognitive Sciences*, *20*(6), 414–424. https://doi.org/10.1016/j.tics.2016.03.013

Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage*, *16*(2), 331–348. https://doi.org/10.1006/nimg.2002.1087

Philippot, P., Chapelle, G., & Blairy, S. (2002). Respiratory feedback in the generation of emotion. *Cognition and Emotion*, *16*(5), 605–627. https://doi.org/10.1080/02699930143000392

Piccinini, G., & Craver, C. (2011). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. *Synthese*, *183*(3), 283–311. https://doi.org/10.1007/s11229–011-9898-4

Pillaud, N., & Ric, F. (2020). The impact of automatic evaluation on mood: An awareness-dependent effect. *Cognition and Emotion*, *34*(7), 1457–1472. https://doi.org/10.1080/02699931.2020.1775555

Pillutla, M. M., & Murnighan, J. K. (1996). Unfairness, anger, and spite: Emotional rejections of ultimatum offers. *Organizational Behavior and Human Decision Processes*, *68*(3), 208–224. https://doi.org/10.1006/obhd.1996.0100

Pineda, D. (2015a). Are emotions valent embodied appraisals? www.terapiacognitiva.eu/cpc/dwl/embodied/02_david_pineda.pdf

Pineda, D. (2015b). Emotions, appraisals, and embodied appraisals. *Crítica (México, DF)*, *47*(140), 3–30. https://doi.org/10.22201/iifs.18704905e.2015.494

Pitcher, G. (1965). Emotion. *Mind*, *74*(295), 326–346. https://doi.org/10.1093/mind/lxxiv.295.326

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research and experience: Vol. 1. Theories of emotion* (pp. 3–33). Academic Press. https://doi.org/10.1016/b978–0-12-558701-3.50007-7

Plutchik, R. (1984). Emotions: A general psychoevolutionary theory. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 197–219). Erlbaum. https://doi.org/10.4324/9781315798806

Plutchik, R. (2001). The nature of emotions. *American Scientist*, *89*, 344–350. https://doi.org/10.1511/2001.4.344

Pober, J. (2018). What emotions really are (in the theory of constructed emotions). *Philosophy of Science*, *85*(4), 640–659. https://doi.org/10.1086/699158

Pool, E., Brosch, T., Delplanque, S., & Sander, D. (2016). Attentional bias for positive emotional stimuli: A meta-analytic investigation. *Psychological Bulletin*, *142*(1), 79–106. https://doi.org/10.1037/bul0000026

Port, R. F., & Van Gelder, T. (Eds.). (1995). *Mind as motion: Explorations in the dynamics of cognition*. MIT Press. https://doi.org/10.7551/mitpress/4622.001.0001

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3, Pt.1), 353–363. https://doi.org/10.1037/h0025953

Power, M. J., & Dalgleish, T. (1997). *Cognition and emotion: From order to disorder*. Psychology Press. https://doi.org/10.4324/9781315708744

Powers, W.T. (1973a). *Behavior: The control of perception*. Aldine.

Powers, W. T. (1973b). Feedback: Beyond behaviorism: Stimulus-response laws are wholly predictable within a control-system model of behavioral organization. *Science*, *179*(4071), 351–356. https://doi.org/10.1126/science.179.4071.351

Price, C. (2006). Affect without object: Moods and objectless emotions. *European Journal of Analytic Philosophy*, *2*(1), 49–68.

Prinz, J. J. (2004a). *Gut reactions: A perceptual theory of emotion*. Oxford University Press.

Prinz, J. J. (2004b). Which emotions are basic? In D. Evans & P. Cruse (Eds.), *Emotion, evolution, and rationality* (pp. 69–88). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198528975.003.0004

Prinz, J. J. (2006a). Is emotion a form of perception? *Canadian Journal of Philosophy Supplementary Volume*, *32*, 136–160. https://doi.org/10.1353/cjp.2007.0035

Prinz, J. J. (2006b). Putting the brakes on enactive perception. *Psyche*, *12*(1), 1–19.

Prinz, J. J. (2010). For valence. *Emotion Review*, *2*(1), 5–13. https://doi.org/10.1177/1754073909345546

Prinz, J. J., & Barsalou, L. W. (2000). Steering a course for embodied representation. In E. Dietrich & A. Markman (Eds.), *Cognitive dynamics: Conceptual change in humans and machines* (pp. 51–77). MIT Press.

Prinz, W. (1990). A common coding approach to perception and action. In O. Neumann & W. Prinz (Eds.), *Relationships between perception and action* (pp. 167–201). Springer-Verlag. https://doi.org/10.1007/978-3-642-75348-0_7

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, *9*(2), 129–154. https://doi.org/10.1080/713752551

Quartz, S. R. (2007). Reason, emotion, and decision-making: Risk and reward computation with feeling. *Trends in Cognitive Sciences, 13*(5), 209–215. https://doi.org/10.1016/j.tics.2009.02.003

Quigley, K. S., & Barrett, L. F. (2014). Is there consistency and specificity of autonomic changes during emotional episodes? Guidance from the Conceptual Act Theory and psychophysiology. *Biological Psychology*, *98*, 82–94. https://doi.org/10.1016/j.biopsycho.2013.12.013

Quigley, K. S., Barrett, L. F., & Weinstein, S. (2002). Cardiovascular patterns associated with threat and challenge appraisals: A within-subjects analysis. *Psychophysiology*, *39*(3), 292–302. https://doi.org/10.1017/S0048577201393046

Quigley, K. S., Lindquist, K. A., & Barrett, L. F. (2014). Inducing and measuring emotion and affect: Tips, tricks, and secrets. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 220–252). Cambridge University Press. https://doi.org/10.1017/cbo9780511996481.014

Quoidbach, J., Berry, E. V., Hansenne, M., & Mikolajczak, M. (2010). Positive emotion regulation and well-being: Comparing the impact of eight savoring and dampening strategies. *Personality and Individual Differences*, *49*(5), 368–373. https://doi.org/10.1016/j.paid.2010.03.048

Railton, P. (2017). At the core of our capacity to act for a reason: The affective system and evaluative model-based learning and control. *Emotion Review*, *9*(4), 335–342. https://doi.org/10.1177/1754073916670021

Raghunathan, R., & Pham, M. T. (1999). All negative moods are not created equal: Motivational influences of anxiety and sadness on decision making. *Organizational Behavior and Human Decision Process, 79*(1), 56–77. https://doi.org/10.1006/obhd.1999.2838

Ratcliffe, M. (2005). William James on emotion and intentionality. *International Journal of Philosophical Studies*, *13*(2), 179–202. https://doi.org/10.1080/09672550500080405

Reisenzein, R. (1983). The Schachter theory of emotion: Two decades later. *Psychological Bulletin*, *94*(2), 239–264. https://doi.org/10.1037/0033-2909.94.2.239

Reisenzein, R. (2000). Exploring the strength of association between the components of emotion syndromes: The case of surprise. *Cognition and Emotion*, *14*(1), 1–38. https://doi.org/10.1080/026999300378978

Reisenzein, R. (2009). Emotional experience in the computational belief–desire theory of emotion. *Emotion Review*, *1*(3), 214–222. https://doi.org/10.1177/1754073909103589

Reisenzein, R. (2012). What is an emotion in the belief-desire theory of emotion? In F. Paglieri, L. Tummolini, R. Falcone, & M. Miceli (Eds.), *The goals of cognition: Essays in honor of Cristiano Castelfranchi* (pp. 181–211). College Publications.

Reisenzein, R., Bördgen, S., Holtbernd, T., & Matz, D. (2006). Evidence for strong dissociation between emotion and facial displays: The case of surprise.

