Nebraska Symposium on Motivation 66

Maital Neta
Ingrid J. Haas *Editors*

Emotion in the Mind and Body

Springer

Nebraska Symposium on Motivation

Volume 66

Series editor

Lisa Crockett
Department of Psychology, University of Nebraska–Lincoln, Lincoln, NE, USA

More information about this series at http://www.springer.com/series/7596

Maital Neta • Ingrid J. Haas
Editors

Emotion in the Mind and Body

Springer

Editors
Maital Neta
Department of Psychology
Center for Brain, Biology, and Behavior
University of Nebraska-Lincoln
Lincoln, NE, USA

Ingrid J. Haas
Department of Political Science
Center for Brain, Biology, and Behavior
University of Nebraska–Lincoln
Lincoln, NE, USA

ISSN 0146-7875
Nebraska Symposium on Motivation
ISBN 978-3-030-27472-6 ISBN 978-3-030-27473-3 (eBook)
https://doi.org/10.1007/978-3-030-27473-3

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Series Preface

We are pleased to offer this volume from the 66th Nebraska Symposium on Motivation.

This year the volume editors are Maital Neta and Ingrid Haas. In addition to overseeing the development of this book, the volume editors coordinated the 66th Symposium, including selecting and inviting the contributors. I would like to express my appreciation to Professors Neta and Haas and the contributors for a stimulating meeting and an excellent series of papers on emotion, an important factor in behavior, motivation, and many forms of psychopathology.

This symposium series is supported by funds provided by the Chancellor of the University of Nebraska–Lincoln, Harvey Perlman, and by funds given in memory of Professor Harry K. Wolfe to the University of Nebraska Foundation by the late Professor Cora L. Friedline. We are also grateful for the University of Nebraska Foundation's support via the Friedline bequest. This symposium volume, like those in the recent past, is dedicated in memory of Professor Wolfe, who brought psychology to the University of Nebraska. After studying with Professor Wilhelm Wundt in Germany, Professor Wolfe returned to his native state to establish the first undergraduate laboratory in psychology in the nation. As a student at the University of Nebraska, Professor Friedline studied psychology under Professor Wolfe.

Lincoln, NE, USA — Lisa J. Crockett

Contents

Chapter 1
Movere: Characterizing the Role of Emotion and Motivation in Shaping Human Behavior

Maital Neta and Ingrid J. Haas

From the moment we start our day, our lives are saturated with affective value (e.g., disappointment at the ringing of an alarm clock, enjoyment in that first cup of hot coffee). In this way, emotion is a lens through which we perceive and interact with the world, coloring our day-to-day experiences and driving our behavior. When we encounter new information (e.g., novel people, sounds, flavors), we readily sort this information into emotional valence categories: good or bad, reward or punishment, approach or avoid. This task of understanding and processing our emotional responses can have an enormous impact on many aspects of our lives, including shaping our social relationships, our long-term goal pursuit (e.g., occupational choices, efficacy expectations), and more. The field of affective science examines the nature of our emotional experience, expression, and the mechanisms with which we regulate these processes.

Emotion, like motivation, is derived from the Latin word *movere* (to move) and is one of the primary forces that activates or energizes our behaviors. Both emotion and motivation have important influences on many social and cognitive processes and can shape the way we navigate our social world. For example, emotion can change one's sensory and perceptual experiences (Balcetis & Dunning, 2006; Siegel, Wormwood, Quigley, & Barrett, 2018; Vuilleumier, 2005) and have an impact on many higher-level cognitive processes, including attention (Vuilleumier, 2005), learning and memory (Kensinger & Corkin, 2003; Phelps, 2004; Um, Plass, Hayward, & Homer, 2012), reasoning (Damasio, 1996; Jung, Wranke, Hamburger, & Knauff, 2014; Kunda, 1990), problem-solving (Isen, Daubman, &

M. Neta (✉)
Department of Psychology, Center for Brain, Biology, and Behavior, University of Nebraska-Lincoln, Lincoln, NE, USA
e-mail: maitalneta@unl.edu

I. J. Haas
Department of Political Science, Center for Brain, Biology, and Behavior, University of Nebraska-Lincoln, Lincoln, NE, USA

M. Neta, I. J. Haas (eds.), *Emotion in the Mind and Body*, Nebraska Symposium on Motivation 66, https://doi.org/10.1007/978-3-030-27473-3_1

Nowicki, 1987), and even financial decision-making (Lerner, Li, & Weber, 2013) and the expression of political attitudes (Haas, 2016; Haas & Cunningham, 2014). Distinct emotions, even those with similar valence, can produce differences in cognitive processing. For example, anger has been shown to increase reliance on stereotypical and heuristic thinking when compared to sadness or neutral mood (Bodenhausen, Sheppard, & Kramer, 1994).

As is easily observed by following current events, emotion also has important implications for political behavior, as recent research has shown that emotion can contribute to political polarization, attraction to "fake news," and the spread of misinformation. For example, recent research using Twitter data demonstrated that the spread of information online is both highly polarized and faster when moral content is combined with emotional words (Brady, Wills, Jost, Tucker, & Van Bavel, 2017). Emotion can also motivate political action, as we have seen in recent years when, for example, school shootings in places like Parkland, Florida, have led to increased calls for action on gun control (Meyer, 2019). Recent work has even suggested that there are ideological differences in emotional responses (e.g., Hibbing, Smith, & Alford, 2014), leading some to conclude that political ideology may have biological origins. Thus, the study of emotion has important implications for society more broadly. In this volume, we provide a brief sampling of some of the areas of research that have been dedicated to elucidating the role of emotion and motivation in shaping human behavior.

The Field of Affective Science

Throughout history, scholars have been asking questions about the role of emotion in human experience and have approached this work in different ways. Philosophers like David Hume wondered about the relationship between emotion and reason (Hume, 1740). Many of these individuals proposed versions of mind-body dualism or the idea that the mind and body existed separately—that emotion literally resided in the gut, while reason lived in the mind (e.g., Descartes, 1641). Experimental psychology adopted an empirical approach to studying emotion that was largely inspired by these earlier questions. Relying largely on empirical observation, experimental manipulation, and self-report, psychologists observe how people and animals respond to emotional situations and ask people directly about their emotional experiences.

More recently, work on emotion has begun to integrate biology, physiology, and genetics. This work moves away from self-report and allows for the examination of emotional experience via physiological responses, using a variety of measurement techniques including heart rate, skin conductance, electromyography (EMG), and eye tracking. There has also been substantial research looking at both individual and genetic variation in emotional responses, with an appreciation that we are not all created equal and individual differences are perhaps the norm rather than the exception (e.g., Hamann & Canli, 2004; John & Gross, 2004; Neta, Norris, & Whalen,

2009; Tomarken, Davidson, Wheeler, & Doss, 1992). Neuroscience has provided emotion researchers with new methods and new technologies, including electroencephalography (EEG), magnetoencephalography (MEG), and structural and functional MRI. These methods allow for the examination of the neural representation of emotional experience and have provided researchers with new ways to examine and understand the mind-body connection (e.g., Damasio, 1996).

Research on emotion has increased significantly over the past two decades with many fields contributing—including psychology, neuroscience, medicine, sociology, political science, and computer science. The resulting interdisciplinary field of affective science examines the role of emotion and motivation in social decision-making and the underlying (e.g., neural) mechanisms that support these processes. The numerous theories that attempt to explain the origin, neurobiology, experience, structure, and function of emotions have only fostered more intense research on this topic.

At this point, the challenge is to integrate our understanding of emotion across levels of analysis, from how we think about the psychology and conscious experience of emotion, to biology, genetics, and neuroscience, and then extrapolating to what this means for society, group behavior, and politics more broadly. Arguably, it is not enough to simply study emotional phenomena as they exist at multiple levels of analysis, but engage in work that attempts to synthesize across levels of analysis. Scholars have labeled this approach integrative multilevel analysis, or understanding how multiple levels of analysis are related to one another and using each level to constrain theory at other levels (Cacioppo & Berntson, 1992). This approach is based on the idea that the translation across levels of analysis may not always be easy or straightforward—that in some cases translation across levels is nonadditive (Marr, 1982). The field of affective science has illustrated some of the challenges of this work, for example, by showing that emotion categories developed by social psychologists (Ekman, 1992) do not necessarily map on to emotional experience at the neural or computational level (Touroutoglou, Lindquist, Dickerson, & Barrett, 2015) and that there may be some aspects of emotion processing that are shared across emotion categories (Phan, Wager, Taylor, & Liberzon, 2002).

Some important and related areas of work include an examination of these processes across the lifespan, as well as in both healthy and clinical populations. Lifespan research in affective science has exploded with many new findings related to the positivity effect in aging (Mather & Carstensen, 2005; Neta & Tong, 2016) and age-related changes in decision-making and motivation (Samanez-Larkin & Knutson, 2015; see also Knutson & Srirangarajan, this volume). On the other end of the lifespan, emotional valence conveyed through facial expressions is reliably identified in early childhood (Bruce et al., 2000; Widen & Russell, 2008), but there are profound developmental changes in emotional behavior that support functional social behavior (Denham, 1998; John & Gross, 2004; Petro, Tottenham, & Neta, submitted; Saarni, 1984) and predict mental health outcomes (Reef, Diamantopoulou, van Meurs, Verhulst, & van der Ende, 2011). Along with behavioral changes, there are neurobiological and structural brain changes during puberty (Blakemore, Burnett, & Dahl, 2010), such as decreased amygdala reactivity (Guyer et al., 2008;

Swartz, Carrasco, Wiggins, Thomason, & Monk, 2014), and more inverse frontoamygdalar connectivity (Gee et al., 2013; Perlman & Pelphrey, 2011) thought to reflect a downregulation of the amygdala (Swartz et al., 2014). Taken together, there are many important age-related changes in emotion across the lifespan that must be considered (see e.g., Nook & Somerville, this volume).

Further, much of the extant work on emotion in healthy groups is aimed at better understanding situations of dysfunction and identifying treatment targets (see, e.g., Kragel & Wager, this volume). However, this translational work must be tested in clinical populations. Notably, as evidenced by the Research Domain Criteria (RDoC) put forth by the National Institute of Mental Health, emotion and motivation are central systems impacted across many areas of psychopathology. For example, impairments in emotion regulation are associated with depression (Beck, 1979), anxiety (Amstadter, 2008), psychosis (Livingstone, Harper, & Gillanders, 2009), and addiction (Goldstein & Volkow, 2011). While significant progress has been made to understand the mechanisms underlying emotion-related disorders (Davidson, Abercrombie, Nitschke, & Putnam, 1999), there is much work to be done. Recent research examining impairment as a function of these broader systems is making great strides in localizing the causes and identifying treatment targets for specific disorders (see, e.g., Barch, Pagliaccio, Luking, Moran, & Culbreth, this volume).

While much progress has been made, there is still plenty of challenging work to be done. As is apparent from the collection of chapters included in this edited volume, the participants in the 2018 Nebraska Symposium on Motivation are among the most prominent researchers in the field who are currently tackling these difficult problems.

Looking Ahead

The symposium and the chapters in this volume focused on an array of research questions essential to moving the study of emotion forward, such as understanding the basic function of emotion and motivation—why are these phenomena so essential to human experience and what function do they serve? How does the role of emotion and motivation in guiding human behavior shift across the lifespan? How do we regulate our emotions, both successfully and unsuccessfully? What does the study of the brain tell us about our emotional experiences? And, in general, what happens when emotion goes awry? How do variations in both emotional and motivational experience contribute to psychopathology? Given the broad approach to understanding emotion and motivation and the evolution toward a deep science approach that examines these phenomena across many levels of analysis, this volume presents a number of answers to these important questions, capitalizing on a variety of methodological tools to approach this work.

The volume begins with a developmental perspective from Erik Nook and Leah Somerville that considers the changes in emotion concepts, or how we internally

represent emotion (Chap. 2). Although there has been extensive research aimed at defining emotion using a variety of dimensional models (e.g., the Circumplex Model using dimensions of valence and arousal; Russell, 1980) or instead focusing on a set of discrete "core" emotion categories (Ekman, 1992), there has been less attention paid to the ways these concepts change over the lifespan. Nook and Somerville take us on an exploration of emotion concept development and touch on the ways that this developmental process could relate to emotion regulation as well as risk factors for psychopathology.

Kevin Ochsner then expands on the concept of emotion regulation by presenting a model that builds on prior work examining the self-regulation of emotion (Gross, 1998), but adds an important component related to the social regulation of emotion (Chap. 3). First, he explores the stages of emotion regulation that must take place even before regulation begins (i.e., identifying one's current emotional state, evaluating a need for regulation, and selecting a regulatory strategy) and offers evidence for the brain systems that support each stage. Then, he applies this model to the social regulation of emotion, or instances in which we identify the emotional state of another and ultimately offer regulatory support. In the end, Ochsner links these findings to related areas of work (e.g., developmental changes in emotion regulation and dysfunction evident in specific clinical populations), including broader accounts for social and self-regulatory phenomena.

Elaine Fox and Robert Keers provide a theoretical framework to explain how genetic, environmental, and cognitive factors interact in the development of emotional disorders and psychological well-being (Chap. 4). Specifically, they focus on how cognitive and genetic vulnerabilities interact to influence risk for developing mental health disorders (e.g., anxiety and depression). They provide a comprehensive overview of the various cognitive and genetic approaches in the field for studying emotional disorders and well-being, including an expansion of their own Cognitive Bias (CogBIAS) Hypothesis (Fox & Beevers, 2016). They round out their chapter with a discussion of the barriers in existing research and how future work may overcome these barriers to offer a more in-depth understanding of the etiology, maintenance, and, ultimately, treatment of emotional disorders.

Building on the approach to considering psychopathology, Deanna Barch and colleagues examine the pathways linking hedonics to motivated behavior (Chap. 5). This chapter provides a more in-depth focus on problems with cognitive processing (e.g., deficits in reward) and how these problems contribute to dysfunction, particularly focusing on the distinction between depression and schizophrenia. The authors argue that it is important to understand how similar behaviors in different psychopathological groups may arise from different motivational underpinnings, and thus different interventions may be necessary for re-establishing healthy function. Importantly, this work also considers various measures (i.e., self-report, physiological, and neural) that converge to offer a clearer picture of underlying mechanisms of healthy and aberrant function. In the end, they offer a picture of open questions and goals for future work.

To link motivation and behavior more broadly, James Gross and colleagues offer a comprehensive model of motivation that considers the integration of information

that contributes to action and goal pursuit (Chap. 6). This valuation systems perspective relies on hierarchical perception and action loops that use ascending and descending feedback control to match mental models to the world and behavior to goals. This model represents motivation as emergent, arising from feedback loops that work between action affordances and action tendencies; constructive, arising from a negotiation between the person and the environment; and allostatic, flexibly adjusting the state of the system in anticipation of changes in the environment. Ultimately, this chapter offers an integrated model that accounts for highly abstract concepts such as self-identity and self-regulation.

Expanding on work that attempts to link motivation and behavior, Brian Knutson and Tara Srirangarajan offer a deep science framework for examining motivation that focuses on anticipatory affect, or the experience of increased arousal before uncertain goal outcomes (Chap. 7). This anticipatory affect could be associated with positive or negative arousal, which is associated with appetitive (approach) or aversive (avoid) motivational states. This chapter presents a new perspective on motivation that considers broad science, examining these processes across many functional domains, but more so, it unpacks the deep science, examining these processes across different levels of analysis, from physiology to process to purpose, and beyond. This chapter presents the implications and the limits of this deep science approach and offers suggestions for the integration of this approach in future work.

Finally, Philip Kragel and Tor Wager describe a novel approach to reducing complex neuroimaging data to measures that can characterize emotion states in a manner that is substantially stronger than approaches to date (Chap. 8). These models consider important characteristics of a brain representation, including its sensitivity (i.e., does it respond the same way every time?), specificity (i.e., does it respond to other salient events or only to this particular state?), and whether it is necessary (i.e., if the expression of this brain representation is suppressed, then is this state eliminated?) and sufficient (i.e., if the representation is activated exogenously, is the state recreated even in the absence of a stimulus?). The chapter presents three examples of such brain representations—one for pain, negative emotion, and discrete emotional experiences—and demonstrates that brain models can and do robustly and reproducibly predict an individual's affective experiences and even help to identify targets for interventions when necessary.

Summary

Although considerable progress has been made in recent years, it is clear that our understanding of the roles of emotion and motivation in shaping human behavior remains far from complete. This volume addresses an important breadth of questions related to these roles at various stages of life, differences in normal versus aberrant function, in linking levels of analysis, and in forging new approaches that are reproducible and generalizable. Notably, throughout this volume, questions are posed for directing future research, and models are presented that outline clear

predictions for this future work. Understanding emotion and motivation across many domains (breadth) and across levels of analysis (depth) promises to help us better understand the psychological triad—the way we think, feel, and behave.

Acknowledgments We were honored and delighted to organize the 66th annual Nebraska Symposium on Motivation. We would not have been able to accomplish this task without the help of many people. We would like to thank the University of Nebraska-Lincoln Chancellor Harvey Perlman and the late Professor Cora L. Friedline's bequest to the University of Nebraska Foundation in memory of Professor Harry K. Wolfe. The symposium would not be possible without their generous gifts. We would also like to thank Professor Lisa J. Crockett, the incoming series editor, for her support in putting this program together and Debra A. Hope, the outgoing series editor, for her advice when getting started, including selecting and inviting speakers. Last but not least, we would like to thank Pam Waldvogel for her incredible behind-the-scenes support throughout this process, and our graduate student assistants, Catherine C. Brown and Nicholas R. Harp. Thank you all so much for all your hard work and commitment to helping to make the symposium a success.

References

Amstadter, A. (2008). Emotion regulation and anxiety disorders. *Journal of Anxiety Disorders, 22*(2), 211–221.

Balcetis, E., & Dunning, D. (2006). See what you want to see: Motivational influences on visual perception. *Journal of Personality and Social Psychology, 91*, 612–625.

Barch, D. M., Pagliaccio, D., Luking, K., Moran, E. K., & Culbreth, A. J. (this volume). Pathways to motivational impairments in psychopathology: Common versus unique elements across domains. In M. Neta & I. J. Haas (Eds.), *Emotion in the mind and body*. Cham, Switzerland: Springer.

Beck, A. T. (Ed.). (1979). *Cognitive therapy of depression*. New York, NY: Guilford Press.

Blakemore, S.-J., Burnett, S., & Dahl, R. E. (2010). The role of puberty in the developing adolescent brain. *Human Brain Mapping, 31*, 926–933.

Bodenhausen, G. V., Sheppard, L., & Kramer, G. P. (1994). Negative affect and social perception: The differential impact of anger and sadness. *European Journal of Social Psychology, 24*, 45–62.

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences of the United States of America, 114*(28), 7313–7318. https://doi.org/10.1073/pnas.1618923114

Bruce, V., Campbell, R. N., Doherty-Sneddon, G., Langton, S., McAuley, S., & Wright, R. (2000). Testing face processing skills in children. *The British Journal of Developmental Psychology, 18*, 319–333.

Cacioppo, J. T., & Berntson, G. G. (1992). Social psychological contributions to the decade of the brain: Doctrine of multilevel analysis. *American Psychologist, 47*(8), 1019–1028.

Damasio, A. R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions: Biological Sciences, 351*, 1413–1420.

Davidson, R. J., Abercrombie, H., Nitschke, J. B., & Putnam, K. (1999). Regional brain function, emotion and disorders of emotion. *Current Opinion in Neurobiology, 9*(2), 228–234.

Denham, S. A. (1998). *Emotional development in young children*. New York, NY: Guilford Press.

Descartes, R. (1641). Meditations on first philosophy. In J. Cottingham, R. Stoothoff, & D. Murdoch (Eds.), *The philosophical writings of René Descartes* (Vol. II, pp. 1–62). Cambridge, England: Cambridge University Press.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*(3/4), 169–200.

Fox, E., & Beevers, C. G. (2016). Differential sensitivity to the environment: contribution of cognitive biases and genes to psychological wellbeing. *Molecular Psychiatry, 21*(12), 1657–1662. https://doi.org/10.1038/mp.2016.114

Gee, D., Humphreys, K., Flannery, J., Goff, B., Telzer, E. H., Shapiro, M., … Tottenham, N. (2013). A developmental shift from positive to negative connectivity in human amygdala–prefrontal circuitry. *The Journal of Neuroscience, 33*, 4584–4593.

Goldstein, R. Z., & Volkow, N. D. (2011). Dysfunction of the prefrontal cortex in addiction: Neuroimaging findings and clinical implications. *Nature Reviews Neuroscience, 12*(11), 652.

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*, 271–299.

Guyer, A. E., Monk, C. S., McClure-Tone, E. B., Nelson, E. E., Roberson-Nay, R., Adler, A. D., … Ernst, M. (2008). A developmental examination of amygdala response to facial expressions. *Journal of Cognitive Neuroscience, 20*, 1565–1582.

Haas, I. J. (2016). The impact of uncertainty, threat, and political identity on support for political compromise. *Basic and Applied Social Psychology, 38*(3), 137–152.

Haas, I. J., & Cunningham, W. A. (2014). The uncertainty paradox: Perceived threat moderates the effect of uncertainty on political tolerance. *Political Psychology, 35*(2), 291–302.

Hamann, S., & Canli, T. (2004). Individual differences in emotion processing. *Current Opinion in Neurobiology, 14*(2), 233–238.

Hibbing, J. R., Smith, K. B., & Alford, J. R. (2014). Differences in negativity bias underlie variations in political ideology. *Behavioral and Brain Sciences, 37*(3), 297–307. https://doi.org/10.1017/S0140525X13001192

Hume, D. (1740). *A treatise of human nature*.

Isen, A. M., Daubman, K. A., & Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. *Journal of Personality and Social Psychology, 52*, 1122–1131. https://doi.org/10.1037/0022-3514.52.6.1122

John, O. P., & Gross, J. J. (2004). Healthy and unhealthy emotion regulation: Personality processes, individual differences, and life span development. *Journal of Personality, 72*(6), 1301–1334.

Jung, N., Wranke, C., Hamburger, K., & Knauff, M. (2014). How emotions affect logical reasoning: Evidence from experiments with mood-manipulated participants, spider phobics, and people with exam anxiety. *Frontiers in Psychology, 5*, 570. https://doi.org/10.3389/fpsyg.2014.00570

Kensinger, E. A., & Corkin, S. (2003). Effect of negative emotional content on working memory and long-term memory. *Emotion, 3*, 378–393. https://doi.org/10.1037/1528-3542.3.4.378

Knutson, B., & Srirangarajan, T. (this volume). Towards a deep science of affect and motivation. In M. Neta & I. J. Haas (Eds.), *Emotion in the mind and body*. Cham, Switzerland: Springer.

Kragel, P., & Wager, T. D. (this volume). Reproducible, generalizable brain models of affective processes. In M. Neta & I. J. Haas (Eds.), *Emotion in the mind and body*. Cham, Switzerland: Springer.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*, 480–498.

Lerner, J. S., Li, Y., & Weber, E. U. (2013). The financial costs of sadness. *Psychological Science, 24*(1), 72–79.

Livingstone, K., Harper, S., & Gillanders, D. (2009). An exploration of emotion regulation in psychosis. *Clinical Psychology & Psychotherapy, 16*(5), 418–430.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York, NY: Henry Holt.

Mather, M., & Carstensen, L. L. (2005). Aging and motivated cognition: The positivity effect in attention and memory. *Trends in Cognitive Sciences, 9*(10), 496–502.

Meyer, D. S. (2019, February 14). One year after the Parkland shooting, is the #NeverAgain movement on track to succeed? *The Washington Post.* Retrieved March 15, 2019, from https://www.washingtonpost.com/news/monkey-cage/wp/2019/02/14/one-year-after-the-parkland-shooting-is-the-neveragain-movement-on-track-to-succeed/?utm_term=.86c2d1eeae95

Neta, M., Norris, C. J., & Whalen, P. J. (2009). Corrugator muscle responses are associated with individual differences in positivity-negativity bias. *Emotion, 9*(5), 640.

Neta, M., & Tong, T. T. (2016). Don't like what you see? Give it time: Longer reaction times associated with increased positive affect. *Emotion, 16*(5), 730.

Nook, E. C., & Somerville, L. H. (this volume). Emotion concept development from childhood to adulthood. In M. Neta & I. J. Haas (Eds.), *Emotion in the mind and body*. Cham, Switzerland: Springer.

Perlman, S. B., & Pelphrey, K. A. (2011). Developing connections for affective regulation: Age-related changes in emotional brain connectivity. *Journal of Experimental Child Psychology, 108*, 607–620.

Petro, N. M., Tottenham, N., & Neta, M. (submitted). Positive valence bias is associated with inverse frontoamygdalar connectivity and less depressive symptoms in developmentally mature children.

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

Phelps, E. A. (2004). Human emotion and memory: interactions of the amygdala and hippocampal complex. *Current Opinion in Neurobiology, 14*, 198–202. https://doi.org/10.1016/j.conb.2004.03.015

Reef, J., Diamantopoulou, S., van Meurs, I., Verhulst, F. C., & van der Ende, J. (2011). Developmental trajectories of child to adolescent externalizing behavior and adult DSM-IV disorder: Results of a 24-year longitudinal study. *Social Psychiatry and Psychiatric Epidemiology, 46*, 1233–1241.

Russell, J. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*, 1161–1178.

Saarni, C. (1984). An observational study of children's attempts to monitor their expressive behavior. *Child Development, 55*, 1504–1513.

Samanez-Larkin, G. R., & Knutson, B. (2015). Decision making in the ageing brain: Changes in affective and motivational circuits. *Nature Reviews Neuroscience, 16*(5), 278.

Siegel, E. H., Wormwood, J. B., Quigley, K. S., & Barrett, L. F. (2018). Seeing what you feel: Affect drives visual perception of structurally neutral faces. *Psychological Science, 29*(4), 496–503.

Swartz, J. R., Carrasco, M., Wiggins, J. L., Thomason, M. E., & Monk, C. S. (2014). Age-related changes in the structure and function of prefrontal cortex–amygdala circuitry in children and adolescents: A multi-modal imaging approach. *NeuroImage, 86*, 212–220.

Tomarken, A. J., Davidson, R. J., Wheeler, R. E., & Doss, R. C. (1992). Individual differences in anterior brain asymmetry and fundamental dimensions of emotion. *Journal of Personality and Social Psychology, 62*(4), 676.

Touroutoglou, A., Lindquist, K. A., Dickerson, B. C., & Barrett, L. F. (2015). Intrinsic connectivity in the human brain does not reveal networks for 'basic' emotions. *Social Cognitive and Affective Neuroscience, 10*(9), 1257–1265. https://doi.org/10.1093/scan/nsv013

Um, E., Plass, J. L., Hayward, E. O., & Homer, B. D. (2012). Emotional design in multimedia learning. *Journal of Education & Psychology, 104*, 485–498. https://doi.org/10.1037/a0026609

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences, 9*(12), 585–594.

Widen, S. C., & Russell, J. A. (2008). Children acquire emotion categories gradually. *Cognitive Development, 23*, 291–312.

Chapter 2
Emotion Concept Development from Childhood to Adulthood

Erik C. Nook and Leah H. Somerville

Introduction

The transformation from young child to mature adult brings with it changes in nearly every domain of psychological functioning. Central to this psychological growth is a collection of cognitive processes that evolve with development to shape our experiences across many domains. This includes the development of emotions, which can be defined as experiences that involve changes in subjective affect (i.e., feeling tone), behavior, and peripheral physiology in response to internal mental events or external stimuli. From infants' cries to the complex emotional states faced during adolescence, humans undergo complex transformations in emotional experiences that pervade everyday life and influence day-to-day well-being.

Researchers have charted several outward signs of emotional development. For example, children's initial emotional responses to frustrating situations shift from angry outbursts to bids for aid from a parent across their first few years (Cole et al., 2011). However, much less work has examined the development of underlying processes that shape the emotions we experience, perceive, and comprehend at different stages of development. Understanding the underlying processes that shape emotional development is crucial, both for building a mechanistic theory of emotional development and for understanding age-constrained mechanisms that threaten emotional well-being. Clarifying how internal emotional processes develop is therefore central to understanding how risks to emotional health and well-being change from childhood to adolescence to adulthood.

There is a rich historical literature from cognitive developmental traditions demonstrating that development brings an expansion and deepening of conceptual representations with age. In this chapter, we apply this perspective to ask how and

E. C. Nook · L. H. Somerville (✉)
Department of Psychology and Center for Brain Science, Harvard University, Cambridge, MA, USA
e-mail: somerville@fas.harvard.edu

M. Neta, I. J. Haas (eds.), *Emotion in the Mind and Body*, Nebraska Symposium on Motivation 66, https://doi.org/10.1007/978-3-030-27473-3_2

when emotion concepts develop and how emotion concept development shapes the characteristic emotional experiences of children, adolescents, and adults. To do so, we first describe normative changes in emotional experience as individuals transition from childhood to adulthood. Next, we introduce emotion concepts and their relationship with emotion perception and experience based on both theory and recent empirical work. We then focus on an emerging set of findings that begin to reveal the nature of emotion concept development, which reinforce the conclusion that emotion concept development underpins key changes in emotional experience from childhood to adulthood. Finally, we suggest fruitful avenues for future research on emotion concept development.