Journal of Personality and Social Psychology, *91*(2), 295–315. https://doi.org/10.1037/0022-3514.91.2.295

Reisenzein, R., & Junge, M. (2012). Language and emotion from the perspective of the computational belief-desire theory of emotion. In P. A. Wilson (Ed.), *Dynamicity in emotion concepts* (Vol. 27, pp. 37–59). Peter Lang. https://doi.org/10.3726/978-3-653-01466-2/3

Reisenzein, R., & Schönpflug, W. (1992). Stumpf's cognitive-evaluative theory of emotion. *American Psychologist*, *47*(1), 34–45. https://doi.org/10.1037/0003-066x.47.1.34

Reisenzein, R., & Stephan, A. (2014). More on James and the physical basis of emotion. *Emotion Review*, *6*(1), 35–46. https://doi.org/10.1177/1754073913501395

Reisenzein, R., Studtmann, M., & Horstmann, G. (2013). Coherence between emotion and facial expression: Evidence from laboratory experiments. *Emotion Review*, *5*(1), 16–23. https://doi.org/10.1177/1754073912457228

Richardson, D. R., Vandenberg, R. J., & Humphries, S. A. (1986). Effect of power to harm on retaliative aggression among males and females. *Journal of Research in Personality*, *20*(4), 402–419. https://doi.org/10.1016/0092-6566(86)90122-4

Rimé, B. (2009). Emotion elicits the social sharing of emotion: Theory and empirical review. *Emotion Review*, *1*(1), 60–85. https://doi.org/10.1177/1754073908097189

Riskind, J. H. (1984). They stoop to conquer: Guiding and self-regulatory functions of physical posture after success and failure. *Journal of Personality and Social Psychology*, *47*, 479–493. https://doi.org/10.1037/0022-3514.47.3.479

Riskind, J. H., & Gotay, C. C. (1982). Physical posture: Could it have regulatory or feedback effects on motivation and emotion? *Motivation and Emotion*, *6*, 273–298. https://doi.org/10.1007/bf00992249

Robbins, P., & Aydede, M. (2009). A short primer on situated cognition. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 3–10). Cambridge University Press. https://doi.org/10.1017/cbo9780511816826.001

Robbins, T. W., & Costa, R. M. (2017). Habits. *Current Biology*, *27*(22), R1200–R1206. https://doi.org/10.1016/j.cub.2017.09.060

Roberts, R. C. (1988). What an emotion is: A sketch. *Philosophical Review*, *97*(2), 183–209. https://doi.org/10.2307/2185261

Roberts, R. C. (2003). *Emotions: An essay in aid of moral psychology*. Cambridge University Press. https://doi.org/10.1017/cbo9780511610202

Roberts, R. C. (2013). *Emotions in the moral life*. Cambridge University Press. https://doi.org/10.1017/cbo9781139061469

Robinson, M. D., & Clore, G. L. (2001). Simulation, scenarios, and emotional appraisal: Testing the convergence of real and imagined reactions to emotional stimuli. *Personality and Social Psychology Bulletin*, *27*(11), 1520–1532. https://doi.org/10.1177/01461672012711012

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, *128*(6), 934–960. https://doi.org/10.1037/0033-2909.128.6.934

Roch-Levecq, A. C. (2006). Production of basic emotions by children with congenital blindness: Evidence for the embodiment of theory of mind. *British Journal of Developmental Psychology*, *24*(3), 507–528. https://doi.org/10.1348/026151005x50663

Rodogno, R. (2016). Prudential value or well-being. In T. Brosch & D. Sander (Eds.), *Handbook of value: Perspectives from economics, psychology, and sociology* (pp. 287–311). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198716600.003.0014

Rolls, E. T. (1999). *The brain and emotion*. Oxford University Press.

Rolls, E. T. (2005). *Emotion explained*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198570035.001.000

Romero, F. (2015). Why there isn't inter-level causation in mechanisms. *Synthese*, *192*(11), 3731–3755. https://doi.org/10.1007/s11229–015-0718-0

Rorty, A. O. (1980). Explaining emotions. In A. O. Rorty (Ed.), *Explaining emotions* (pp. 103–126). University of California Press. https://doi.org/10.2307/2025425

Rorty, A. O. (1982). From passions to emotions and sentiments. *Philosophy, 57*(220), 159–172. https://doi.org/10.1017/S0031819100050749

Rosch, E. (1999). Principles of categorization. In E. Rosch & B. B. Loyd (Eds.), *Cognition and categorization* (pp. 27–48). Erlbaum. https://doi.org/10.1016/B978–1-4832-1446-7.50028-5

Roseman, I. J. (1984). Cognitive determinants of emotion: A structural theory. *Review of Personality & Social Psychology*, *5*, 11–36.

Roseman, I. J. (2001). A model of appraisal in the emotion system: Integrating theory, research, and applications. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 68–91). Oxford University Press.

Roseman, I. J. (2011). Emotional behaviors, emotivational goals, emotion strategies: Multiple levels of organization integrate variable and consistent responses. *Emotion Review*, *3*(4), 434–443. https://doi.org/10.1177/1754073911410744

Roseman, I. J. (2013). Appraisal in the emotion system: Coherence in strategies for coping. *Emotion Review*, *5*(2), 141–149. https://doi.org/10.1177/1754073912469591

Roseman, I. J., Dhawan, N., Rettek, S. I., Naidu, R. K., & Thapa, K. (1995). Cultural differences and cross-cultural similarities in appraisals and emotional responses. *Journal of Cross-Cultural Psychology*, *26*(1), 23–48. https://doi.org/10.1177/0022022195261003

Roseman, I. J., & Evdokas, A. (2004). Appraisals cause experienced emotions: Experimental evidence. *Cognition and Emotion*, *18*(1), 1–28. https://doi.org/10.1080/02699930244000390

Roseman, I. J., & Smith, C. A. (2001). Appraisal theory: Overview, assumptions, varieties, controversies. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 3–34). Oxford University Press.

Rosen, J. B., & Donley, M. P. (2006). Animal studies of amygdala function in fear and uncertainty: Relevance to human research. *Biological Psychology*, *73*(1), 49–60. https://doi.org/10.1016/j.biopsycho.2006.01.007

Rosenberg, E. L., & Ekman, P. (1994). Coherence between expressive and experiential systems in emotion. *Cognition and Emotion*, *8*(3), 201–229. https://doi.org/10.1080/02699939408408938

Rossberg-Gempton, I., & Poole, G. D. (1993). The effect of open and closed postures on pleasant and unpleasant emotions. *The Arts in Psychotherapy*, *20* (1), 75–82. https://doi.org/10.1016/0197-4556(93)90034-Y

Rossi, M., & Tappolet, C. (2019). What kind of evaluative states are emotions? The attitudinal theory vs. the perceptual theory of emotions. *Canadian Journal of Philosophy*, *49*(4), 544–563. https://doi.org/10.1080/00455091.2018.1472516

Ruckmick, C. A. (1936). *The psychology of feeling and emotion*. McGraw-Hill. https://doi.org/10.1037/10770-000

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2). MIT Press. https://doi.org/10.7551/mitpress/5236.001.0001

Russell, J. A. (1979). Affective space is bipolar. *Journal of Personality and Social Psychology, 37*(3), 345–356. https://doi.org/10.1037/0022-3514.37.3.345

Russell, J. A. (1991). In defense of a prototype approach to emotion concepts. *Journal of Personality and Social Psychology*, *60*(1), 37–47. https://doi.org/10.1037/0022-3514.60.1.37

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, *115* (1), 102–141. https://doi.org/10.1037/0033-2909.115.1.102

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. https://doi.org/10.1037/0033-295x.110.1.145

Russell, J. A. (2005). Emotion in human consciousness is built on core affect. *Journal of Consciousness Studies*, *12*(8–9), 26–42.

Russell, J. A. (2006). Emotions are not modules. *Canadian Journal of Philosophy*, *36* (Suppl. 1), 53–71. https://doi.org/10.1353/cjp.2007.0037

Russell, J. A. (2009). Emotion, core affect, and psychological construction. *Cognition and Emotion*, *23*(7), 1259–1283. https://doi.org/10.1080/02699930902809375

Russell, J. A. (2012). From a psychological constructionist perspective. In P. Zachar & R. D. Ellis (Eds.), *Categorical versus dimensional models of affect: A seminar on the theories of Panksepp and Russell* (pp. 79–118). John Benjamins. https://doi.org/10.1075/ceb.7.03rus

Russell, J. A. (2014). My psychological constructionist perspective, with a focus on conscious affective experience. In L. F. Barrett & J. A. Russell (Eds.), *The psychological construction of emotion* (pp. 183–208). Guilford.