Age-Related Changes in Normative Emotional Experiences

The transition from childhood to adolescence, which begins around physical puberty, brings tremendous change in nearly every important arena of daily life. Adolescents' social groups grow, become more complex, and exhibit higher fluctuations in affiliation and status (Cairns, Leung, Buchanan, & Cairns, 1995) at the same time as social-evaluative concern increases (Westenberg, Drewes, Goedhart, Siebelink, & Treffers, 2004). Adolescents spend more time unmonitored by parents (Barnes, Hoffman, Welte, Farrell, & Dintcheff, 2007) and are thus challenged to make increasingly independent decisions about how to navigate the world based on a limited experience base. In many cultures, concerns with academic and personal achievement become salient as adolescents face stressful, life-altering decisions concerning future educational and occupational goals (Csikszentmihalyi & Larson, 1984). Physical changes such as growth spurts, pubertal hormonal surges (Blakemore, Burnett, & Dahl, 2010; Forbes & Dahl, 2010; Sisk & Zehr, 2005; Steinberg & Morris, 2001), and shifts in endogenous sleep patterns that "mismatch" school and work schedules (Peper & Dahl, 2013) are also common in adolescence. Models of adolescent development refer to these simultaneous, important shifts in demands as a "pile-up" of emotional stressors (Petersen, 1988; Petersen et al., 1993). Indeed, this characterization highlights the challenges adolescents face in managing concurrently changing bodies, relationships, and responsibilities.

It is worth considering whether differences in adolescents' daily affective states are a straightforward by-product of the intense, stressful, and uncertain environments they live in, rather than an underlying developing process. If this were the case, then laboratory measures of response to emotional provocation would not distinguish adolescents from older and younger ages. In fact, data suggest that adolescents do indeed differ from other ages in laboratory studies of emotion. Thus, as we detail here and elsewhere (McLaughlin, Garrad, & Somerville, 2015; Somerville & McLaughlin, 2018), there is ample reason to believe that layered beneath unique stressors of adolescent life lie distinct response profiles in emotional subprocesses that contribute to changes in emotional reactivity from childhood to adulthood. Prior work has granted insight into age-related changes from childhood to adulthood

in two "ingredients" of emotional experience: subjective emotional experience (i.e., affect) and physiological reactivity (see also Bailen, Green, & Thompson, 2018).

Given the major life changes that adolescents face, it is perhaps no surprise that adolescents experience distinct emotional states compared to children and adults. Experience-sampling studies indicate that compared to children, adolescents experience higher levels of negative affect, more variability (or lability) in the valence of their daily emotional experiences, more intense emotions, and when compared to adults, and more frequent bouts of mixed positive and negative affect (Larson, Moneta, Richards, & Wilson, 2002; Riediger, Schmiedek, Wagner, & Lindenberger, 2009). In addition, stressors elicit stronger negative affect among adolescents than children (Larson & Ham, 1993), suggesting a tighter coupling between stressful events and negative emotional experience. Flook (2011) found that in adolescent girls, there was a bidirectional relationship between negative mood and more negative interpersonal events and fewer positive interpersonal events, whereas a positive mood was related to fewer negative interpersonal events. Future work will be needed to infer causal directionality between these measures, but this work indicates that changes in daily mood are associated with shifts in adolescents' quality of life.

A second key "ingredient" contributing to affect is physiological reactivity to emotional antecedents. This includes activation of the sympathetic division of the autonomic nervous system (ANS) and the hypothalamic-pituitary-adrenal (HPA) axis. Evidence from both animal and human studies indicates that adolescence is characterized by heightened physiological reactivity to environmental stimuli, including reactions to stressful experiences. Rodent models have revealed key linkages between the systemic hormonal changes that are a hallmark of puberty on the one hand and physiological reactivity in the ANS and HPA axis on the other (Sisk & Zehr, 2005). In humans, ANS and HPA axis responses to social evaluation and performance-related stressors are greater among adolescents than children (Gunnar, Wewerka, Frenn, Long, & Griggs, 2009; Stroud et al., 2009). A similar developmental pattern has been observed in other indirect indices of physiological arousal, such that adolescents exhibit greater pupil dilation in response to social rejection than children (Silk et al., 2012), and they report stronger reductions in self-esteem when given a mix of positive and negative feedback from peers (Rodman, Powers, & Somerville, 2017). Other work suggests that even subtle situations involving the possibility of social evaluation, such as being observed on a video camera by a peer, generate greater autonomic arousal and self-reported emotion (i.e., embarrassment) in adolescents as compared to children and adults (Somerville et al., 2013). This is consistent with a broad set of findings indicating that adolescents exhibit unique patterns of neural activation when thinking about the cognitive and emotional states of other people as compared to children or adults (Blakemore, 2008). These findings suggest that individuals at different developmental stages might use distinct strategies to infer the internal mental states of others, which may ultimately contribute to their distinct profiles of emotional responses in social situations. In sum, biological mechanisms appear to sensitize adolescents' physiological responses to emotional provocation, which likely have widespread effects on adolescents' emotional experiences.

Emotion Concepts

An emotion concept is an individual's internal representation of what defines any given emotion (Adolphs, 2017; Barrett, 2006; Kousta, Vigliocco, Vinson, Andrews, & Del Campo, 2011). Emotion concepts are expressed through emotion words which are thought to help organize the semantic knowledge of a given emotion, including its cognitive and contextual causes, body sensations, prototypical facial expressions, and behavioral consequences (Lindquist, Satpute, & Gendron, 2015). Modern theories and emerging evidence suggest that emotion concepts exert a crucial influence in shaping how people experience or "construct" their own and others' emotions (Barrett, 2006; Brooks & Freeman, 2018; Lindquist & Barrett, 2008; Lindquist, Satpute, & Gendron, 2015; Nook, Lindquist, & Zaki, 2015). In other words, emotion concepts play a foundational role in emotional experiences.

There has been a surge of interest in understanding how conceptual processes shape emotional experiences, expressed most prominently through the constructionist theory of emotion (Barrett, 2006; Lindquist & Gendron, 2013; Lindquist, MacCormack, & Shablack, 2015; Lindquist, Satpute, & Gendron, 2015). This theory posits that people experience emotions when they use conceptual knowledge to parse, or conceptualize, their own internal affective state (i.e., the collection of interoceptive sensations that continuously vary along axes of arousal and valence). This conceptualization process effectively constructs a discrete experience of feeling a particular emotion. For example, when confronted by a snake on a hiking trail, one uses their understanding of the emotion "fear" to call their racing heart and sudden urge to flee an instance of fear. An analogous conceptualized process occurs when people parse the affect others express in their facial expressions, vocal expressions, or body movements into emotion types. For instance, when observing another person with wide eyes and stretched lips running through the woods, an observer uses the same concept of "fear" to construct the notion that the other individual is afraid of something. A growing body of research gives insight into these processes by testing how emotion concepts shape one's own emotional experiences and one's perceptions of others' emotions. We first review evidence of conceptual processing in emotion perception and emotion experience before discussing how these conceptual processes vary across development.

Emotion Concepts and Emotion Perception

Several studies testing adult samples have found that the accessibility of emotion concepts shifts how people perceive others' emotions. Lindquist, Barrett, Bliss-Moreau, and Russell (2006) used a semantic satiation paradigm (cf. Smith & Klein, 1990) to show that temporarily reducing access to emotion concepts interferes with emotion perception. Participants repeated an emotion category word (e.g., "anger") out loud either three times to prime the concept of that emotion or 30 times to satiate the concept. Repeating a word 30 times is thought to temporarily disconnect the

semantic meaning of a word from the label itself (Black, 2004). Conversely, repeating a word three times is thought to prime the concept associated with the word by activating it within one's semantic network (Collins & Loftus, 1975; Neely, 1977). Participants then indicated whether they believed two faces expressed the same emotion. Participants were slower and less accurate at judging whether faces represented the same category if they had repeated the emotion word that is typically applied to one or both of the faces 30 times compared to when they repeated it only three times. However, there was no effect of satiation on reaction time or accuracy when the word that was satiated did not apply to the emotions expressed by either face. Hence, satiating the meaning of an emotion word interfered with the perception of expressions that fell within that emotion category.

Gendron, Lindquist, Barsalou, and Barrett (2012) used a similar paradigm to show that satiating an emotion category expressed by a face eliminated typical repetition priming effects for that face. This implies that temporarily inhibiting emotion concepts interferes even with the perceptual encoding of emotional faces. In fact, studies of patients with brain damage have shown that loss of emotion concept knowledge is associated with an inability to sort faces into typical emotion categories (Lindquist, Gendron, Barrett, & Dickerson, 2014). Two patients with semantic dementia—who in these cases lost understanding of the semantic definitions of emotion words due to degeneration of the left anterior temporal lobe—sorted faces into piles based on valence (i.e., positive, neutral, and negative), rather than into specific emotion types (i.e., anger, disgust, fear, etc.). Similarly, LEW—a stroke patient who lost his ability to name objects—produced disorganized piles of emotional faces (Roberson, Davidoff, & Braisby, 1999). Importantly, in all of these cases, patients were not severely impaired on control tasks involving perceptual matching. Atypical performance only arose on tasks that required free sorting of emotional stimuli (i.e., the independent generation and application of emotion concepts).

Studies have also investigated how introducing or priming emotion concepts (rather than ablating or satiating them) affects emotion perception. Fugate, Gouzoules, and Barrett (2010) showed participants chimpanzee facial expressions either with or without nonsense words for different types of expressions. Only participants who were given labels (around which an emotion concept for these novel expressions could be organized) later showed categorical perception of chimpanzee faces morphed from one expression to another. Another set of studies used emotion concept priming to investigate conceptual processes in emotion perception. Participants saw an emotion expression for 1s (the cue), a blank screen for 200 ms, and then either another emotion expression or an emotion word for 1s (the target). When the target was on the screen, participants responded as quickly and accurately as they could to indicate whether it expressed the same emotion as the cue. Interestingly, responses were more accurate when the target was an emotion word, rather than a facial expression, and reaction time analyses showed that facial expression cues primed congruent emotion words more than congruent emotion expressions (Nook et al., 2015). If participants were matching stimuli based on visual qualities alone, one would expect the reverse relationship: faces should prime other

faces more than emotion words. This result suggests instead that participants activated emotion concepts when categorizing the emotion of the cue, leading to stronger priming of emotion words (which are unambiguously tied to emotion concepts) than emotional faces.

The authors' second study (Nook et al., 2015) replicated these results and extended them by asking participants to indicate at the end of each trial what expression they were shown as the cue from a set of "morphed" expressions. Participants reliably misreported their perceptions of the cue expressions: they tended to select expressions that were more like the emotion of the target than what they had actually observed, regardless of whether the target was a face or a word (e.g., participants' perception of the cue face was more "angry" if the concept for anger had been activated by a target that was either the word anger or an angry expression than if the target activated the concept for sadness). This result echoed similar results showing that associating facial expressions with emotion category words shapes visual representations of those expressions toward prototypical expressions of faces for those categories (Halberstadt, 2003, 2005; Halberstadt & Niedenthal, 2001). Like Bruner's classic study on top-down processes in color perception—priming the concept of a tomato can make an ambiguous color swatch seem more red (Bruner, Postman, & Rodrigues, 1951)—these studies suggest that conceptual processes also shape how we see our emotional world.

Several studies have explored the role of top-down processing in emotion perception by showing that categorical perceptions of faces change radically depending on the context in which the faces are situated (see Hassin, Aviezer, & Bentin, 2013 for a review). For example, Aviezer et al. (2008) showed that the same face is seen as expressing a variety of emotions, depending on the situation in which it is embedded (e.g., a "disgust face" can be said to express fear when pasted onto a body reacting in fear). Carroll and Russell (1996) found similar results by pairing the same emotion expression with a variety of verbal vignettes. These studies suggest that people do not automatically recognize innate emotion categories "encoded" in faces like "light emanating from a lighthouse" (Russell, Bachorowski, & Fernandez-Dols, 2003). Instead, contextual information guides what concepts people use to parse ambiguous facial affect.

Evidence from the functional neuroimaging literature also supports a role of concepts in emotion perception. A recent study showed participants faces that ranged from calm to afraid and asked them to either rate faces using a continuous sliding scale (ranging from calm to afraid) or a categorical scale (in which participants selected either "calm" or "afraid" (Satpute et al., 2016). Categorizing faces that expressed only a moderate amount of fear as "afraid," rather than "calm," was associated with increased amygdala activity compared to rating these ambiguous faces on a continuous scale. Thus, the mere act of parsing faces into the category "afraid" was associated with increased neural activity in a region reliably involved in processing fear expressions (Adolphs et al., 2005; Costafreda, Brammer, David, & Fu, 2007; Whalen & Phelps, 2009). This effect parallels broader research showing that categorization influences perceptual processing of non-emotional stimuli

such as colors and consonants (Feldman, Griffiths, & Morgan, 2009). Fox, Moon, Iaria, and Barton (2009) also found that repetition suppression (a natural reduction in neural activity when the same psychological process is repeated; see Grill-Spector, Henson, & Martin, 2006) occurred in the fusiform face area and the posterior superior temporal sulcus only when participants repeatedly viewed faces that they perceived as representing the same emotion category, regardless of the actual expressions of the faces. This study shows that participants' categorical perceptions of faces are reflected in repetition-suppression-related changes in neural activity and that emotion concepts might shape even the fundamental perceptual signals used to process emotional cues.

Recently, Brooks and Freeman (2018) adopted an individual differences approach and found converging evidence from three studies showing that how people represent emotion concepts is related to how they perceive facial expressions of emotion. Participants provided a measure of their conceptual representations of emotions by rating how similarly they think a set of emotions are to each other. In two studies, participants then saw a series of facial expressions and had to drag the face to either a "correct" or "incorrect" emotion label (e.g., they dragged a prototypical sad expression either to the label "sad" or the label "surprised"). Interestingly, the path that participants dragged when identifying the emotions of faces was related to their conceptual representation of emotions. For example, participants who considered sadness and surprise to be conceptually similar also tended to drag a prototypical sad expression more toward the "surprised" label than participants who saw little conceptual overlap in these emotions. Similarly, a reverse-correlation paradigm (Dotsch & Todorov, 2012) revealed that participants who had strong conceptual overlap in emotion categories have internal representations of paradigmatic emotion expressions that are more similar to each other than those who have less conceptual overlap across emotion categories.

Finally, studies of surprised facial expressions provide an interesting testing ground for studying how conceptual processes shape emotion perception, as surprised expressions are affectively ambiguous and can be interpreted as expressing either positive or negative valence. Indeed, studies demonstrate that negative interpretations of surprised facial expressions are associated with increased amygdala activity and that priming negative emotion concepts before viewing surprised expressions increases amygdala reactivity to these faces (Kim et al., 2004), similar to what is observed in individuals who spontaneously tend to interpret surprise as negatively valenced (Kim, Somerville, Johnstone, Alexander, & Whalen, 2003). Relatedly, negative interpretations of surprise faces are thought to be relatively fast (Neta & Whalen, 2010). Allowing participants to engage in extended elaborative processing of surprised expressions increases the extent to which they are interpreted as positive rather than negative (Neta & Tong, 2016). Thus one's interpretation of the very same affective information presented in a surprised facial expression can be modulated by conceptual priming or elaborative processing, supporting the notion that conceptual forces influence how people parse affective information expressed by others.