Russell, J. A. (2017). Toward a broader perspective on facial expressions: Moving on from basic emotion theory. In J. M. Fernández-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp. 93–105). Oxford

University Press. https://doi.org/10.1093/acprof:oso/9780190613501.003.0006

Russell, J. A., Bachorowski, J., & Fernández-Dols, J. M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology*, *54*, 329–349. https://doi.org/10.1146/annurev.psych.54.101601.145102

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*(5), 805–819. https://doi.org/10.1037/0022-3514.76.5.805

Ryle, G. (1949). *The concept of mind*. Hutchinson. https://doi.org/10.4324/9780203875858

Saarimäki, H., Ejtehadian, L. F., Glerean, E., Jääskeläinen, I. P., Vuilleumier, P., Sams, M., & Nummenmaa, L. (2018). Distributed affective space represents multiple emotion categories across the human brain. *Social Cognitive and Affective Neuroscience*, *13*(5), 471–482. https://doi.org/10.1093/scan/nsy018

Saarimäki, H., Glerean, E., Smirnov, D., Mynttinen, H., Jääskeläinen, I. P., Sams, M., & Nummenmaa, L. (2020). Classification of emotions based on functional connectivity patterns of the human brain. *bioRxiv*. https://doi.org/10.1101/2020.01.17.910869

Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I. P., Lampinen, J., Vuilleumier, P., Hari, R., Sams, M., & Nummenmaa, L. (2016). Discrete neural signatures of basic emotions. *Cerebral Cortex*, *26*(6), 2563–2573. https://doi.org/10.1093/cercor/bhv086

Salmela, M. (2008). How to evaluate the factual basis of emotional appraisals? In L. C. Charland & P. Zachar (Eds.), *Facts and value in emotion* (pp. 35–51). John Benjamins. https://doi.org/10.1075/ceb.4.03sal

Salmela, M. (2011). Can emotion be modelled on perception? *Dialectica*, *65*(1), 1–29. https://doi.org/10.1111/j.1746-8361.2011.01259.x

Salmela, M., & Nagatsu, M. (2016). Collective emotions and joint action: Beyond received and minimalist approaches. *Journal of Social Ontology*, *2*(1), 33–57. https://doi.org/10.1515/jso-2015-0020

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton University Press. https://doi.org/10.1515/9780691221489

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: An evolved system for relevance detection. *Reviews in the Neurosciences*, *14*(4), 303–316. https://doi.org/10.1515/revneuro.2003.14.4.303

Sander, D., Grandjean, D., & Scherer, K. R. (2018). An appraisal-driven componential approach to the emotional brain. *Emotion Review*, *10*(3), 219–231. https://doi.org/10.1177/1754073918765653

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, *300*(5626), 1755–1758. https://doi.org/10.1126/science.1082976

Sartre, J. P. (1939). *Esquisse d'une théorie des émotions* [Sketch of a theory of emotions]. Hermann.

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations.

Proceedings of the National Academy of Sciences, *107*(6), 2408–2412. https://doi.org/10.1073/pnas.0908239106
Sauter, D. A., LeGuen, O., & Haun, D. B. M. (2011). Categorical perception of emotional facial expressions does not require lexical categories. *Emotion*, *11*(6), 1479–1483. https://doi.org/10.1037/a0025336
Sauter, D. A., & Russell, J. A. (2022). What do nonverbal expressions tell us about emotion. In A. Scarantino (Ed.), *Routledge handbook of emotion theory*. Routledge.
Scarantino, A. (2003). Affordances explained. *Philosophy of Science*, *70*(5), 949–961. https://doi.org/10.1086/377380
Scarantino, A. (2010). Insights and blindspots of the cognitivist theory of emotions. *British Journal for the Philosophy of Science*, *61*(4), 729–768. https://doi.org/10.1093/bjps/axq011
Scarantino, A. (2012a). Functional specialization does not require a one-to-one mapping between brain regions and emotions. *Behavioral and Brain Sciences*, *35*(3), 161–162. https://doi.org/10.1017/s0140525x11001749
Scarantino, A. (2012b). How to define emotions scientifically. *Emotion Review*, *4*(4), 358–368. https://doi.org/10.1177/1754073912445810
Scarantino, A. (2014). The motivational theory of emotions. In J. D'Arms & D. Jacobson (Eds.), *Moral psychology and human agency: Philosophical essays on the science of ethics* (pp. 156–185). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198717812.003.0008
Scarantino, A. (2015). Basic emotions, psychological construction and the problem of variability. In J. Russell & L. Barrett (Eds.), *The psychological construction of emotion* (pp. 334–376). Guilford.
Scarantino, A. (2016). The philosophy of emotions and its impact on affective science. In M. Lewis, J. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (4th ed., pp. 3–65). Guilford.
Scarantino, A. (2017a). Do emotions cause actions, and if so how? *Emotion Review*, *9*(4), 326–334. https://doi.org/10.1177/1754073916679005
Scarantino, A. (2017b). How to do things with emotional expressions: The theory of affective pragmatics. *Psychological Inquiry*, *28*(2–3), 165–185. https://doi.org/10.1080/1047840x.2017.1328951
Scarantino, A. (2018). Are LeDoux's survival circuits basic emotions under a different name? *Current Opinion in Behavioral Sciences*, *24*, 75–82. https://doi.org/10.1016/j.cobeha.2018.06.001
Scarantino, A. (2019). Affective pragmatics extended: From natural to overt expressions of emotions. In U. Hess & S. Hareli (Eds.), *The social nature of emotion expression: What emotions can tell us about the world* (pp. 49–81). Springer. https://doi.org/10.1007/978-3-030-32968-6_4
Scarantino, A., & de Sousa, R. (2018, Winter). Emotion. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/win2018/entries/emotion/
Scarantino, A., & Griffiths, P. E. (2011). Don't give up on basic emotions. *Emotion Review*, *3*(4), 444–454. https://doi.org/10.1177/1754073911410745

Scarantino, A., & Nielsen, M. (2015). Voodoo dolls and angry lions: How emotions explain arational actions. *Philosophical Studies*, *172*(11), 2975–2998. https://doi.org/10.1007/s11098–015-0452-y

Schachter, S. (1964). The interaction of cognitive and physiological determinants of emotional state. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 1, pp. 49–80). Academic Press. https://doi.org/10.1016/S0065–2601(08)60048-9

Schachter, S., & Singer, J. E. (1962). Cognitive, social, and physiological determinants of an emotional state. *Psychological Review*, *69*(5), 379–399. https://doi.org/10.1037/h0046234

Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. In K. R. Scherer & P. E. Ekman (Eds.), *Approaches to emotion* (pp. 293–317). Erlbaum. https://doi.org/10.4324/9781315798806

Scherer, K. R. (1988). Criteria for emotion-antecedent appraisal: A review. In V. Hamilton, G. H. Bower, & N. H. Frijda (Eds.), *Cognitive perspectives on emotion and motivation* (pp. 89–126). Kluwer Academic Publishers. https://doi.org/10.1007/978-94-009-2792-6_4

Scherer, K. R. (1993a). Neuroscience projections to current debates in emotion psychology. *Cognition and Emotion, 7*(1), 1–41. https://doi.org/10.1080/02699939308409174

Scherer, K. R. (1993b). Studying the emotion-antecedent appraisal process: An expert system approach. *Cognition and Emotion, 7*(3–4), 325–355. https://doi.org/10.1080/02699939308409192

Scherer, K. R. (1994a). An emotion's occurrence depends on the relevance of an event to the organism's goal/need hierarchy. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 227–231). Oxford University Press.

Scherer, K. R. (1994b). Toward a concept of "modal emotions." In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 25–31). Oxford University Press.

Scherer, K. R. (1997a). Profiles of emotion-antecedent appraisals: Testing theoretical predictions across cultures. *Cognition and Emotion, 11*(2), 113–150. https://doi.org/10.1080/026999397379962

Scherer, K. R. (1997b). The role of culture in emotion-antecedent appraisal. *Journal of Personality and Social Psychology*, *73*(5), 902–922. https://doi.org/10.1037/0022-3514.73.5.902

Scherer, K. R. (2000). Emotions as episodes of subsystem synchronization driven by nonlinear appraisal processes. In M. D. Lewis & I. Granic (Eds.), *Emotion, development and selforganization: Dynamic systems approaches to emotional development* (pp. 70–99). Cambridge University Press. https://doi.org/10.1017/cbo9780511527883.005

Scherer, K. R. (2001a). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92–120). Oxford University Press.

Scherer, K. R. (2001b). The nature and study of appraisal: A review of the issues. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 369–391). Oxford University Press.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 695–729. https://doi.org/10.1177/0539018405058216

Scherer, K. R. (2009a). Emotions are emergent processes: They require a dynamic computational architecture. *Philosophical Transactions of the Royal Society: B, 364*, 3459–3474. https://doi.org/10.1098/rstb.2009.0141

Scherer, K. R. (2009b). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion, 23*(7), 1307–1351. https://doi.org/10.1080/02699930902928969

Scherer, K. R. (2022). Theory convergence in emotion science is timely and realistic. *Cognition and Emotion, 36*(2), 154–170. https://doi.org/10.1080/02699931.2021.1973378

Scherer, K. R., & Brosch, T. (2009). Culture-specific appraisal biases contribute to emotion dispositions. *European Journal of Personality*, *23*(3), 265–288. https://doi.org/10.1002/per.714

Scherer, K. R., & Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion, 7*(1), 113–130. https://doi.org/10.1037/1528-3542.7.1.113

Scherer, K. R., & Ellsworth, P. C. (2013). Author reply: The unbearable heaviness of feeling. *Emotion Review*, *5*(2), 189–191. https://doi.org/10.1177/1754073912468301

Scherer, K. R., & Fontaine, J. R. (2019). The semantic structure of emotion words across languages is consistent with componential appraisal models of emotion. *Cognition and Emotion*, *33*(4), 673–682. https://doi.org/10.1080/02699931.2018.1481369

Scherer, K. R., & Meuleman, B. (2013). Human emotion experiences can be predicted on theoretical grounds: Evidence from verbal labeling. *PLoS ONE*, *8*(3), e58166. https://doi.org/10.1371/journal.pone.0058166

Scherer, K. R., & Moors, A. (2019). The emotion process: Event appraisal and component differentiation. *Annual Review of Psychology*, *70*, 719–745. https://doi.org/10.1146/annurev-psych-122216-011854

Scherer, K. R., Mortillaro, M., & Mehu, M. (2013). Understanding the mechanisms underlying the production of facial expression of emotion: A componential perspective. *Emotion Review*, *5*(1), 47–53. https://doi.org/10.1177/1754073912451504

Scherer, K. R., Mortillaro, M., & Mehu, M. (2017). Facial expression is driven by appraisal and generates appraisal inference. In J. M. Fernández-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp. 353–73). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190613501.003.0019

Scherer, K. R., Mortillaro, M., Rotondi, I., Sergi, I., & Trznadel, S. (2018). Appraisal-driven facial actions as building blocks for emotion inference. *Journal of Personality and Social Psychology, 114*(3), 358–379. https://doi.org/10.1037/pspa0000107

Schnabel, K., Asendorpf, J. B., & Greenwald, A. G. (2008). Assessment of individual differences in implicit cognition: A review of IAT measures. *European Journal of Psychological Assessment*, *24*(4), 210–217. https://doi.org/10.1027/1015-5759.24.4.210

Schneider, K., & Josephs, I. (1991). The expressive and communicative functions of preschool children's smiles in an achievement-situation. *Journal of Nonverbal Behavior*, *15*(3), 185–198. https://doi.org/10.1007/BF01672220

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences, 4*(8), 299–309. https://doi.org/10.1016/s1364–6613(00)01506-0

Schwabe, L., & Wolf, O. T. (2011). Stress-induced modulation of instrumental behavior: From goal-directed to habitual control of action. *Behavioural Brain Research*, *219*(2), 321–328. https://doi.org/10.1016/j.bbr.2010.12.038

Schwager, S., & Rothermund, K. (2013). Counter-regulation triggered by emotions: Positive/negative affective states elicit opposite valence biases in affective processing. *Cognition and Emotion*, *27*(5), 839–855. https://doi.org/10.1080/02699931.2012.750599

Schwager, S., & Rothermund, K. (2014). On the dynamics of implicit emotion regulation: Counter-regulation after remembering events of high but not of low emotional intensity. *Cognition and Emotion*, *28*(6), 971–992. https://doi.org/10.1080/02699931.2013.866074

Schwartz, C. E., Wright, C. I., Shin, L. M., Kagan, J., Whalen, P. J., McMullin, K. G., & Rauch, S. L. (2003). Differential amygdalar response to novel versus newly familiar neutral faces: A functional MRI probe developed for studying inhibited temperament. *Biological Psychiatry*, *53*(10), 854–862. https://doi.org/10.1016/s0006–3223(02)01906-6

Schwarz, N., & Clore, J. L. (1996). Feelings and phenomenal experiences. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: A handbook of basic principles* (pp. 433–465). Guilford.

Schwarz, N., & Clore, G. L. (2003). Mood as information: 20 years later. *Psychological Inquiry*, *14*(3–4), 296–303. https://doi.org/10.1080/1047840x.2003.9682896

Schwarz, N., & Strack, F. (1991). Context effects in attitude surveys: Applying cognitive theory to social research. *European Review of Social Psychology*, *2*(1), 31–50. https://doi.org/10.1080/14792779143000015

Searle, J. R. (1958). Proper names. *Mind, 67*(266), 166–173. https://doi.org/10.1093/mind/lxvii.266.166

Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge University Press. https://doi.org/10.1017/cbo9781139173452

Seligman, M. E. (1968). Chronic fear produced by unpredictable electric shock. *Journal of Comparative and Physiological Psychology*, *66*(2), 402–411. https://doi.org./10.1037/h0026355

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17*(11), 565–573. https://doi.org/10.1016/j.tics.2013.09.007

Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology*, *2*, 395. https://doi.org/10.3389/fpsyg.2011.00395
Severo, M. C., Walentowska, W., Moors, A., & Pourtois, G. R. C. (2017). Goal impact influences the evaluative component of performance monitoring: Evidence from ERPs. *Biological Psychology*, *129*, 90–102. https://doi.org/10.1016/j.biopsycho.2017.08.052
Severo, M. C., Walentowska, W., Moors, A., & Pourtois, G. R. C. (2018). Goals matter: Amplification of the motivational significance of the feedback when goal impact is increased. *Brain and Cognition*, *128*, 56–72. https://doi.org/10.1016/j.bandc.2018.11.002
Shadmehr, R., & Ahmed, A. A. (2020). Precis of Vigor: Neuroeconomics of movement control. *Behavioral and Brain Sciences*, 1–10. https://doi.org/10.1017/s0140525x20000667
Shan, C., Gong, S., & McOwan, P. W. (2009). Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*, *27*(6), 803–816. https://doi.org/10.1016/j.imavis.2008.08.005
Shand, A. F. (1914). *The foundations of character: Being a study of the tendencies of the emotions and sentiments*. Macmillan. https://doi.org/10.1037/10666-000
Shargel, D. (2015). Emotions without objects. *Biology & Philosophy*, *30*(6), 831–844. https://doi.org/10.1007/s10539–014-9473-8
Shargel, D., & Prinz, J. J. (2018). An enactivist theory of emotional content. In H. Naar & F. Teroni (Eds.), *The ontology of emotions* (pp. 110–129). Cambridge University Press. https://doi.org/10.1017/9781316275221.007
Shariff, A. F., & Tracy, J. L. (2011a). Emotion expressions: On signals, symbols, and spandrels – A response to Barrett (2011). *Current Directions in Psychological Science*, *20*(6), 407–408. https://doi.org/10.1177/0963721411429126
Shariff, A. F., & Tracy, J. L. (2011b). What are emotion expressions for? *Current Directions in Psychological Science*, *20*(6), 395–399. https://doi.org/10.1177/0963721411424739
Shaver, P., Schwartz, J., Kirson, D., & O'connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology*, *52*(6), 1061–1086. https://doi.org/10.1037/0022-3514.52.6.1061
Sheeran, P., Webb, T. L., & Gollwitzer, P. M. (2005). The interplay between goal intentions and implementation intentions. *Personality and Social Psychology Bulletin*, *31*(1), 87–98. https://doi.org/10.1177/0146167204271308
Sherman, J. W. (2006). On building a better process model: It's not only how many, but which ones and by which means? *Psychological Inquiry*, *17*(3), 173–184. https://doi.org/10.1207/s15327965pli1703_3
Sherrington, C. S. (1900). Experiments on the value of vascular and visceral factors for the genesis of emotion. *Proceedings of the Royal Society of London*, *6*, 390–403. https://doi.org/10.1098/rspl.1899.0118
Shiota, M. N. (2022). Basic and discrete emotion theories. In A. Scarantino (Ed.), *Routledge handbook of emotion theory*. Routledge.

Shiota, M. N., Neufeld, S. L., Yeung, W. H., Moser, S. E., & Perea, E. F. (2011). Feeling good: Autonomic nervous system responding in five positive emotions. *Emotion, 11*(6), 1368–1378. https://doi.org/10.1037/a0024278

Siegel, E. H., Sands, M. K., Van den Noortgate, W., Condon, P., Chang, Y., Dy, J., Quigley, K. S., & Barrett, L. F. (2018). Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories. *Psychological Bulletin, 144*(4), 343–393. https://doi.org/10.1037/bul0000128

Siemer, M. (2008). Beyond prototypes and classical definitions: Evidence for a theory-based representation of emotion concepts. *Cognition and Emotion*, *22* (4), 620–632. https://doi.org/10.1080/02699930701440067

Siemer, M., Mauss, I., & Gross, J. J. (2007). Same situation – different emotions: How appraisals shape our emotions. *Emotion, 7*(3), 592–600. https://doi.org/10.1037/1528-3542.7.3.592

Silvia, P. J. (2005). What is interesting? Exploring the appraisal structure of interest. *Emotion, 5*(1), 89–102. https://doi.org/10.1037/1528–3542.5.1.89

Simon, H. A. (1983). *Reason in human affairs*. Stanford University Press.

Skinner, B. F. (1945). The operational analysis of psychological terms. *Psychological Review, 52*(5), 270–277. https://doi.org/10.1037/h0062535

Skinner, B. F. (1953). *Science and human behavior*. Simon & Schuster.

Skinner, B. F. (1974). *About behaviorism*. Knopf.

Skinner, B. F. (1977). Why I am not a cognitive psychologist. *Behaviorism*, *5*(2), 1–10.