Emotion Concepts and Emotion Experience

Studies in adults have demonstrated that conceptual processes influence how people parse their own affect into discrete categories. Lindquist and Barrett (2008) asked participants to write about an image that portrayed two men talking, one with a fear-like expression and the other with an anger-like expression. Participants either wrote about the man on the left (fear-prime condition), the man on the right (anger-prime condition), or both as if they were having a neutral conversation about nature (neutral-prime condition). Following the prime, participants either underwent a negative mood induction or a neutral mood induction. Even though all participants in the negative mood induction reported similar levels of affect (using a computerized circumplex rating scale), only those who had primed the concept of "fear" were less willing to perform a set of risky behaviors in the future. This suggested that priming the concept of fear affected how participants parsed the negative affect of the mood induction, consequently shifting their threshold for risky decision-making. Importantly, the fear prime did not affect decision-making in the neutral mood induction condition. Thus, the consequences of experiencing fear only occurred when participants were given both its concept and its affective "raw materials."

Functional neuroimaging data also suggest that emotion labels can shift how people parse their affective experiences. One study showed participants negative images and provided false feedback that the participants' "neural activity" revealed that they were feeling either fear, disgust, or fascination (Oosterwijk, Lindquist, Adebayo, & Barrett, 2015). Even though participants were instructed to not let this feedback shape their emotional experiences, they later reported feeling more interested when re-rating the images that were previously paired with the term "fascinated." The study by Satpute and colleagues described previously also involved a condition in which participants categorized (or continuously rated) their emotions in response to negative images (Satpute et al., 2016). Participants either rated their emotions on a continuous scale (from "bad" to "good") or categorically (as "bad," "neutral," or "good"). Categorizing one's own emotional responses to moderately negative images as "bad" was associated with increased activity in the amygdala and insula. Hence, parsing a moderate degree of one's own negative affect as "bad" was associated with increased activity in regions reliably involved in interoception and attention to one's own emotions (Craig, 2003; Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012; Zaki, Davis, & Ochsner, 2012).

Akin to the work of Lindquist and Barrett (2008), Brooks (2014) demonstrated that high arousal affective experiences associated with test-taking can be labeled either as "excitement" or "anxiety." In fact, instructing participants to label their affect as "excitement" before a host of evaluative tasks reliably boosted their performance. Others report similar results, although these researchers did not instruct participants to use specific emotion terms (Jamieson, Mendes, Blackstock, & Schmader, 2010). Thus, how we label affect has downstream consequences for our behavior, particularly in stressful situations. Interestingly, Lieberman and colleagues have found that merely pairing negative stimuli with negative words (called "affect

labeling") can serve a regulatory function (Torre & Lieberman, 2018). A growing body of work shows that affect labeling reduces psychophysiological (Tabibnia, Lieberman, & Craske, 2008), neural (Lieberman et al., 2007), behavioral (Kircanski, Lieberman, & Craske, 2012), and self-reported negative (Lieberman, Inagaki, Tabibnia, & Crockett, 2011) responses to aversive stimuli. However, studies have also produced inconsistent results, with affect labeling sometimes intensifying negative affect (Ortner, 2015). Thus, the mere act of using a linguistic concept to categorize one's affect appears to change the impact of that affective experience, potentially by reducing uncertainty about what one is feeling, increasing self-reflection, or making that affective experience more abstract (see Torre & Lieberman, 2018 for discussion).

Emotion Concept Development

As described in the previous section, research on emotion concepts has focused on the role they play in the experience and perception of emotions. However, much less attention has been paid to how these concepts emerge over the lifespan. Here, we highlight work characterizing when and how emotion concepts change over child and adolescent development. This work is rooted in the assumptions that emotion concepts (a) undergo ontogenetic transformation throughout childhood and adolescence and (b) emerge in lockstep with general developments of related cognitive processes that enable more complex, linguistically rich, and abstract conceptual representations. Wherever possible, these two points are elaborated in our discussion of relevant research, but specifically charting relations between emotional and non-emotional concept development remains a very interesting direction for future research.

While little work has explored the development of emotion concepts directly, previous research has reported on the developmental trajectories of constituent processes that support emotion perception and emotion experience. Data demonstrate that very young children (around 3.5 years of age) categorize facial expressions depicting a variety of emotions primarily into two categories corresponding to "positive" and "negative" (Widen, 2013). With increasing age, these researchers found a gradual separation of emotion categories. This expansion is reflected in data indicating an expansion in the use of specific emotion categories and concepts over the following few years (Widen, 2013). Hence, young children attend primarily to the valence of emotional expressions in others and learn to separate expressions based on other qualities as they develop.

Constructionist theories would posit that this developmental trend could be explained by an underlying development of emotion concepts. Specifically, children may represent emotion concepts primarily within a "positive vs. negative" dichotomy, where valence is the primary (or even only) dimension on which emotions are organized. However, this landscape may become richer with age, with emotion representations becoming more multidimensional. If so, this shift in emotion concept

representations should produce concomitant shifts in emotional experiences as well as emotional perception. Relatively little work has examined emotion concept representations themselves across development, but studies in which children were asked to sort emotion words and emotion faces into piles based on similarity have shown that (similar to adults) young children organize emotion concepts and emotion faces along valence and arousal dimensions (Bullock & Russell, 1984; Russell & Bullock, 1985; Russell & Ridgeway, 1983). However, these studies did not test whether the "weight" people place on the valence and arousal dimensions of emotions varies from childhood to adulthood. Such a test would provide insight into whether the circumplex representation of emotions might evolve from a "good vs. bad" dichotomy to more multidimensional representations.

Change in Emotion Concept Representation from Childhood to Adulthood

To address this gap, we conducted a study to empirically characterize age-related changes in emotion concept representations in a cross-sectional sample of children, adolescents, and adults aged 6–25 years (Nook, Sasse, Lambert, McLaughlin, & Somerville, 2017). First, we used an emotion vocabulary assessment in which participants verbally defined a set of emotion words (Nook et al., in press). Participants who could not demonstrate comprehension of emotion words used in later tasks were excluded from analyses. Participants then completed a task adapted from prior work (Barrett, 2004; Suvak et al., 2011) that assessed how people mentally organized emotion concepts. In this task, participants rated the degree to which the meaning of each pair of ten emotion words (e.g., fear and anger) was similar to or different from each other by sliding them nearer to or further from each other on a computer screen. This simple task permitted analyses (using multidimensional scaling methods) to derive the underlying semantic organization of the emotion words, with conceptually similar words spaced physically nearer to one another. The analysis also resolved the underlying dimensions upon which the emotion concepts were organized. The key question for this study was whether the use of the underlying dimensions varied systematically across age.

Results demonstrated that while children, adolescents, and adults organized their semantic knowledge about emotions consistently with the affective circumplex model (Posner, Russell, & Peterson, 2005; Russell, 1980), their reliance on the valence and arousal dimensions differed with age. With increasing age, individuals attended less to the valence dimension and more to the arousal dimension when organizing their semantic understanding of emotions. Although these linear effects were present across our full age range of 6–25, Fig. 2.1 provides a categorical visualization to clarify this finding. Children's emotion organization in the left panel exhibits a "wider" *y*-axis (indicating increased focus on the valence dimension) and a "flatter" *y*-axis (indicating reduced attention to the arousal dimension) compared

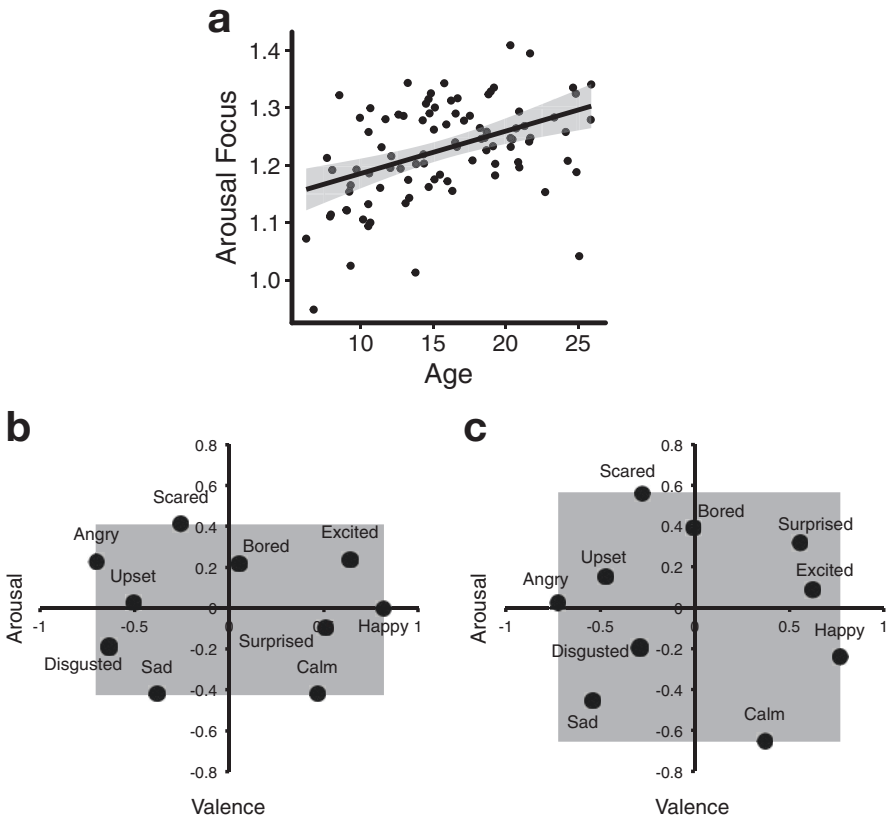

Fig. 2.1 Multidimensional semantic organization of emotions increases from childhood to adulthood. (**a**) Attention to the second dimension of emotions (arousal) increased linearly across age. (**b**) Average emotion representation in youngest ten children (mean age = 8.04 years, range = 6.24–9.22). (**c**) Average emotion representation in oldest ten young adults (mean age = 24.58 years, range = 22.73–25.91). Age-related change in emotion representations is evident in the expansion in focus on arousal within the definitional space of emotion terms, indicating expansion in multidimensional emotion representations. Although age-related differences in attention to the valence and arousal dimensions were continuous across the sample's age range, we present representations from young and old participants separately for illustration. Figure adapted from Nook, Sasse, et al. (2017)

to adults' organization in the right panel. Thus, these data showed that the semantic organization of emotion concepts becomes more multidimensional with age.

Because we posited that emotion concepts would build on the development of foundational cognitive processes, we additionally tested potential mediators of emotion concept development. The first potential mediator was general verbal knowledge. Recent theoretical and empirical work suggests that increases in one's verbal repertoire (Baron-Cohen, Golan, Wheelwright, Granader, & Hill, 2010; Farkas & Beron, 2004) may foster emotion concept development. As reviewed above, the presence of linguistic labels facilitates one's ability to learn new conceptual

distinctions, including distinctions between emotions (Fugate et al., 2010; Lupyan, 2016). Thus, increasing verbal knowledge across age may provide the linguistic footholds needed to create nuanced and distinct concepts for emotions, allowing people to expand from a valence-bound "positive vs. negative" emotional dichotomy. Indeed, research on the related phenomenon of emotion understanding (i.e., one's understanding of the myriad psychological processes involved in the production, experience, display, and regulation of emotions) shows that this skill is robustly related to verbal knowledge (Beck, Kumschick, Eid, & Klann-Delius, 2012; de Rosnay, Pons, Harris, & Morrell, 2004; De Stasio, Fiorilli, & Di Chiacchio, 2014; Pons, Harris, & de Rosnay, 2004). We tested participants' verbal knowledge using the Wechsler vocabulary assessment (Wechsler, 1991, 1999) as a potential mediator of age-related differences in multidimensional emotion representations.

Second, we tested whether general intellectual development—not verbal knowledge in particular—might drive emotion concept development. Classic measures of intellectual abilities including the Wechsler assessments (Wechsler, 1991, 1999) conceptualize verbal knowledge and fluid reasoning as separate components of intellectual ability. Fluid reasoning refers to the ability to flexibly deduce and apply rules to solve novel problems, and prior work suggests that this skill may also contribute to emotion concept development (De Stasio et al., 2014). Hence, we tested the specificity of the relationship between verbal knowledge and emotion concept development by also assessing how emotion concept development relates to fluid reasoning as assessed by the Wechsler indices of matrix reasoning (Wechsler, 1991, 1999).

Finally, the development of multidimensional emotion representation could also scaffold on development of general abilities to represent multiple dimensions simultaneously. Piagetian theory postulates that children tend to fixate on a single concrete perceptual dimension and neglect other dimensions, a phenomenon called centration (Piaget, 1952). Indeed, empirical studies demonstrate that people gradually develop an understanding that stimuli can have multiple dimensions as they age. For example, children tend to organize and represent animal species primarily in terms of their size, but these representations develop to include the more abstract dimensions of domesticity and predativity across adolescence and into adulthood (Howard & Howard, 1977). Hence, because representations of other stimulus classes (such as animals) grow increasingly multidimensional across development, it is possible that multidimensional emotion representations arise through a similar domain-general cognitive developmental process. To test this potential mediator, we also administered a control task to participants that assessed multidimensional representation of non-emotional cues (i.e., shapes that varied in both size and shading). After verifying that shading represented the second dimension of participants' representation of shapes in this control task, we quantified the focus placed on this second dimension (i.e., size) as an index of non-emotional multidimensionality.

As expected, all potential mediators, verbal knowledge, fluid reasoning, and shading focus (i.e., the weight placed on the second dimension in the perceptual similarities control task), increased with age. However, only verbal knowledge significantly mediated the relationship between age and arousal focus in the parallel

mediation analysis (Fig. 2.2). Therefore, while fluid reasoning and multidimensionality of semantic representations grow with age, developments in arousal focus are not merely a by-product of cognitive developments in these domains. Rather, they are likely a product of richer understanding of emotion concepts, supported by age-related increases in verbal abilities.

Altogether, this study demonstrated that emotion representations are not static across life. Instead, moving from childhood to adulthood, people develop an understanding that emotions are more than positive or negative and instead vary on multiple dimensions. As emotion concepts play a central role in both perceiving others' emotions and experiencing one's own emotions (reviewed above), children's tendency to lump emotion concepts within valences could influence their ability to make fine-grained distinctions between different emotion types.

These data also indicate that emotion representations have a prolonged developmental trajectory. Because most studies on the development of emotion perception, emotion experience, and emotion understanding are constrained to childhood, little is known about emotion conceptualization in adolescence. The results from this study revealed that there are continued changes in emotion conceptualization throughout late adolescence and into early adulthood, a finding that prompts new questions about the role of emotion concept development in the social and affective changes that occur during adolescence (Somerville, Jones, & Casey, 2010). Additionally, the fact that we found that emotion concept development scaffolded on general verbal knowledge supported emerging arguments on the role of language in emotion concept formation. However, more research is needed to determine the

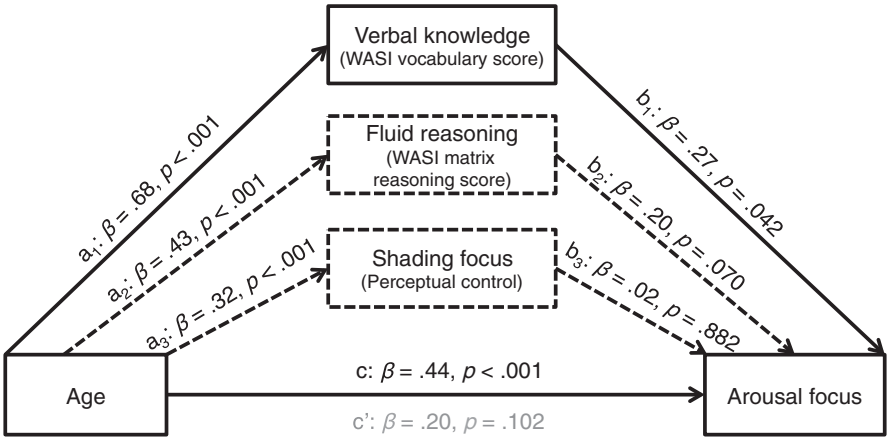

Fig. 2.2 Increasing verbal knowledge mediated increased arousal focus across age, over and above fluid reasoning and general abilities to represent two perceptual dimensions simultaneously. Adapted from Nook, Sasse, et al. (2017)

exact process by which general language development translates into multidimensional emotional concept representations. On the one hand, verbal development could play a "bootstrapping" role in giving children the conceptual footholds needed to learn subtle distinctions between abstract emotion types (Carey, 2011; Lupyan, 2012). It is also possible that an enhanced general awareness of the multitude of emotion terms (i.e., learning that English labels "annoyance," "frustration," and "aggravation" as emotional states that are separate from "anger") may expand concepts underlying emotion words already learned.