Sloman, A., & Croucher, M. (1981). *You don't need to have a soft skin to have a warm heart: Towards a computational analysis of motives and emotions*. Cognitive Science Research Paper No. SCRP-004, Cognitive Studies Programme, School of Social Sciences, University of Sussex at Brighton, UK.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3–22. https://doi.org/10.1037/0033-2909.119.1.3

Smidt, K. E., & Suvak, M. K. (2015). A brief, but nuanced, review of emotional granularity and emotion differentiation research. *Current Opinion in Psychology*, *3*, 48–51. https://doi.org/10.1016/j.copsyc.2015.02.007

Smith, C. A. (1989). Dimensions of appraisal and physiological response in emotion. *Journal of Personality and Social Psychology, 56*(3), 339–353. https://doi.org/10.1037/0022-3514.56.3.339

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology, 48*(4), 813–838. https://doi.org/10.1037/0022–3514.48.4.813

Smith, C. A., & Kirby, L. D. (2000). Consequences require antecedents: Toward a process model of emotion elicitation. In J. P. Forgas (Ed.), *Feeling and thinking: The role of affect in social cognition* (pp. 83–106). Cambridge University Press. https://doi.org/10.5860/choice.37-5959

Smith, C. A., & Kirby, L. D. (2001). Toward delivering on the promise of appraisal theory. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 121–138). Oxford University Press.

Smith, C. A., & Kirby, L. D. (2009). Putting appraisal in context: Toward a relational model of appraisal and emotion. *Cognition and Emotion*, *23*(7), 1352–1372. https://doi.org/10.1080/02699930902860386

Smith, C. A., & Lazarus, R. S. (1993). Appraisal components, core relational themes, and the emotions. *Cognition and Emotion*, *7*(3–4), 233–269. https://doi.org/10.1080/02699939308409189

Smith, C. A., & Pope, L. K. (1992). Appraisal and emotion: The interactional contributions of dispositional and situational factors. In M. S. Clark (Ed.), *Emotion and social behavior* (pp. 32–62). Sage.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*(2), 108–131. https://doi.org/10.1207/s15327957pspr0402_01

Smith, E. R., & Neumann, R. (2005). Emotion processes considered from the perspective of dual-process models. In L. F. Barrett, P. M. Niedenthal, & P. Winkielman (Eds.), *Emotion and consciousness* (pp. 287–311). Guilford.

Smith, H. L. (2020). *Biting your tongue: Experimental effects of reducing anger expression* [Unpublished doctoral dissertation]. Florida State University.

Smith, J. (2014). Are emotions embodied evaluative attitudes? Critical review of Julien A. Deonna and Fabrice Teroni's "The emotions: A philosopical introduction." *Disputatio, 6*(38), 93–106. https://doi.org/10.2478/disp-2014-0005

Smith, M. (1987). The Humean theory of motivation. *Mind*, *96*(381), 36–61. www.jstor.org/stable/2253760

Smith, P. K., & Bargh, J. A. (2008). Nonconscious effects of power on basic approach and avoidance tendencies. *Social Cognition*, *26*(1), 1–24. https://doi.org/10.1521/soco.2008.26.1.1

Smith, R., Thayer, J. F., Khalsa, S. S., & Lane, R. D. (2017). The hierarchical basis of neurovisceral integration. *Neuroscience & Biobehavioral Reviews*, *75*, 274–296. https://doi.org/10.1016/j.neubiorev.2017.02.003

So, J., Achar, C., Han, D., Agrawal, N., Duhachek, A., & Maheswaran, D. (2015). The psychology of appraisal: Specific emotions and decision-making. *Journal of Consumer Psychology*, *25*(3), 359–371. https://doi.org/10.1016/j.jcps.2015.04.003

Solomon, R. C. (1973). Emotions and choice. *The Review of Metaphysics*, *27*(1), 20–41. https://doi.org/10.1093/0195145496.003.0001

Solomon, R. C. (1984). Getting angry: The Jamesian theory of emotion in anthropology. In R. A. Shweder & R. A. LeVine (Eds.), *Culture theory: Essays on mind, self, and emotion* (pp. 238–55). Cambridge University Press. https://doi.org/10.1093/0195145496.003.0006

Solomon, R. C. (1993). *The passions: Emotions and the meaning of life*. Hacket.

Solomon, R. C. (2001). Against valence ("positive" and "negative" emotions). In R. C. Solomon, *Not passion's slave: Emotions and choice* (pp. 135–147). Oxford University Press. https://doi.org/10.1093/0195145496.003.0010

Solomon, R. C. (2002). Back to basics: On the very idea of basic emotions. *Journal for the Theory of Social Behaviour*, *32*(2), 115–144. https://doi.org/10.1111/1468-5914.00180

Solomon, R. C., & Stone, L. D. (2002). On "positive" and "negative" emotions. *Journal for the Theory of Social Behaviour*, *32*(4), 417–443. https://doi.org/10.1111/1468-5914.00196

Solowij, N., Hall, W., & Lee, N. (1992). Recreational MDMA use in Sydney: A profile of "Ecstasy" users and their experiences with the drug. *British Journal of Addiction*, *87*(8), 1161–1172. https://doi.org/10.1111/j.1360-0443.1992.tb02003.x

Soteriou, M. (2018). The ontology of emotion. In H. Naar & F. Teroni (Eds.), *The ontology of emotions* (pp. 71–89). Cambridge University Press. https://doi.org/10.1017/9781316275221.005

Spencer, H. (1855). *Principles of psychology*. Longman, Brown, Green & Longmans. https://doi.org/10.1037/14065-000

Spinoza, B. (1982). *Ethics* (S. Shirley, Trans.). Hackett (Original work published 1677).

Sprangers, M. A. G., & Schwartz, C. E. (1999). Integrating response shift into health-related quality of life research: A theoretical model. *Social Science & Medicine*, *48*(11), 1507–1515. https://doi.org/10.1016/s0277–9536(99)00045-3

Spruyt, A., Gast, A., & Moors, A. (2011). The sequential priming paradigm: A primer. In K. C. Klauer, A. Voss, & C. Stahl (Eds.), *Cognitive methods in social psychology* (pp. 48–77). Guilford.

Spruyt, A., Tibboel, H., De Schryver, M., & De Houwer, J. (2018). Automatic stimulus evaluation depends on goal relevance. *Emotion*, *18*(3), 332–341. https://doi.org/10.1037/emo0000361

Stein, D. G., & Rosen, J. J. (1974). *Motivation and emotion*. Macmillan.

Stephens, C. L., Christie, I. C., & Friedman, B. H. (2010). Autonomic specificity of basic emotions: Evidence from pattern classification and cluster analysis. *Biological Psychology*, *84*(3), 463–473. https://doi.org/10.1016/j.biopsycho.2010.03.014

Stich, S. P. (1983). *From folk psychology to cognitive science: The case against belief*. MIT Press.

Storbeck, J., & Clore, G. L. (2007). On the interdependence of cognition and emotion. *Cognition and Emotion*, *21*(6), 1212–1237. https://doi.org/10.1080/02699930701438020

Storbeck, J., & Robinson, M. D. (2004). Preferences and inferences in encoding visual objects: A systematic comparison of semantic and affective priming. *Personality and Social Psychology Bulletin*, *30*(1), 81–93. https://doi.org/10.1177/0146167203258855

Storbeck, J., Robinson, M. D., & McCourt, M. E. (2006). Semantic processing precedes affect retrieval: The neurological case for cognitive primacy in visual processing. *Review of General Psychology*, *10*(1), 41–55. https://doi.org/10.1037/1089-2680.10.1.41

Storm, C., & Storm, T. (1987). A taxonomic study of the vocabulary of emotions. *Journal of Personality and Social Psychology*, *53*(4), 805–816. https://doi.org/10.1037/0022-3514.53.4.805

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*(3), 220–247. https://doi.org/10.1207/s15327957pspr0803_1

Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology*, *54*(5), 768–777. https://doi.org/10.1037/0022–3514.54.5.768

Swartz, N. (1997). *Definitions, dictionaries, and meanings.* www.sfu.ca/~swartz/definitions.htm

Sze, J. A., Gyurak, A., Yuan, J. W., Levenson, R. W. (2010). Coherence between emotional experience and physiology: Does body awareness training have an impact? *Emotion*, *10*(6), 803–814. https://doi.org/10.1037/a0020146

Sznycer, D., Cosmides, L., & Tooby, J. (2017). Adaptationism carves emotions at their functional joints. *Psychological Inquiry*, *28*(1), 56–62. https://doi.org/10.1080/1047840x.2017.1256132

Tamir, M. (2016). Why do people regulate their emotions? A taxonomy of motives in emotion regulation. *Personality and Social Psychology Review*, *20*(3), 199–222. https://doi.org/10.1177/1088868315586325