Change in Emotion Concepts and Emotion Experience from Childhood to Adulthood

As reviewed previously, there are normative changes in a variety of emotional experiences that are characteristic of developmental transitions between childhood, adolescence, and adulthood. However, there has been scant research evaluating how emotion concept development shapes characteristic emotional experiences across the lifespan. Here, we review an emerging line of research on a particular aspect of emotional experiences which is thought to be tightly linked to emotion concepts—emotion differentiation.

Emotion differentiation (also called emotion granularity) refers to how specifically people experience their emotions. This construct is a natural corollary of the constructionist theory. If emotions occur when people parse inherently ambiguous affect into discrete types using emotion concepts, then an individual's "emotional spectrum" will be influenced by the emotion concepts they have at their disposal. Kashdan, Barrett, and McKnight (2015) described individual differences in emotion differentiation by quoting two responses to the September 11th attacks on New York City. An individual likely high in emotion differentiation wrote: "My first reaction was terrible sadness. … But the second reaction was that of anger, because you can't do anything with the sadness. I felt a bunch of things I couldn't put my finger on. Maybe anger, confusion, fear." By contrast, a reaction indicating low emotion differentiation example was: "I just felt bad on September 11th. Really bad." Thus, emotion granularity captures the specificity with which individuals conceptualize their affect according to discrete emotion types. Greater differentiation is evident when an individual assembles a unique combination of specific emotional experiences in response to an emotional provocation; conversely, low differentiation is associated with constructing a more general affective experience (i.e., "bad") that is less specific across instances.

Empirical investigations of emotion differentiation have focused on how individual differences in people's ability to separate their emotional states relate to other trait-level measures. These studies originally used experience sampling methods to measure emotion differentiation (Barrett, Gross, Christensen, & Benvenuto, 2001; Kashdan, Ferssizidis, Collins, & Muraven, 2010; Tugade, Fredrickson, & Barrett,

2004). Participants repeatedly rated how much they felt a series of emotions (e.g., anger, disgust, fear, sadness, disappointment, frustration) over several days, and intraclass correlations were used to quantify how specifically participants separated their affective states across ratings. If participants consistently rated all negative emotions as either high (on "bad" days) or low (on "good" days), intraclass correlations between these emotions would be high. This would indicate low emotion differentiation, as participants were not parsing their affect into specific emotions, but rather reported feeling all (or none) within that valence. By contrast, if participants tended to feel different combinations of emotions at each rating (i.e., ratings for each emotion type seemed to move independently of each other), intraclass correlations between these emotions would be low, indicating high emotion differentiation. Scholars have since used repeated ratings of negative images in a short lab session to compute measures of emotion differentiation that largely replicate findings gathered through much longer and more expensive experience sampling methods (Erbas, Ceulemans, Lee Pe, Koval, & Kuppens, 2014).

These studies repeatedly show that higher emotion differentiation is associated with a host of positive mental health benefits (see Kashdan et al., 2015 for a review). For example, high emotion differentiation is associated with increased use of emotion regulation (Barrett et al., 2001), reduced drinking to cope (Kashdan et al., 2010), reduced aggression when angry (Pond et al., 2012), and reduced self-harm in people with borderline personality disorder (Zaki, Coifman, Rafaeli, Berenson, & Downey, 2013). Low emotion differentiation has also been identified in people with depression (Demiralp et al., 2012), social anxiety disorder (Kashdan & Farmer, 2014), autism (Erbas, Ceulemans, Boonen, Noens, & Kuppens, 2013), and schizophrenia (Kimhy et al., 2014). Scholars interpret these results as evidence that emotion differentiation may facilitate emotion regulation (Kashdan et al., 2015), but the specific mechanisms of this process have not been defined. For example, being able to specifically identify one's emotions might boost psychological well-being by allowing people to implement adaptive emotion regulation strategies that are ideal for the situation at hand (e.g., using mindful acceptance when frustrated by uncontrollable traffic but using interpersonal problem-solving skills when disappointed by an employee's poor work). In line with this reasoning, evidence suggests that cognitive reappraisal strategies may only be helpful in situations where controllability over stressors is low (Troy, Ford, McRae, Zarolia, & Mauss, 2017; Troy, Shallcross, & Mauss, 2013). However, many other possibilities might explain the relations between high emotion differentiation and well-being (e.g., general intelligence may be a third variable), so future work is needed to adjudicate between possible mechanisms. Additionally, it must be highlighted that all of these studies are correlational (or quasi-experimental) in nature, and so directionality and causality for the relationship between emotion differentiation and outcomes cannot be ascertained from these data alone.

Prior research could support two hypotheses concerning the development of emotion differentiation. Converging evidence presented in the previous section indicates that children represent their own and others' emotions within a broad "positive" vs. "negative" dichotomy and thus may struggle to make fine-grained

distinctions between emotions within each valence. In fact as described above, children's emotion concepts are strongly focused on valence, and this focus shifts to other dimensions (i.e., arousal) through adolescence and into adulthood (Nook, Sasse, et al., 2017). This increasing complexity in emotion representation might contribute to greater emotion differentiation with age. These findings motivate the hypothesis that emotion differentiation may increase from childhood to adulthood as emotion representations shift from a broad valence dichotomy to more specific emotion concepts that are highly differentiable.

A competing hypothesis is that emotion differentiation follows a quadratic trajectory such that it reaches a nadir in adolescence. This hypothesis is based on the finding that children not only report an absence of mixed emotions, they also struggle to understand that emotions can co-occur (Harter & Buddin, 1987; Wintre & Vallance, 1994). For example, children expect people to feel either angry or sad, not both angry and sad. Interestingly, reporting only one emotion at a time is one "route" to high emotion differentiation, as it involves specifically identifying one individual emotion that is being experienced. For example, children would experience sadness and anger as discrete and differentiated experiences, precisely because they do not co-occur. Even if children conceptualize these emotions more similarly to each other than adults do, childhood may be a period of high emotion differentiation if children's tendency to report feeling only one emotion at a time "trumps" their underlying conceptual similarity.

Thus, emotion differentiation may decrease from childhood to adolescence as children shift away from experiencing emotions as mutually exclusive. Adolescence would be a period of low emotion differentiation in which emotions co-occur at greater frequency (Harter & Buddin, 1987; Wintre & Vallance, 1994). However, because emotion concepts continue to become more refined from adolescence to adulthood, emotion differentiation may rise within this period as young adults learn to separate co-experienced emotions using increasingly defined emotion concepts. Hence, adults may also have high emotion differentiation but through a different "route" than children (i.e., because they can specifically identify emotions, including those that occur simultaneously). These two developmental processes (i.e., reduced single emotion experience from childhood to adolescence and increased familiarity parsing co-experienced emotions from adolescence to adulthood) would ultimately result in a quadratic relationship between age and emotion differentiation.

We conducted a study to chart the development of negative emotion differentiation using a standardized emotion differentiation laboratory task in a cross-sectional sample of individuals aged 5–25 years (Nook, Sasse, Lambert, McLaughlin, & Somerville, 2018). Participants completed a task in which they viewed mildly aversive images to evoke negative affect, and on each image, they were asked to rate the degree to which they experienced five specific negative emotions: angry, disgusted, sad, scared, and upset. Images were selected to depict a wide variety of negative scenes that would be tolerable even to young children. Again, data from participants who failed to show that they comprehended emotion words used in this task were excluded from analyses. Following work described above, intraclass

correlations were used to quantify how specifically participants separated their affective states across images. Lower values indicated that participants experienced each emotion as a unique type across trials and thus were able to make fine-grained distinctions between affective instances, interpreted as higher granularity (Fig. 2.3).

In addition to adjudicating between the linear and non-linear trajectories outlined above, we assessed two potential cognitive mechanisms that could explain age-related changes in emotion differentiation: *average emotion intensity*, operationalized as the average level of endorsement of all emotion categories, and the tendency to experience emotions one at a time (called *single emotion experience*), operationalized as a larger difference in ratings between the most endorsed emotion term and the others. These calculations, based on the primary task ratings, are separate from the primary dependent variable (intraclass correlations).

Results demonstrated that emotion differentiation (reverse-scored ICC across emotion ratings) exhibited a quadratic relationship with age: it decreased from childhood to adolescence and increased from adolescence to adulthood. The age of nadir was estimated to be mid-adolescence (15.77 years) (Fig. 2.4).

We then examined the degree to which the two mediators described above (average emotion intensity and single emotion experience) could explain the inverted-U

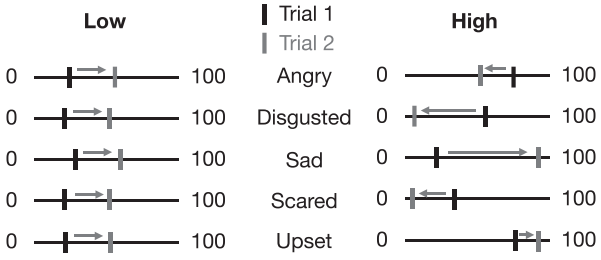

Fig. 2.3 Negative emotion granularity (i.e., differentiation) was computed by reverse-scoring the intraclass correlation of emotion ratings across several trials. Two trials are displayed to clarify that this measure depends on intercorrelations between emotion ratings across trials. Adapted from Nook et al. (2018)

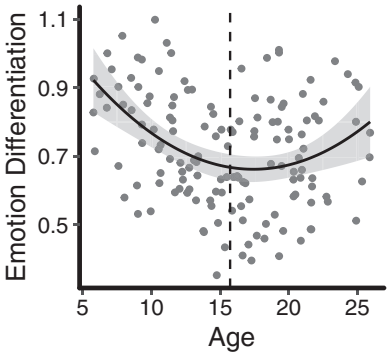

Fig. 2.4 Negative emotion differentiation exhibited a significant quadratic (solid line) relationship with age. A separate analysis (Simonsohn, 2017) revealed that the change point occurred at age 15.77 (vertical dashed line). Adapted from Nook et al. (2018)

relationship observed with emotion differentiation. First, the overall intensity of affective response did not vary by age, which disqualified it from tests of mediation. Conversely, single emotion experience did significantly mediate the decrease in emotion granularity from childhood to adolescence. However, this mediator did not explain increased emotion differentiation from adolescence to adulthood.

Contrary to the hypothesis that emotion differentiation increases across development, this study revealed that emotion differentiation is high in early childhood. Even though children tend to place excess emphasis on whether emotions are merely positive or negative (Nook, Sasse, et al., 2017; Pons et al., 2004; Widen, 2013), their tendency to report experiencing emotions in isolation (i.e., in a mutually exclusive fashion; Wintre & Vallance, 1994) "trumps" this reduced multidimensionality of emotion concepts and ultimately yields differentiated emotional experiences. However, experiencing emotions one at a time represents a different "route" to high emotion granularity than the ability to differentiate several emotions that occur simultaneously. Hence, our results not only demonstrate that emotion granularity has a non-linear developmental trajectory, they also reveal that the predominant route to emotion differentiation varies depending on one's developmental stage.

One important question that these data illuminate is why children tend to report experiencing one emotion at a time. Drawing on the constructionist theory of emotion (Barrett, 2006), this may occur either because children's core affective experiences are naturally parceled into discrete types or because children apply a single emotion label to categorize their experienced affect. The first explanation seems unlikely because robust evidence suggests that core affect—one's internal somatic and physiological sensations—does not share a one-to-one mapping with specific emotion types (Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 2000). Instead, children may either believe emotions can only occur in a singular fashion (leading them to apply only a single emotion concept at a time) or they lack the ability to represent their core affect as fitting multiple emotion concepts simultaneously (Hoemann, Gendron, & Barrett, 2017). Hence, one possibility is that children are still developing the psychological foundations for representing multiple co-occurring emotions. Future work is needed to investigate this hypothesis more directly. The extent to which this developmental process scaffolds on non-emotional cognitive processes (i.e., a general tendency to assign experiences or stimuli to only one category) should be investigated.

Intriguingly, our data also suggest that adolescence is a period in which emotions co-occur with greater frequency, but these emotions are poorly differentiated. In concert with the studies summarized at the beginning of this chapter, this finding contributes to basic understandings of the socioemotional changes that arise during adolescence (Somerville & McLaughlin, 2018). However, at a more applied level, the novel experience of simultaneous emotions during adolescence could produce meta-emotions such as confusion or potentially interfere with effective emotion regulation, as adolescents struggle to select optimal strategies for regulating simultaneously experienced emotions. Although speculative, our results suggest that low emotion differentiation—which has previously been associated with psychopathology (Kashdan et al., 2015)—may be one factor that contributes to the increased

onset of mental illness in adolescence (Kessler, Berglund, Demler, Jin, & Walters, 2005; Lee et al., 2014). Difficulty applying emotion concepts to parse ambiguous affect (Nook et al., 2015) may contribute to the spike in psychopathology that occurs in adolescence. In line with this notion, prior work demonstrates that a high incidence of co-occurring emotions is associated with non-suicidal self-injury in adolescents (Andrewes, Hulbert, Cotton, Betts, & Chanen, 2017), whereas high emotion differentiation protects people with borderline personality disorder from self-injury (Zaki et al., 2013). However, future research is needed to empirically evaluate whether low emotion differentiation in adolescence contributes to increased risk for the onset of mental illness in this developmental period.

Emotion Concept Development: Looking Ahead

While the research described in the preceding sections suggests that emotion concepts underpin key facets of emotional development from childhood to adulthood, there remains a host of open questions that will need to be addressed before arriving at a comprehensive understanding of emotion concept development. Here we highlight some pressing questions for expanding our understanding of emotion concept development.

Do Emotion Words Hold the Same Meaning to Children, Adolescents, and Adults?