Tamir, M., & Millgram, Y. (2017). Motivated emotion regulation: Principles, lessons, and implications of a motivational analysis of emotion regulation. In A. J. Elliot (Ed.), *Advanced in motivation science* (pp. 207–247). Elsevier. https://doi.org/10.1016/bs.adms.2016.12.001

Tamir, M., Vishkin, A., & Gutentag, T. (2020). Emotion regulation is motivated. *Emotion*, *20*(1), 115–119. https://doi.org/10.1037/emo0000635

Tapia, E., Breitmeyer, B. G., & Shooner, C. R. (2010). Role of task-directed attention in nonconscious and conscious response priming by form and color. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(1), 74–87. https://doi.org/10.1037/a0017166

Tappolet, C. (2000). *Émotions et valeurs [Emotions and values]*. Presses Universitaires de France. https://doi.org/10.3917/puf.tappo.2000.01

Tappolet, C. (2010). Emotion, motivation, and action: The case of fear. In P. Goldie (Ed.), *The Oxford handbook of philosophy of emotion* (pp. 325–345). Cambridge University Press. https://doi.org/10.1093/oxfordhb/9780199235018.003.0015

Tappolet, C. (2016). *Emotions, values, and agency*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199696512.001.0001

Tappolet, C. (2018). The metaphysics of moods. In H. Naar & F. Teroni (Eds.), *The ontology of emotions* (pp. 169–177). Cambridge University Press. https://doi.org/10.1017/9781316275221.009

Tappolet, C. (2020). Emotions inside out: The nonconceptual content of emotions. In C. Demmerling & D. Schröder (Eds.), *Concepts in thought, action, and emotion* (pp. 257–276). Routledge.

Tassinary L. G., & Cacioppo, J. T. (1992). Unobservable facial actions and emotion. *Psychological Science*, *3*(1), 28–33. https://doi.org/10.1111/j.1467-9280.1992.tb00252.x

Teachman, B. A., Clerkin, E. M., Cunningham, W. A., Dreyer-Oren, S., & Werntz, A. (2019). Implicit cognition and psychopathology: Looking back and looking forward. *Annual Review of Clinical Psychology*, *15*, 123–148. https://doi.org/10.1146/annurev-clinpsy-050718-095718

Teasdale, J. D. (1999). Multi-level theories of cognition-emotion relations. In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 665–682). Wiley. https://doi.org/10.1002/0470013494.ch31

Teasdale, J. D., & Barnard, P. J. (1993). *Affect, cognition, and change: Re-modelling depressive thought*. Erlbaum.

Teroni, F. (2007). Emotions and formal objects. *Dialectica*, *61*(3), 395–415. https://doi.org/10.1111/j.1746-8361.2007.01108.x

Teroni, F. (2021). Evaluative theories in psychology and philosophy of emotion. *Mind and Language*. https://doi/org./10.1111/mila.12374

Thagard, P. R. (1978). The best explanation: Criteria for theory choice. *Journal of Philosophy*, *75*(2), 76–92. https://doi.org/10.2307/2025686

Thalberg, I. (1964). Emotion and thought. *American Philosophical Quarterly*, *1*(1), 45–55. www.jstor.org/stable/20009118

Thayer, R. E. (1989). *The biopsychology of mood and activation*. Oxford University Press.

Thorndike, A. (1911). *Animal intelligence: Experimental studies*. Macmillan. https://doi.org/10.5962/bhl.title.55072

Tobler, P. N., O'Doherty, J. P., Dolan, R. J., & Schultz, W. (2006). Reward value coding distinct from attitude-related uncertainty coding in human reward systems. *Journal of Neurophysiology*, *97*(2), 1621–1632. https://doi.org/10.1152/jn.00745.2006

Tolman, E. C. (1932). *Purposive behavior in animals and men*. Century.

Tomaka, J., Blascovich, J., Kelsey, R. M., & Leitten, C. L. (1993). Subjective, physiological, and behavioral effects of threat and challenge appraisal. *Journal of Personality and Social Psychology*, *65*(2), 248–260. https://doi.org/10.1037/0022-3514.65.2.248

Tomaka, J., Blascovich, J., Kibler, J., & Ernst, J. M. (1997). Cognitive and physiological antecedents of threat and challenge appraisal. *Journal of Personality and Social Psychology*, *73*(1), 63–72. https://doi.org/10.1037/0022–3514.73.1.63

Tomkins, S. S. (1962). *Affect, imagery, consciousness: Vol. 1. The positive affects*. Springer. https://doi.org/10.1037/14351-000

Tomkins, S. S. (1963). *Affect, imagery, consciousness: Vol. 2. The negative affects*. Springer.

Tomkins, S. S. (1970). Affects as primary motivational system. In M. B. Arnold (Ed.), *Feelings and emotions: The Loyola symposium* (pp. 101–110). Academic Press. https://doi.org/10.1016/b978–0-12-063550-4.50013-9

Tomkins, S. S. (1995). *Exploring affect: The selected writings of Silvan S. Tomkins* (E. V. Demos, Ed.). Cambridge University Press. https://doi.org/10.1017/cbo9780511663994

Tomonaga, M., Tanaka, M., Matsuzawa, T., Myowa-Yamakoshi, M., Kosugi, D., Mizuno, Y., Okamoto, S., Yamaguchi, M. K., & Bard, K. A. (2004). Development of social cognition in infant chimpanzees (Pan troglodytes):

Face recognition, smiling, gaze, and the lack of triadic interactions. *Japanese Psychological Research*, *46*(3), 227–235. https://doi.org/10.1111/j.1468-5584.2004.00254.x

Tooby, J., & Cosmides, L. (1990). The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology*, *11*(4–5), 375–424. https://doi.org/10.1016/0162-3095(90)90017-z

Topolinski, S., & Strack, F. (2015). Corrugator activity confirms immediate negative affect in surprise. *Frontiers in Psychology*, *6*, 134. https://doi.org/10.3389/fpsyg.2015.00134

Touroutoglou, A., Andreano, J., Dickerson, B. C., & Barrett, L. F. (2020). The tenacious brain: How the anterior mid-cingulate contributes to achieving goals. *Cortex*, *123*, 12–29. https://doi.org/10.1016/j.cortex.2019.09.011

Tracy, J. L. (2014). An evolutionary approach to understanding distinct emotions. *Emotion Review*, *6*(4), 308–312. https://doi.org/10.1177/1754073914534478

Tracy, J. L., & Matsumoto, D. (2008). The spontaneous expression of pride and shame: Evidence for biologically innate nonverbal displays. *Proceedings of the National Academy of Sciences*, *105*(33), 11655–11660. https://doi.org/10.1073/pnas.0802686105

Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review*, *3*(4), 397–405. https://doi.org/10.1177/1754073911410747

Tracy, J. L., Shariff, A. F., & Cheng, J. T. (2010). A naturalist's view of pride. *Emotion Review*, *2*(2), 163–177. https://doi.org/10.1177/1754073909354627

Tracy, J. L., Shariff, A. F., Zhao, W., & Henrich, J. (2013). Cross-cultural evidence that the nonverbal expression of pride is an automatic status signal. *Journal of Experimental Psychology: General*, *142*(1), 163–180. https://doi.org/10.1037/a0028412

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*(2), 440–463. https://doi.org/10.1037/a0018963

Tucker, C. (2013). Seemings and justification: An introduction. In C. Tucker (Ed.), *Seemings and justification: New essays on dogmatism and phenomenal conservatism* (pp. 1–29). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199899494.003.000

Turillo, C. J., Folger, R., Lavelle, J. J., Umphress, E. E., & Gee, J. O. (2002). Is virtue its own reward? Self-sacrificial decisions for the sake of fairness. *Organizational Behavior and Human Decision Processes*, *89*(1), 839–865. https://doi.org/10.1016/S0749–5978(02)00032-8

Turner, S. A., Jr., & Silvia, P. J. (2006). Must interesting things be pleasant? A test of competing appraisal structures. *Emotion*, *6*(4), 670–674. https://doi.org/10.1037/1528-3542.6.4.670

Turner, T. J., & Ortony, A. (1992). Basic emotions: Can conflicting criteria converge? *Psychological Review*, *99*(3), 566–571. https://doi.org/10.1037/0033-295x.99.3.566

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Tye, M. (2008). The experience of emotion: An intentionalist theory. *Revue Internationale de Philosophie*, *243*(1), 25–50. https://doi.org/10.3917/rip.243.0025

Tymula, A., Belmaker, L. A. R., Roy, A. K., Ruderman, L., Manson, K., Glimcher, P. W., & Levy, I. (2012). Adolescents' risk-taking behavior is driven by tolerance to ambiguity. *Proceedings of the National Academy of Sciences*, *109*(42), 17135–17140. https://doi.org/10.1073/pnas.1207144109

Valins, S. (1966). Cognitive effects of false heart-rate feedback. *Journal of Personality and Social Psychology*, *4*(4), 400–408. https://doi.org/10.1037/h0023791

Van de Cruys, S. (2017) Affective value in the predictive mind. In T. Metzinger & W. Wiese (Eds.), *Philosophy and predictive processing* (Vol. 24, pp. 1–21). MIND Group. https://doi.org/10.15502/9783958573253

Van Dessel, P., De Houwer, J., Roets, A., & Gast, A. (2016). Failures to change stimulus evaluations by means of subliminal approach and avoidance training. *Journal of Personality and Social Psychology*, *110*(1), e1–e15. https://doi.org/10.1037/pspa0000039

van Gaal, S., de Lange, F. P., & Cohen, M. X. (2012). The role of consciousness in cognitive control and decision making. *Frontiers in Human Neuroscience*, *6*, 121. https://doi.org/10.3389/fnhum.2012.00121

van Kleef, G. A. (2009). How emotions regulate social life: The emotions as social information (EASI) model. *Current Directions in Psychological Science*, *18*(3), 184–188. https://doi.org/10.1111/j.1467-8721.2009.01633.x

van Kleef, G. A., van Doorn, E. A., Heerdink, M. W., & Koning, L. F. (2011). Emotion is for influence. *European Review of Social Psychology*, *22*(1), 114–163. https://doi.org/10.1080/10463283.2011.627192

van Reekum, C. M., & Scherer, K. R. (1997). Levels of processing in emotion-antecedent appraisal. In G. Matthews (Ed.), *Cognitive science perspectives on personality and emotion* (pp. 259–300). Elsevier Science. https://doi.org/10.1016/S0166–4115(97)80123-9

Varela, F. J. (1991). Whence the origin of perception? A cartography of current ideas. In F. Varela & J. P. Dupuy (Eds.), *Understanding origins: Contemporary ideas on the origin of life, mind and society* (pp. 235–265). Kluwer.

Velasco, P. F., & Loev, S. (2020). Affective experience in the predictive mind: A review and new integrative account. *Synthese*, 1–36. https://doi.org/10.1007/s11229–020-02755-4

Velleman, J. D. (2000). *The possibility of practical reason*. Clarendon Press. https://doi.org/10.3998/maize.13240734.0001.001

Velten Jr., E. (1968). A laboratory task for induction of mood states. *Behaviour Research and Therapy*, *6*(4), 473–482. https://doi.org/10.1016/0005-7967(68)90028-4

Verduyn, P., Delvaux, E., Van Coillie, H., Tuerlinckx, F., & Van Mechelen, I. (2009). Predicting the duration of emotional experience: Two experience sampling studies. *Emotion*, *9*(1), 83–91. https://doi.org/10.1037/a0014610

Verduyn, P., Van Mechelen, I., Tuerlinckx, F., Meers, K., & Van Coillie, H. (2009). Intensity profiles of emotional experience over time. *Cognition and Emotion*, *23*(7), 1427–1443. https://doi.org/10.1080/02699930902949031

Verplanck, W. S. (1954). Burrhus F. Skinner. In W. K. Estes et al. (Eds.), *Modern learning theory: A critical analysis of five examples* (pp. 267–316). Appleton-Century-Crofts. https://doi.org/10.1037/10626-003

Vervliet, B., & Boddez, Y. (2020). Memories of 100 years of human fear conditioning research and expectations for its future. *Behaviour Research and Therapy*, *135*, 103732. https://doi.org/10.1016/j.brat.2020.103732

Vick, S. J., Waller, B. M., Parr, L. A., Pasqualini, M. C. S., & Bard, K. A. (2007). A cross-species comparison of facial morphology and movement in humans and chimpanzees using the facial action coding system (FACS). *Journal of Nonverbal Behavior*, *31*(1), 1–20. https://doi.org/10.1007/s10919–006-0017-z

Vishkin, A., Hasson, Y., Millgram, Y., & Tamir, M. (2020). One size does not fit all: Tailoring cognitive reappraisal to different emotions. *Personality and Social Psychology Bulletin*, *46*(3), 469–484. https://doi.org/10.1177/0146167219861432

Vogt, J., De Houwer, J., Crombez, G., & Van Damme, S. (2013). Competing for attentional priority: Temporary goals versus threats. *Emotion*, *13*(3), 587–598. https://doi.org/10.1037/a0027204

Volz, K. G., & Hertwig, R. (2016). Emotions and decisions: Beyond conceptual vagueness and the rationality muddle. *Perspectives on Psychological Science*, *11*(1), 101–116. https://doi.org/10.1177/1745691615619608

Voorspoels, W., Vanpaemel, W., & Storms, G. (2011). A formal ideal-based account of typicality. *Psychonomic Bulletin & Review*, *18*, 1006–1014. https://doi.org/10.3758/s13423–011-0122-9

Vytal, K., & Hamann, S. (2010). Neuroimaging support for discrete neural correlates of basic emotions: A voxel-based meta-analysis. *Journal of Cognitive Neuroscience*, *22*(12), 2864–2885. https://doi.org/10.1162/jocn.2009.21366

Wagenmakers, E. J., Beek, T., Dijkhoff, L., Gronau, Q. F., Acosta, A., Adams Jr., R. B., Albohn, D. N., Allard, E. S., Benning, S. D., Blouin-Hudon, E.-M., Bulnes, L. C., Caldwell, T. L., Calin-Jageman, R. J., Capaldi, C. A., Carfagno, N. S., Chasten, K. T., Cleeremans, A., Connell, L., DeCicco, J. M., . . . Zwaan, R. A. (2016). Registered replication report: Strack, Martin, & Stepper (1988). *Perspectives on Psychological Science*, *11*(6), 917–928. https://doi.org/10.1177/1745691616674458

Wager, T. D., Kang, J., Johnson, T. D., Nichols, T. E., Satpute, A. B., & Barrett, L. F. (2015). A Bayesian model of category-specific emotional brain responses. *PLoS Computational Biology*, *11*(4), e1004066. https://doi.org/10.1371/journal.pcbi.1004066

Wakefield, J., & Dreyfus, H. (1991). Intentionality and the phenomenology of action. In E. Lepore & R. Van Gulick (Eds.), *John Searle and his critics* (pp. 259–270). Blackwell.

Walentowska, W., Moors, A., Paul, K., & Pourtois, G. (2016). Goal relevance influences performance monitoring at the level of the FRN and P3 components. *Psychophysiology*, *53*(7), 1020–1033. https://doi.org/10.1111/psyp.12651

Watson, J. B. (1919). A schematic outline of the emotions. *Psychological Review*, *26*(3), 165–196. https://doi.org/10.1037/h0072509

Watson, J. B. (1929). Behaviorism: The modern note in psychology. In J. B. Watson & W. McDougall (Eds.), *The battle of behaviorism: An exposition and an exposure*. Norton.

Webb, T. L., Miles, E., & Sheeran, P. (2012). Dealing with feeling: A meta-analysis of the effectiveness of strategies derived from the process model of emotion regulation. *Psychological Bulletin*, *138*(4), 775–808. https://doi.org/10.1037/a0027600

Webb, T. L., Schweiger Gallo, I., Miles, E., Gollwitzer, P. M., & Sheeran, P. (2012). Effective regulation of affect: An action control perspective on emotion regulation. *European Review of Social Psychology*, *23*(1), 143–186. https://doi.org/10.1080/10463283.2012.718134

Weber, M. (1947). *The theory of social and economic organization*. Free Press.

Wegner, D. M., & Bargh, J. A. (1998). Control and automaticity in social life. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 1, pp. 446–496). McGraw-Hill.