Research described thus far examines how emotion concepts are organized and used to parse affective states across development. In these studies, emotions are measured using verbal labels such as anger, fear, and happiness. However, it is plausible that these emotion terms themselves hold different meanings to individuals depending on their age. While the studies on emotion concepts that use verbal labels described in the previous section ensured that participants had a basic understanding of each word used in the experiment, it is possible—and likely—that the manner in which individuals conceptualize the meaning of an emotion changes from childhood to adulthood. In particular, individuals could differ in how abstractly they conceptualize emotions. For example, one could define anger as an aversive state provoked by being the target of injustice, or one could define anger as what is felt when someone steals your toy. Neither definition is wrong, as comprehending an emotion word means being able to connect that word with a culturally agreed-upon concept of what characterizes or defines that emotion (Bloom, 2000; Yin & Csibra, 2015). However, it is obvious that these two meanings differ in their level of abstraction. Psychologists have long known that any given situation can be represented either concretely (i.e., with attention to low-level physical, observable, and situationally bound details) or abstractly (i.e., with attention to higher-level principles

that generalize beyond specific situations; Trope & Liberman, 2000). However, whether emotions are likewise conceptualized concretely or abstractly has received scant attention in developmental, affective, or cognitive theory.

Emotion concept abstraction may thus vary across individuals and developmental stages. In fact, both verbal and non-verbal abstraction abilities develop from childhood to adulthood (Crone et al., 2009; Dumontheil, 2014; Ferrer, O'Hare, & Bunge, 2009; Joelson & Herrmann, 1978; Ponari, Norbury, & Vigliocco, 2018), meaning it is likely that the conceptual representations underlying emotion terms similarly become more abstract with age. Such a progression would align with classic Piagetian theories positing that development proceeds from a concrete sensorimotor focus to a more abstract hypothetico-deductive focus (Demetriou et al., 2018; Inhelder & Piaget, 1958). Thus, the very emotion concepts we seek to understand are likely evolving in their meaning over age, and this evolution could contribute to the broad developmental differences we observe in emotion experiences. Future work obtaining and quantifying data about individual's conceptual understanding of emotion terms is needed to address this possibility and its implications for emotion perception and experience (see recent investigation by Nook et al., in press).

As such, researchers should pay added attention to the content of participants' emotion representations. Methods that capture participants' actual definitions of emotions would be helpful (similar to what was used in Nook, Sasse, et al., 2017; Nook et al., 2018). Although we speculate about developments in emotion abstraction here, there are myriad understudied aspects of emotion definition development that merit further study. Other aspects include (1) the "sociality" of emotion concepts (i.e., how tied they are to interpersonal situations) or (2) the valence vs ambivalence of emotion concepts (i.e., how separately they see "negative" and "positive" emotions or how much they consider the fact that emotions can feel both negative and positive).

How Does Emotion Concept Development Relate to Emotion Regulation?

A potentially impactful extension to basic research on emotion concept development is to evaluate how emotion concepts influence emotion regulation. As described above, emotion concepts and emotion regulation intersect, as applying an alternative concept (such as reconceptualizing nervousness as excitement) can alter emotion experiences in desirable ways (Brooks, 2014; Jamieson et al., 2010), effectively functioning as an emotion regulation strategy (Gross & Barrett, 2011). However, there is a dearth of research on how developmental changes in emotion concepts shape what emotion regulation strategies are chosen and how effective these strategies are in modulating emotion.

That said, a parallel stream of research has examined the development of emotion regulation efficacy, largely observing that emotion regulation abilities improve

with age. One effective technique for regulating emotions—called *cognitive reappraisal*—involves changing the meaning of an affective stimulus in a way that renders your response more consistent with your goals (Gross, 1998). For example, critical feedback from a parent or peer can be interpreted as helpful guidance for improvement rather than an indication that one is flawed. Substantial evidence demonstrates that cognitive reappraisal is an effective method for regulating emotions (Gross, 1998; Ochsner & Gross, 2005) and that greater use of cognitive reappraisal is associated with the absence of several forms of psychopathology (Aldao, Nolen-Hoeksema, & Schweizer, 2010).

However, evidence from several studies demonstrates that the beneficial effects of cognitive reappraisal might be less accessible to younger individuals, due to ongoing development of reappraisal efficacy. Some studies have demonstrated that the effectiveness of reappraisal in decreasing negative affect improves from childhood to young adulthood, especially when the affective cues contain social content involving negative interactions between people (McRae et al., 2012; Silvers et al., 2012, 2017; Silvers, Shu, Hubbard, Weber, & Ochsner, 2015). These findings converge with neurodevelopmental evidence that brain areas that are crucial for supporting cognitive reappraisal (Buhle et al., 2014) undergo protracted development through this age window (Ahmed, Bittencourt-Hewitt, & Sebastian, 2015). Interestingly however, not all studies have found age-related improvements in cognitive reappraisal success across childhood and adolescence, either when downregulating negative affect in response to aversive vignettes or images (Ahmed, Somerville, & Sebastian, 2018; Van Cauwenberge, Van Leeuwen, Hoppenbrouwers, & Wiersema, 2017) or when downregulating craving using reappraisal in response to appetizing foods (Giuliani & Pfeifer, 2015; Silvers et al., 2014). This suggests that under some conditions, reappraisal is effective even for children; more research is needed to specify the particular affective challenges for which even young individuals can benefit from efficacious reappraisal.

Given emotion regulation ability is improving (at least under some conditions) with age, what role does emotion concept development play in this emerging ability? Although research has not yet directly addressed this question, interesting linkages have been made between emotion regulation and the highly related domain of emotion language development. Decades of empirical research demonstrate that a child's general and emotional vocabularies are related to their ability to manage distressing situations, as well as their executive functioning, mental health, social likability, and academic outcomes (Cole, Armstrong, & Pemberton, 2010; Fabes, Eisenberg, Hanish, & Spinrad, 2001; Kuhn, Willoughby, Vernon-Feagans, & Blair, 2016; Roben, Cole, & Armstrong, 2013; Salmon, O'Kearney, Reese, & Fortune, 2016; Trentacosta & Izard, 2007). Hence, developing a functional emotion lexicon (possibly reflecting, or fostering, the development of emotion concepts) appears to be an important ingredient to overall well-being.

Combining this thinking with the ideas in the previous subsection, it is possible that emotion regulation ability might be related to emotion abstraction (i.e., the ability to represent an emotional definition in general terms, outside of concrete here-and-now situations). Indeed, prior work demonstrates that (1) psychological

distancing involves representing stimuli at higher levels of abstraction (Liberman & Förster, 2009; Soderberg, Callahan, Kochersberger, Amit, & Ledgerwood, 2015) and (2) psychological distancing facilitates emotion regulation (Ayduk & Kross, 2010; Kross et al., 2014; Nook, Schleider, & Somerville, 2017) Combining these ideas prompts the interesting possibility that the ability to view one's emotional experiences abstractly facilitates certain strategies for effective emotion regulation. The development of emotion abstraction could thus facilitate the beneficial effects of a variety of emotion regulation strategies, a potentially fruitful topic for future research.

How Does Emotion Concept Development Relate to Psychopathology and Its Treatment?

Implications of emotion concept development for mental health have been referenced throughout this chapter. However, at their base, all of these connections are speculative, and clear mechanistic accounts of how developmental differences in emotion concepts relate to psychopathology or psychological resilience have not been clearly established. For example, emotion differentiation correlates with the absence of mental illness (Kashdan et al., 2015) and is low in adolescence (Nook et al., 2018), but we do not know if low emotion differentiation in adolescence might explain increased risk for psychopathology in this age range (Kessler et al., 2005). Likewise, we know that verbal knowledge is associated with both social and psychological well-being (Salmon et al., 2016) and that verbal knowledge mediates the development of multidimensional emotion representations (Nook, Sasse, et al., 2017), but we do not know if (or how) more refined emotion concepts might facilitate resilience and well-being across the lifespan.

Answering these questions poses several interesting methodological challenges. Historically, measuring emotions has been a difficult challenge, but the proliferation of laboratory tasks (some of which are described above) as well as advances in experience sampling methods (in which participants provide real-time insight about their momentary experiences; e.g., Kalokerinos, Résibois, Verduyn, & Kuppens, 2017; Wu et al., 2017) has made valuable progress against this problem. A greater challenge is establishing causality in relations between emotion concepts and mental illness. Given that the link between emotion conceptualization and mental health is primarily correlational in nature, longitudinal and experimental approaches could provide an extra level of insight into how exactly emotion concepts relate to mental health and rule out the possibility of confounding variables.

Connecting emotion concept development to psychopathology is a particularly exciting direction for future research for several reasons. Mental illness across the lifespan constitutes a severe source of personal and economic burden worldwide (Patel, Flisher, Hetrick, & McGorry, 2007). Thus, extending our understanding of how exactly emotion concepts relate to psychopathology could germinate tools for

intervening to reduce this burden. Although shaping patients' understanding of emotions is a central part of several psychological treatments such as cognitive-behavior therapy (Beck, 2011), the Unified Protocol (Barlow et al., 2017), and dialectical behavior therapy (Linehan et al., 1999), we do not have a mechanistic understanding of how these interventions function or which components of these interventions are most impactful. Hence, advancing our understanding in this domain could help bolster the impact of tailoring already existing treatments to the emotional abilities of individuals of different ages.

Additionally, because psychotherapists largely use language to treat patients, this line of research focuses on a zone in which affective, linguistic, cognitive, and clinical questions intersect. If emotion language development is ongoing throughout adolescence, this motivates the need to constrain and shape language used during therapy to match the developmental stage of the individual in treatment. Thus, greater insight into how therapists use their language to enhance their patients' emotional well-being represents a potent testing ground for exploring questions that are interesting to basic and applied scientists across disciplines.

Conclusion

Here, we have argued that emotion concepts underpin emotion perceptions and emotional experiences. An emerging area of research concerning emotion concept development is beginning to reveal how the representation and application of emotion concepts continues to evolve through childhood and adolescence. Moreover, these developments have a tangible impact on emotional experiences across the lifespan and could relate to age-specific health risk factors including risks to mental health that evolve over childhood and adolescence.

References

Adolphs, R. (2017). How should neuroscience study emotions? By distinguishing emotion states, concepts, and experiences. *Social Cognitive and Affective Neuroscience, 12*, 24–31.

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature, 433*(7021), 68–72.

Ahmed, S. P., Bittencourt-Hewitt, A., & Sebastian, C. L. (2015). Neurocognitive bases of emotion regulation development in adolescence. *Developmental Cognitive Neuroscience, 15*, 11–25.

Ahmed, S. P., Somerville, L. H., & Sebastian, C. L. (2018). Using temporal distancing to regulate emotion in adolescence: Modulation by reactive aggression. *Cognition and Emotion, 32*(4), 812–826.

Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review, 30*(2), 217–237.

Andrewes, H. E., Hulbert, C. A., Cotton, S. M., Betts, J., & Chanen, A. M. (2017). An ecological momentary assessment investigation of complex and conflicting emotions in youth with borderline personality disorder. *Psychiatry Research, 252*, 102–110.

Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., … Bentin, S. (2008). Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychological Science, 19*, 724–732.

Ayduk, Ö., & Kross, E. (2010). From a distance: Implications of spontaneous self-distancing for adaptive self-reflection. *Journal of Personality and Social Psychology, 98*, 809–829.

Bailen, N. H., Green, L. M., & Thompson, R. J. (2018). Understanding emotion in adolescents: A review of emotional frequency, intensity, instability, and clarity. *Emotion Review*, *11* (advanced online publication).

Barlow, D. H., Farchione, T. J., Sauer-Zavala, S., Latin, H. M., Ellard, K. K., Bullis, J. R., … Cassiello-Robbins, C. (2017). *Unified protocol for transdiagnostic treatment of emotional disorders: Therapist guide*. New York, NY: Oxford University Press.

Barnes, G. M., Hoffman, J. H., Welte, J. W., Farrell, M. P., & Dintcheff, B. A. (2007). Adolescents' time use: Effects of substance use, delinquency and sexual activity. *Journal of Youth and Adolescence, 36*, 697–710.

Baron-Cohen, S., Golan, O., Wheelwright, S., Granader, Y., & Hill, J. (2010). Emotion word comprehension from 4 to 16 years old: A developmental survey. *Frontiers in Evolutionary Neuroscience, 2*, 109.

Barrett, L. F. (2004). Feelings or words? Understanding the content in self-report ratings of experienced emotion. *Journal of Personality and Social Psychology, 87*, 266–281.

Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review, 10*, 20–46.

Barrett, L. F., Gross, J. 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 & Emotion, 15*, 713–724.

Beck, J. S. (2011). *Cognitive behavior therapy: Basics and beyond*. New York, NY: Guilford Press.

Beck, L., Kumschick, I. R., Eid, M., & Klann-Delius, G. (2012). Relationship between language competence and emotional competence in middle childhood. *Emotion, 12*, 503–514.

Black, S. R. (2004). A review of semantic satiation. *Advances in Psychology Research, 26*, 95–106.

Blakemore, S. J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience, 9*, 267–277.

Blakemore, S. J., Burnett, S., & Dahl, R. E. (2010). The role of puberty in the developing adolescent brain. *Human Brain Mapping, 31*(6), 926–933.

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Brooks, A. W. (2014). Get excited: Reappraising pre-performance anxiety as excitement. *Journal of Experimental Psychology: General, 143*, 1144–1158.

Brooks, J. A., & Freeman, J. B. (2018). Conceptual knowledge predicts the representational structure of facial emotion perception. *Nature Human Behaviour, 2*(8), 581–591.

Bruner, J. S., Postman, L., & Rodrigues, J. (1951). Expectation and the perception of color. *The American Journal of Psychology, 64*, 216–227.

Buhle, J. T., Silvers, J. A., Wager, T. D., Lopez, R., Onyemekwu, C., Kober, H., … Ochsner, K. N. (2014). Cognitive reappraisal of emotion: A meta-analysis of human neuroimaging studies. *Cerebral Cortex, 24*, 2981–2990.

Bullock, M., & Russell, J. A. (1984). Preschool children's interpretation of facial expressions of emotion. *International Journal of Behavioral Development, 7*, 193–214.

Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., & Ito, T. A. (2000). The psychophysiology of emotion. In R. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 173–191). New York, NY: Guilford Press.

Cairns, R. B., Leung, M.-C., Buchanan, L., & Cairns, B. D. (1995). Friendships and social networks in childhood and adolescence: Fluidity, reliability, and interrelations. *Child Development, 66*, 1330–1345.

Carey, S. (2011). *The origin of concepts*. New York, NY: Oxford University Press.

Carroll, J. M., & Russell, J. (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of Personality and Social Psychology, 70*, 205–218.

Cole, P. M., Armstrong, L. M., & Pemberton, C. K. (2010). The role of language in the development of emotion regulation. In S. D. Calkins & M. A. Bell (Eds.), *Human brain development. Child development at the intersection of emotion and cognition* (pp. 59–77). Washington, DC: American Psychological Association.

Cole, P. M., Tan, P. Z., Hall, S. E., Zhang, Y., Crnic, K. A., Blair, C. B., & Li, R. (2011). Developmental changes in anger expression and attention focus: Learning to wait. *Developmental Psychology, 47*(4), 1078–1089.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407–428.