Wells, G. L., & Petti, R. E. (1980). The effects of overt head movements on persuasion: Compatibility and incompatibility of responses. *Basic and Applied Social Psychology*, *1*(3), 219–230. https://doi.org/10.1207/s15324834basp0103_2

Wentura, D., Rothermund, K., & Bak, P. (2000). Automatic vigilance: The attention-grabbing power of approach- and avoidance-related social information. *Journal of Personality and Social Psychology*, *78*(6), 1024–1037. https://doi.org/10.1037/0022-3514.78.6.1024

Wetzel, L. (2018, Fall). Types and tokens. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. https://plato.stanford.edu/archives/fall2018/entries/types-tokens/

Whalen, P. J. (2007). The uncertainty of it all. *Trends in Cognitive Sciences*, *11*(12), 499–500. https://doi.org/10.1016/j.tics.2007.08.016

Whiting, D. (2011). The feeling theory of emotion and the object-directed emotions. *European Journal of Philosophy*, *19*(2), 281–303. https://doi.org/10.1111/j.1468-0378.2009.00384.x

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron*, *40*(3), 655–664. https://doi.org/10.1016/s0896-6273(03)00679-2

Widen, S. C., & Russell, J. A. (2010). Descriptive and prescriptive definitions of emotion. *Emotion Review*, *2*(4), 377–378. https://doi.org/10.1177/1754073910374667

Wiech, K., Kalisch, R., Weiskopf, N., Pleger, B., Stephan, K. E., & Dolan, R. J. (2006). Anterolateral prefrontal cortex mediates the analgesic effect of expected and perceived control over pain. *Journal of Neuroscience*, *26*(44), 11501–11509. https://doi.org/10.1523/jneurosci.2568-06.2006

Wiegman, I. (2020). Emotional actions without goals. *Erkenntniss*, 1–31. https://doi.org/10.1007/s10670-019-00200-8

Wiener, N. (1948). *Cybernetics: Control and communication in the animal and the machine*. MIT Press. https://doi.org/10.1037/13140-000

Wierzbicka, A. (1986). Human emotions: Universal or culture-specific? *American Anthropologist*, *88*(3), 584–594. https://doi.org/10.1525/aa.1986.88.3.02a00030

Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. *Psychological Bulletin*, *120*(1), 3–24. https://doi.org/10.1037/0033-2909.120.1.3

Williams, K. D. (2007). Ostracism. *Annual Review of Psychology*, *58*, 425–452. https://doi.org/10.1146/annurev.psych.58.110405.085641

Williams, K. D., & Jarvis, B. (2006). Cyberball: A program for use in research on interpersonal ostracism and acceptance. *Behavior Research Methods*, *38*(1), 174–180. https://doi.org/10.3758/bf03192765

Williams, L. E., Bargh, J. A., Nocera, C. C., & Gray, J. R. (2009). The unconscious regulation of emotion: Nonconscious reappraisal goals modulate emotional reactivity. *Emotion*, *9*(6), 847–854. https://doi.org/10.1037/a0017745

Wilson, F. C. L., & Gregory, J. D. (2018). Overgeneral autobiographical memory and depression in older adults: A systematic review. *Aging & Mental Health*, *22*(5), 575–586. https://doi.org/10.1080/13607863.2017.1326461

Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 35, pp. 345–411). Academic Press. https://doi.org/10.1016/S0065–2601(03)01006-2

Wilson-Mendenhall, C. D., Barrett, L. F., Simmons, W. K., & Barsalou, L. W. (2011). Grounding emotion in situated conceptualization. *Neuropsychologia*, *49*(5), 1105–1127. https://doi.org/10.1016/j.neuropsychologia.2010.12.032

Winkielman, P., & Berridge, K. C. (2004). Unconscious emotion. *Current Directions in Psychological Science*, *13*(3), 120–123. https://doi.org/10.1111/j.0963-7214.2004.00288.x

Winkielman, P., Berridge, K. C., & Wilbarger, J. L. (2005). Unconscious affective reactions to masked happy versus angry faces influence consumption behavior and judgments of value. *Personality and Social Psychology Bulletin*, *31*(1), 121–135. https://doi.org/10.1177/0146167204271309

Witkower, Z., & Tracy, J. L. (2019). Bodily communication of emotion: Evidence for extrafacial behavioral expressions and available coding systems. *Emotion Review*, *11*(2), 184–193. https://doi.org/10.1177/1754073917749880

Wittgenstein, L. (1953). *Philosophical investigations*. Macmillan.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*(5232), 1880–1882. https://doi.org/10.1126/science.7569931

Wood, W., & Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychological Review*, *114*(4), 843–863. https://doi.org/10.1037/0033-295X.114.4.843

Wood, W., & Rünger, D. (2016). Psychology of habit. *Annual Review of Psychology*, *67*, 289–314. https://doi.org/10.1146/annurev-psych-122414-033417

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press. https://doi.org/10.1093/0195155270.001.0001

Worcester, W. L. (1893). Observations on some points of James's psychology: II. Emotion. *The Monist*, *3*(2), 285–298. https://doi.org/10.5840/monist18933224

Woud, M. L., Maas, J., Becker, E. S., & Rinck, M. (2013). Make the manikin move: Symbolic approach–avoidance responses affect implicit and explicit face evaluations. *Journal of Cognitive Psychology*, *25*(6), 738–744. https://doi.org/10.1080/20445911.2013.817413

Wright, C. I., Negreira, A., Gold, A. L., Britton, J. C., Williams, D., & Barrett, L. F. (2008). Neural correlates of novelty and face–age effects in young and elderly adults. *Neuroimage*, *42*(2), 956–968. https://doi.org/10.1016/j.neuroimage.2008.05.015

Wright, P., He, G., Shapira, N. A., Goodman, W. K., & Liu, Y. (2004). Disgust and the insula: fMRI responses to pictures of mutilation and contamination. *Neuroreport*, *15*(15), 2347–2351. https://doi.org/10.1097/00001756-200410250-00009

Wundt, W. (1998). *Outlines of psychology* (C. H. Judd, Trans.). Thoemmes Press (Original work published 1897). https://doi.org/10.1037/12908-000

Ylikoski, P. (2013). Causal and constitutive explanation compared. *Erkenntnis*, *78*(2), 277–297. https://doi.org/10.1007/s10670–013-9513-9

Young, P. T. (1943). *Emotion in man and animal: Its nature and relation to attitude and motive*. Wiley.

Young, P. T. (1949). Emotion as disorganized response: A reply to Professor Leeper. *Psychological Review*, *56*(4), 184–191. https://doi.org/10.1037/h0054176

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2, Pt. 2), 1–27. https://doi.org/10.1037/h0025848

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*(2), 151–175. https://doi.org/10.1037/0003-066x.35.2.151

Zajonc, R. B. (1984). On the primacy of affect. *American Psychologist*, *39*(2), 117–123. https://doi.org/10.1037/0003–066X.39.2.117

Zajonc, R. B., Murphy, S. T., & Inglehart, M. (1989). Feeling and facial efference: Implications of the vascular theory of emotion. *Psychological Review*, *96*(3), 395–416. https://doi.org/10.1037/0033-295X.96.3.395

Index

https://doi.org/10.1017/9781107588882.017 Published online by Cambridge University Press

https://doi.org/10.1017/9781107588882.017 Published online by Cambridge University Press

STUDIES IN EMOTION AND SOCIAL INTERACTION

Titles Published in the Second Series *(continued from page ii)*

Speaking from the Heart: Gender and the Social Meaning of Emotion, by Stephanie A. Shields

The Hidden Genius of Emotion: Lifespan Transformations of Personality, by Carol Magai and Jeannette Haviland-Jones

The Mind and Its Stories: Narrative Universals and Human Emotion, by Patrick Colm Hogan

Feelings and Emotions: The Amsterdam Symposium, edited by Antony S. R. Manstead, Nico H. Frijda, and Agneta H. Fischer

Collective Guilt: International Perspectives, edited by Nyla R. Branscombe and Bertjan Doosje

The Social Life of Emotions, edited by Larissa Z. Tiedens and Colin Wayne Leach

Emotions and Multilingualism, by Aneta Pavlenko

Group Dynamics and Emotional Expression, edited by Ursula Hess and Pierre Philippot

Stigmatization, Tolerance and Repair: An Integrative Psychological Analysis of Responses to Deviance, by Anton J. M. Dijker and Willem Koomen

The Self and Its Emotions, by Kristján Kristjánsson

Face-to-Face Communication over the Internet: Emotions in a Web of Culture, Language and Technology, edited by Arvid Kappas and Nicole C. Krämer

What Literature Teaches Us about Emotion, by Patrick Colm Hogan

Emotion: A Biosocial Synthesis, by Ross Buck

Emotional Mimicry in Social Context, edited by Ursula Hess and Agneta Fischer

The Interpersonal Dynamics of Emotion: Towards an Integrative Theory of Emotions as Social Information, by Gerben A. van Kleef

From Self to Social Relationships: An Essentially Relational Perspective on Social Motivation, by Martijn van Zomeren

The Aesthetics of Emotion: Up the Down Staircase of the Mind-Body, by Gerald C. Cupchik

The Expression of Emotion: Philosophical, Psychological and Legal Perspectives, edited by Catharine Abell and Joel Smith

Emotional Lives: Dramas of Identity in an Age of Mass Media, by E. Doyle McCarthy

Interpersonal Emotion Dynamics in Close Relationships, edited by Ashley K. Randall and Dominik Schoebi

Emotion and Narrative: Perspectives in Autobiographical Storytelling, by Tilmann Habermas

Foundations of Affective Social Learning: Conceptualizing the Social Transmission of Value, edited by Daniel Dukes and Fabrice Clément

Heart to Heart: How Your Emotions Affect Other People, by Brian Parkinson

Beyond Virtue: The Politics of Educating Emotions, by Liz Jackson

Literature and Moral Feeling: A Cognitive Poetics of Ethics, Narrative, and Empathy, by Patrick Colm Hogan

Demystifying Emotions: A Typology of Theories in Psychology and Philosophy, by Agnes Moors