Costafreda, S. G., Brammer, M. J., David, A. S., & Fu, C. H. Y. (2007). Predictors of amygdala activation during the processing of emotional stimuli: A meta-analysis of 385 PET and fMRI studies. *Brain Research Reviews, 58*, 57–70.

Craig, A. D. (2003). Interoception: The sense of the physiological condition of the body. *Current Opinion in Neurobiology, 13*, 500–505.

Crone, E. A., Wendelken, C., van Leijenhorst, L., Honomichl, R. D., Christoff, K., & Bunge, S. A. (2009). Neurocognitive development of relational reasoning. *Developmental Science, 12*, 55–66.

Csikszentmihalyi, M., & Larson, R. (1984). *Being adolescent*. New York, NY: Basic Books.

de Rosnay, M., Pons, P., Harris, P. L., & Morrell, J. M. B. (2004). A lag between understanding false belief and emotion attribution in young children: Relationships with linguistic ability and mothers' mental-state language. *British Journal of Developmental Psychology, 22*, 197–218.

De Stasio, S., Fiorilli, C., & Di Chiacchio, C. (2014). Effects of verbal ability and fluid intelligence on children's emotion understanding. *International Journal of Psychology, 49*, 409–414.

Demetriou, A., Makris, N., Spanoudis, G., Kazi, S., Shayer, M., & Kazali, E. (2018). Mapping the dimensions of general intelligence: An integrated differential-developmental theory. *Human Development, 61*, 4–42.

Demiralp, E., Thompson, R. J., Mata, J., Jaeggi, S. M., Buschkuehl, M., Barrett, L. F., … Jonides, J. (2012). Feeling blue or turquoise? Emotional differentiation in major depressive disorder. *Psychological Science, 23*, 1410–1416.

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science, 3*(5), 562–571.

Dumontheil, I. (2014). Development of abstract thinking during childhood and adolescence: The role of rostrolateral prefrontal cortex. *Developmental Cognitive Neuroscience, 10*, 57–76.

Erbas, Y., Ceulemans, E., Boonen, J., Noens, I., & Kuppens, P. (2013). Emotion differentiation in autism spectrum disorder. *Research in Autism Spectrum Disorders, 7*(10), 1221–1227.

Erbas, Y., Ceulemans, E., Lee Pe, M., Koval, P., & Kuppens, P. (2014). Negative emotion differentiation: Its personality and well-being correlates and a comparison of different assessment methods. *Cognition & Emotion, 28*, 1196–1213.

Fabes, R. A., Eisenberg, N., Hanish, L. D., & Spinrad, T. L. (2001). Preschoolers' spontaneous emotion vocabulary: Relations to likability. *Early Education & Development, 12*, 11–27.

Farkas, G., & Beron, K. (2004). The detailed age trajectory of oral vocabulary knowledge: Differences by class and race. *Social Science Research, 33*, 464–497.

Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review, 116*(4), 752–782.

Ferrer, E., O'Hare, E. D., & Bunge, S. A. (2009). Fluid reasoning and the developing brain. *Frontiers in Neuroscience, 3*, 46–51.

Flook, L. (2011). Gender differences in adolescents' daily interpersonal events and well-being. *Child Development, 82*(2), 454–461.

Forbes, E. E., & Dahl, R. E. (2010). Pubertal development and behavior: Hormonal activation of social and motivational tendencies. *Brain and Cognition, 72*, 66–72.

Fox, C. J., Moon, S. Y., Iaria, G., & Barton, J. J. S. (2009). The correlates of subjective perception of identity and expression in the face network: An fMRI adaptation study. *NeuroImage, 44*, 569–580.

Fugate, J. M. B., Gouzoules, H., & Barrett, L. F. (2010). Reading chimpanzee faces: Evidence for the role of verbal labels in categorical perception of emotion. *Emotion, 10*, 544–554.

Gendron, M., Lindquist, K. A., Barsalou, L., & Barrett, L. F. (2012). Emotion words shape emotion percepts. *Emotion, 12*(2), 314–325.

Giuliani, N. R., & Pfeifer, J. H. (2015). Age-related changes in reappraisal of appetitive cravings during adolescence. *NeuroImage, 108*, 173–181.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences, 10*, 14–23.

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*, 271–299.

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.

Gunnar, M. R., Wewerka, S., Frenn, K., Long, J. D., & Griggs, C. (2009). Developmental changes in hypothalamus-pituitary-adrenal activity over the transition to adolescence: Normative changes and associations with puberty. *Development and Psychopathology, 21*, 69–85.

Halberstadt, J. (2003). The paradox of emotion attribution: Explanation biases perceptual memory for emotional expressions. *Current Directions in Psychological Science, 12*, 197–201.

Halberstadt, J. (2005). Featural shift in explanation-biased memory for emotional faces. *Journal of Personality and Social Psychology, 88*, 38–49.

Halberstadt, J., & Niedenthal, P. (2001). Effects of emotion concepts on perceptual memory for emotional expressions. *Journal of Personality and Social Psychology, 81*, 587–598.

Harter, S., & Buddin, B. J. (1987). Children's understanding of the simultaneity of two emotions: A five-stage developmental acquisition sequence. *Developmental Psychology, 23*, 388–399.

Hassin, R. R., Aviezer, H., & Bentin, S. (2013). Inherently ambiguous: Facial expressions of emotions, in context. *Emotion Review, 5*, 60–65.

Hoemann, K., Gendron, M., & Barrett, L. F. (2017). Mixed emotions in the predictive brain. *Current Opinion in Behavioral Sciences, 15*, 51–57.

Howard, D. V., & Howard, J. H. (1977). A multidimensional scaling analysis of the development of animal names. *Developmental Psychology, 13*, 108–113.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures*. New York, NY: Basic Books.

Jamieson, J. P., Mendes, W. B., Blackstock, E., & Schmader, T. (2010). Turning the knots in your stomach into bows: Reappraising arousal improves performance on the GRE. *Journal of Experimental Social Psychology, 46*, 208–212.

Joelson, J. M., & Herrmann, D. J. (1978). Properties of categories in semantic memory. *The American Journal of Psychology, 91*, 101.

Kalokerinos, E. K., Résibois, M., Verduyn, P., & Kuppens, P. (2017). The temporal deployment of emotion regulation strategies during negative emotional episodes. *Emotion, 17*(3), 450.

Kashdan, T. B., Barrett, L. F., & McKnight, P. E. (2015). Unpacking emotion differentiation: Transforming unpleasant experience by perceiving distinctions in negativity. *Current Directions in Psychological Science, 24*, 10–16.

Kashdan, T. B., & Farmer, A. S. (2014). Differentiating emotions across contexts: Comparing adults with and without social anxiety disorder using random, social interaction, and daily experience sampling. *Emotion, 14*, 629–638.

Kashdan, T. B., Ferssizidis, P., Collins, R. L., & Muraven, M. (2010). Emotion differentiation as resilience against excessive alcohol use: An ecological momentary assessment in underage social drinkers. *Psychological Science, 21*, 1341–1347.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62*(6), 593–602.

Kim, H., Somerville, L. H., Johnstone, T., Alexander, A. L., & Whalen, P. J. (2003). Inverse amygdala and medial prefrontal cortex responses to surprised faces. *Neuroreport, 14*(18), 2317–2322.

Kim, H., Somerville, L. H., Johnstone, T., Polis, S., Alexander, A. L., Shin, L. M., & Whalen, P. J. (2004). Contextual modulation of amygdala responsivity to surprised faces. *Journal of Cognitive Neuroscience, 16*(10), 1730–1745.

Kimhy, D., Vakhrusheva, J., Khan, S., Chang, R. W., Hansen, M. C., Ballon, J. S., … Gross, J. J. (2014). Emotional granularity and social functioning in individuals with schizophrenia: An experience sampling study. *Journal of Psychiatric Research, 53*, 141–148.

Kircanski, K., Lieberman, M. D., & Craske, M. G. (2012). Feelings into words: Contributions of language to exposure therapy. *Psychological Science, 23*, 1086–1091.

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General, 140*, 14–34.

Kross, E., Bruehlman-Senecal, E., Park, J., Burson, A., Dougherty, A., Shablack, H., … Ayduk, O. (2014). Self-talk as a regulatory mechanism: How you do it matters. *Journal of Personality and Social Psychology, 106*, 304–324.

Kuhn, L. J., Willoughby, M. T., Vernon-Feagans, L., & Blair, C. B. (2016). The contribution of children's time-specific and longitudinal expressive language skills on developmental trajectories of executive function. *Journal of Experimental Child Psychology, 148*, 20–34.

Larson, R. W., & Ham, M. (1993). Stress and "storm and stress" in early adolescence: The relationship of negative events with dysphoric affect. *Developmental Psychology, 29*, 130–140.

Larson, R. W., Moneta, G., Richards, M. H., & Wilson, S. (2002). Continuity, stability, and change in daily emotional experience across adolescence. *Child Development, 73*(4), 1151–1165.

Lee, F. S., Heimer, H., Giedd, J. N., Lein, E. S., Šestan, N., Weinberger, D. R., & Casey, B. J. (2014). Adolescent mental health—Opportunity and obligation. *Science, 346*(6209), 547–549.

Liberman, N., & Förster, J. (2009). The effect of psychological distance on perceptual level of construal. *Cognitive Science, 33*, 1330–1341.

Lieberman, M. D., Eisenberger, N. I., Crockett, M. J., Tom, S. M., Pfeifer, J. H., & Way, B. M. (2007). Putting feelings into words: Affect labeling disrupts amygdala activity in response to affective stimuli. *Psychological Science, 18*(5), 421–428.

Lieberman, M. D., Inagaki, T. K., Tabibnia, G., & Crockett, M. J. (2011). Subjective responses to emotional stimuli during labeling, reappraisal, and distraction. *Emotion, 11*, 468–480.

Lindquist, K. A., & Barrett, L. F. (2008). Constructing emotion: The experience of fear as a conceptual act. *Psychological Science, 19*, 898–903.

Lindquist, K. A., Barrett, L. F., Bliss-Moreau, E., & Russell, J. A. (2006). Language and the perception of emotion. *Emotion, 6*, 125–138.

Lindquist, K. A., & Gendron, M. (2013). What's in a word? Language constructs emotion perception. *Emotion Review, 5*, 66–71.

Lindquist, K. A., Gendron, M., Barrett, L. F., & Dickerson, B. C. (2014). Emotion perception, but not affect perception, is impaired with semantic memory loss. *Emotion, 14*, 375–387.

Lindquist, K. A., MacCormack, J. K., & Shablack, H. (2015). The role of language in emotion: Predictions from psychological constructionism. *Frontiers in Psychology, 6*, 444.

Lindquist, K. A., Satpute, A. B., & Gendron, M. (2015). Does language do more than communicate emotion? *Current Directions in Psychological Science, 24*, 99–108.

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.

Linehan, M. M., Schmidt, H., Dimeff, L. A., Craft, J. C., Kanter, J., & Comtois, K. A. (1999). Dialectical behavior therapy for patients with borderline personality disorder and drug-dependence. *American Journal on Addictions, 8*(4), 279–292.

Lupyan, G. (2012). What do words do? Toward a theory of language-augmented thought. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 57, pp. 255–297). Waltham, MA: Academic Press.

Lupyan, G. (2016). The centrality of language in human cognition. *Language Learning, 66*, 516–553.

McLaughlin, K. A., Garrad, M. C., & Somerville, L. H. (2015). What develops during emotional development? A component process approach to identifying sources of psychopathology risk in adolescence. *Dialogues in Clinical Neuroscience, 17*(4), 403.

McRae, K., Gross, J. J., Weber, J., Robertson, E. R., Sokol-Hessner, P., Ray, R. D., … Ochsner, K. N. (2012). The development of emotion regulation: An fMRI study of cognitive reappraisal in children, adolescents and young adults. *Social Cognitive and Affective Neuroscience, 7*(1), 11–22.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106*, 226–254.

Neta, M., & Tong, T. T. (2016). Don't like what you see? Give it time: Longer reaction times associated with increased positive affect. *Emotion, 16*(5), 730–739.

Neta, M., & Whalen, P. J. (2010). The primacy of negative interpretations when resolving the valence of ambiguous facial expressions. *Psychological Science, 21*(7), 901–907.

Nook, E. C., Lindquist, K. A., & Zaki, J. (2015). A new look at emotion perception: Concepts speed and shape facial emotion recognition. *Emotion, 15*, 569–578.

Nook, E. C., Sasse, S. F., Lambert, H. K., McLaughlin, K. A., & Somerville, L. H. (2017). Increasing verbal knowledge mediates development of multidimensional emotion representations. *Nature Human Behaviour, 1*, 881–889.

Nook, E. C., Sasse, S. F., Lambert, H. K., McLaughlin, K. A., & Somerville, L. H. (2018). The nonlinear development of emotion differentiation: Granular emotional experience is low in adolescence. *Psychological Science, 29*, 1346–1357.

Nook, E. C., Schleider, J. L., & Somerville, L. H. (2017). A linguistic signature of psychological distancing in emotion regulation. *Journal of Experimental Psychology: General, 146*, 337–346.

Nook, E. C., Stavish, C. M., Sasse, S. F., Lambert, H. K., Mair, P., McLaughlin, K. A., & Somerville, L. H. (in press). Charting the development of emotion comprehension and abstraction using observer-rated and linguistic measures. *Emotion*.

Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences, 9*(5), 242–249.

Oosterwijk, S., Lindquist, K. A., Adebayo, M., & Barrett, L. F. (2015). The neural representation of typical and atypical experiences of negative images: Comparing fear, disgust and morbid fascination. *Social Cognitive and Affective Neuroscience, 11*(1), 11–22.

Ortner, C. N. M. (2015). Divergent effects of reappraisal and labeling internal affective feelings on subjective emotional experience. *Motivation and Emotion, 39*, 563–570.

Patel, V., Flisher, A. J., Hetrick, S., & McGorry, P. (2007). Mental health of young people: A global public-health challenge. *The Lancet, 369*(9569), 1302–1313.

Peper, J. S., & Dahl, R. E. (2013). The teenage brain: Surging hormones-brain-behavior interactions during puberty. *Current Directions in Psychological Science, 22*(2), 134–139.

Petersen, A. C. (1988). Adolescent development. *Annual Review of Psychology, 39*, 583–607.

Petersen, A. C., Compas, B. E., Brooks-Gunn, J., Stemmler, M., Ey, S., & Grant, K. E. (1993). Depression in adolescence. *American Psychologist, 48*, 155–168.

Piaget, J. (1952). *The child's concept of number*. New York, NY: Norton.

Ponari, M., Norbury, C. F., & Vigliocco, G. (2018). Acquisition of abstract concepts is influenced by emotional valence. *Developmental Science, 21*, e12549.

Pond, R. S., Kashdan, T. B., DeWall, C. N., Savostyanova, A., Lambert, N. M., & Fincham, F. D. (2012). Emotion differentiation moderates aggressive tendencies in angry people: A daily diary analysis. *Emotion, 12*, 326–337.

Pons, F., Harris, P. L., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European Journal of Developmental Psychology, 1*, 127–152.

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology, 17*(3), 715–734.

Riediger, M., Schmiedek, F., Wagner, G. G., & Lindenberger, U. (2009). Seeking pleasure and seeking pain: Differences in prohedonic and contrahedonic motivation from adolescence to old age. *Psychological Science, 20*(12), 1529–1535.

Roben, C. K. P., Cole, P. M., & Armstrong, L. M. (2013). Longitudinal relations among language skills, anger expression, and regulatory strategies in early childhood. *Child Development, 84*, 891–905.

Roberson, D., Davidoff, J., & Braisby, N. (1999). Similarity and categorisation: Neuropsychological evidence for a dissociation in explicit categorisation tasks. *Cognition, 71*(1), 1–42.

Rodman, A. M., Powers, K. E., & Somerville, L. H. (2017). Development of self-protective biases in response to social evaluative feedback. *Proceedings of the National Academy of Sciences, 114*, 13158–13163.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*, 1161–1178.

Russell, J. A., Bachorowski, J. A., & Fernandez-Dols, J. M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology, 54*, 329–349.

Russell, J. A., & Bullock, M. (1985). Multidimensional scaling of emotional facial expressions: Similarity from preschoolers to adults. *Journal of Personality and Social Psychology, 48*, 1290–1298.

Russell, J. A., & Ridgeway, D. (1983). Dimensions underlying children's emotion concepts. *Developmental Psychology, 19*, 795–804.

Salmon, K., O'Kearney, R., Reese, E., & Fortune, C. A. (2016). The role of language skill in child psychopathology: Implications for intervention in the early years. *Clinical Child and Family Psychology Review, 19*, 352–367.

Satpute, A. B., Nook, E. C., Narayanan, S., Shu, J., Weber, J., & Ochsner, K. N. (2016). Emotions in "black and white" or shades of gray? How we think about emotion shapes our perception and neural representation of emotion. *Psychological Science, 27*, 1428–1442.

Silk, J. S., Stroud, L. R., Siegle, G. J., Dahl, R. E., Lee, K. H., & Nelson, E. E. (2012). Peer acceptance and rejection through the eyes of youth: Pupillary, eyetracking and ecological data from the Chatroom interact task. *Social Cognitive and Affective Neuroscience, 7*, 93–105.

Silvers, J. A., Insel, C., Powers, A., Franz, P., Helion, C., Martin, R. E., … Ochsner, K. N. (2017). vlPFC–vmPFC–amygdala interactions underlie age-related differences in cognitive regulation of emotion. *Cerebral Cortex, 27*, 3502–3514.

Silvers, J. A., Insel, C., Powers, A., Franz, P., Weber, J., Mischel, W., … Ochsner, K. N. (2014). Curbing craving: Behavioral and brain evidence that children regulate craving when instructed to do so but have higher baseline craving than adults. *Psychological Science, 25*, 1932–1942.

Silvers, J. A., McRae, K., Gabrieli, J. D. E., Gross, J. J., Remy, K. A., & Ochsner, K. N. (2012). Age-related differences in emotional reactivity, regulation, and rejection sensitivity in adolescence. *Emotion, 12*(6), 1235–1247.

Silvers, J. A., Shu, J., Hubbard, A. D., Weber, J., & Ochsner, K. N. (2015). Concurrent and lasting effects of emotion regulation on amygdala response in adolescence and young adulthood. *Developmental Science, 18*(5), 771–784.

Simonsohn, U. (2017). Two-lines: A valid alternative to the invalid testing of U-shaped relationships with quadratic regressions (pp. 1–36). *SSRN*. Retrieved from https://ssrn.com/abstract=3021690

Sisk, C. L., & Zehr, J. L. (2005). Pubertal hormones organize the adolescent brain and behavior. *Frontiers in Neuroendocrinology, 26*(3–4), 163–174.

Smith, L., & Klein, R. (1990). Evidence for semantic satiation: Repeating a category slows subsequent semantic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 852–861.

Soderberg, C. K., Callahan, S. P., Kochersberger, A. O., Amit, E., & Ledgerwood, A. (2015). The effects of psychological distance on abstraction: Two meta-analyses. *Psychological Bulletin, 141*, 525–548.

Somerville, L. H., Jones, R. M., & Casey, B. J. (2010). A time of change: Behavioral and neural correlates of adolescent sensitivity to appetitive and aversive environmental cues. *Brain and Cognition, 72*, 124–133.

Somerville, L. H., Jones, R. M., Ruberry, E. J., Dyke, J. P., Glover, G., & Casey, B. J. (2013). The medial prefrontal cortex and the emergence of self-conscious emotion in adolescence. *Psychological Science, 24*(8), 1554–1562.

Somerville, L. H., & McLaughlin, K. A. (2018). What develops during emotional development? Normative trajectories and sources of psychopathology risk in adolescence. In A. S. Fox, R. C. Lapate, A. J. Shackman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (2nd ed.). New York, NY: Oxford University Press.

Steinberg, L., & Morris, A. S. (2001). Adolescent development. *Annual Review of Psychology, 52*, 83–110.

Stroud, L. R., Foster, E., Papandonatos, G. D., Handwerger, K., Granger, D. A., Kivlighan, K. T., & Niaura, R. (2009). Stress response and the adolescent transition: Performance versus peer rejection stressors. *Development and Psychopathology, 21*, 47–68.

Suvak, M. K., Litz, B. T., Sloan, D. M., Zanarini, M. C., Barrett, L. F., & Hofmann, S. G. (2011). Emotional granularity and borderline personality disorder. *Journal of Abnormal Psychology, 120*, 414–426.

Tabibnia, G., Lieberman, M. D., & Craske, M. G. (2008). The lasting effect of words on feelings: Words may facilitate exposure effects to threatening images. *Emotion, 8*, 307–317.

Torre, J. B., & Lieberman, M. D. (2018). Putting feelings into words: Affect labeling as implicit emotion regulation. *Emotion Review, 10*(2), 116–124.

Trentacosta, C. J., & Izard, C. E. (2007). Kindergarten children's emotion competence as a predictor of their academic competence in first grade. *Emotion, 7*, 77–88.

Trope, Y., & Liberman, N. (2000). Temporal construal and time-dependent changes in preference. *Journal of Personality and Social Psychology, 79*, 876–889.

Troy, A. S., Ford, B. Q., McRae, K., Zarolia, P., & Mauss, I. B. (2017). Change the things you can: Emotion regulation is more beneficial for people from lower than from higher socioeconomic status. *Emotion, 17*(1), 141–154.

Troy, A. S., Shallcross, A. J., & Mauss, I. B. (2013). A person-by-situation approach to emotion regulation: Cognitive reappraisal can either help or hurt, depending on the context. *Psychological Science, 24*(12), 2505–2514.

Tugade, M. M., Fredrickson, B. L., & Barrett, L. F. (2004). Psychological resilience and positive emotional granularity: Examining the benefits of positive emotions on coping and health. *Journal of Personality, 72*(6), 1161–1190.

Van Cauwenberge, V., Van Leeuwen, K., Hoppenbrouwers, K., & Wiersema, J. R. (2017). Developmental changes in neural correlates of cognitive reappraisal: An ERP study using the late positive potential. *Neuropsychologia, 95*, 94–100.

Wechsler, D. (1991). *The Wechsler intelligence scale for children* (3rd ed.). San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI)*. San Antonio, TX: The Psychological Corporation.

Westenberg, P. M., Drewes, M. J., Goedhart, A. W., Siebelink, B. M., & Treffers, P. D. A. (2004). A developmental analysis of self-reported fears in late childhood through mid-adolescence: Social-evaluative fears on the rise? *Journal of Child Psychology and Psychiatry, 45*(3), 481–495.

Whalen, P. J., & Phelps, E. A. (2009). *The human amygdala*. New York, NY: Guilford Press.

Widen, S. C. (2013). Children's interpretation of facial expressions: The long path from valence-based to specific discrete categories. *Emotion Review, 5*, 72–77.

Wintre, M. G., & Vallance, D. D. (1994). A developmental sequence in the comprehension of emotions: Intensity, multiple emotions, and valence. *Developmental Psychology, 30*, 509–514.

Wu, H., Mata, J., Furman, D. J., Whitmer, A. J., Gotlib, I. H., & Thompson, R. J. (2017). Anticipatory and consummatory pleasure and displeasure in major depressive disorder: An experience sampling study. *Journal of Abnormal Psychology, 126*(2), 149.

Yin, J., & Csibra, G. (2015). Concept-based word learning in human infants. *Psychological Science, 26*, 1316–1324.

Zaki, J., Davis, J. I., & Ochsner, K. N. (2012). Overlapping activity in anterior insula during interoception and emotional experience. *NeuroImage, 62*, 493–499.

Zaki, L. F., Coifman, K. G., Rafaeli, E., Berenson, K. R., & Downey, G. (2013). Emotion differentiation as a protective factor against nonsuicidal self-injury in borderline personality disorder. *Behavior Therapy, 44*, 529–450.

Chapter 3
From the Self to the Social Regulation of Emotion: An Evolving Psychological and Neural Model

Kevin N. Ochsner

Imagine that you have just moved across the country to take a job as a professor at a new and exciting university. Beyond all the usual pragmatic hassles, like organizing the move, finding a place to live, and so on, perhaps the biggest challenges you will face are social and emotional. How you adaptively respond to these challenges will go a long way toward determining the ease of your transition, success in this new job, and your overall well-being. For example, you must meet and get to know all your new colleagues and their relationships to one another, including their relative differences in disposition, status, and friendship. At your new place of residence, you will meet new neighbors and come to understand their connections to one another. At your children's school, you will meet many new parents and children and will come to know the complex web of relationships that ties them all together. And while doing all of this, you must—of course—be working to keep your research program going, mentoring your students, preparing to teach new classes, and establishing your new lab.

Successfully navigating all of these social and emotional challenges requires a combination of three essential abilities. The first is the ability to appraise the personal meaning of all your new encounters and relationships and consequently experience and express the full range of appropriate emotional reactions to them. Emotions can be thought of as readouts of the relevance of people, situations, and stimuli to your goals, wants, and needs. As such, they will provide an essential guide to every aspect of your new life. The second is the ability to perceive and understand other people's behaviors, thoughts, intentions, and emotions, which is commonly referred to with the umbrella term "person perception." This ability will be invaluable to learning about every new individual that you meet—from sizing up their current emotions and thoughts to inferring their enduring dispositions and tendencies to establishing relationships with them. The third is the ability to exert top-down,

K. N. Ochsner (✉)
Department of Psychology, Columbia University, New York, NY, USA
e-mail: ko2132@columbia.edu

M. Neta, I. J. Haas (eds.), *Emotion in the Mind and Body*, Nebraska Symposium on Motivation 66, https://doi.org/10.1007/978-3-030-27473-3_3

cognitive control over both of the above, regulating your emotional responses as need be, as well as regulating the impressions you form of other people so as to ensure that they are accurate. Importantly, you can exert control not just to shape your own emotions and impressions, but those of other people as well, helping your new colleagues and friends to cope with their own social and emotional challenges.

Just as a television can produce a seemingly infinite variety of colors and images from pixels colored red, blue, and green—the variety and complexity of human social and emotional life may arise from interactions between these three essential abilities. Indeed, as illustrated in Fig. 3.1, many topics central to the study of emotion and social behavior lie at the intersection points between these three, "primary colors," including empathy, social cognition, and the self- and social regulation of emotion.

How should we organize our understanding of the psychological processes and brain mechanisms underlying these complex and intersecting abilities? A full answer to this question is beyond the scope of any single chapter—and in fact is the goal of entire disciplines like social and affective neuroscience.

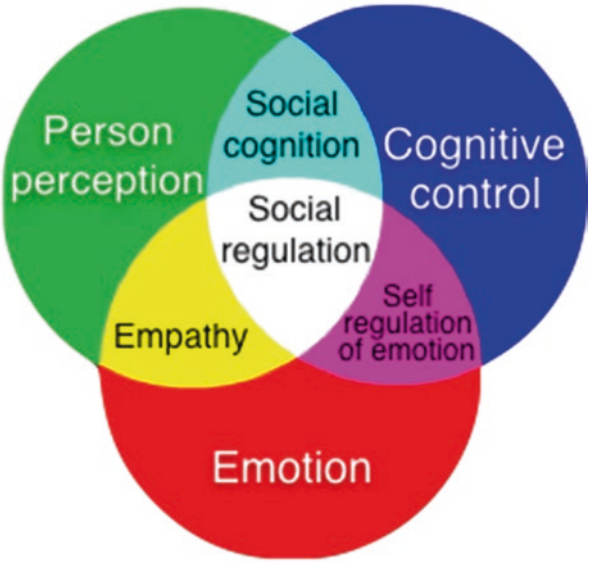

Fig. 3.1 Levels of analysis when studying social and emotional phenomena. At the behavioral level, we conceptualize person perception, cognitive control, and emotion as three "primary colors" of social and emotional life. Just as colored pixels on a screen combine in variegated ways to make a wide array of images, three core abilities can combine in varying ways to give rise to a wide array of social and emotional behaviors. The intersections of each of these domains define individual areas of research, including self-regulation and social regulation, which are the focus of this chapter. At the process level, these behavioral domains map onto varying combinations of underlying psychological processes. For illustrative purposes, these processes are grouped by the three core behavioral domains. At the neural level, each of these processes is supported by the concerted actions of cortical and subcortical brain regions

That said, the more modest goal of this chapter is to describe the development and evolution of a multilevel model of emotion and our capacity to regulate those emotions that is flexible and generalizable to a variety of contexts—ranging from the study of self-regulation to the study of the social regulation of emotion and beyond. Toward this end, the remainder of this chapter is divided into three parts. The first provides an overview of a model of the self-regulation of emotion that has been elaborated in more detail elsewhere (Braunstein, Gross, & Ochsner, 2017; Dore, Silvers, & Ochsner, 2016; Ochsner, Silvers, & Buhle, 2012). This model provides the foundation for the second section, which expands the model to the study of social forms of emotion regulation where one individual attempts to shape and change the emotions of another (Reeck, Ames, & Ochsner, 2016). The third and final section asks what lies ahead for the model and for the study of emotion regulation more generally, considering issues ranging from the continued evolution of the model and its usefulness for other areas of research (Ochsner, 2013, 2014).

The Starting Point: A Multilevel Model of the Self-Regulation of Emotion

For the past 15+ years, behavioral research on the self-regulation of emotion has been guided by James Gross's process model (Gross, 1998, 2015). According to this model, different types of emotion regulation strategies can be understood in terms of the stage of the emotion generation sequence that they impact (see white boxes, Fig. 3.2). Emotion generation proceeds when an emotion eliciting stimulus is perceived in the context of a particular situation, one attends to that stimulus or some aspects of it, they are appraised in terms of their meaning with respect to one's goals, wants, and needs, and depending on the nature of that appraisal, the various components of an emotional response are produced. Situation-focused regulatory strategies impact one's exposure and proximity to stimuli, such as when one moves away from an annoying stimulus or toward one that is desirable. Attention-focused strategies change the way one deploys selective attention to take in information that promotes desired emotional responses and ignore information that promotes undesired responses, such as when you divert your gaze during the scary part of a movie. Cognitive change-focused strategies alter the way one appraises the meaning of a stimulus, such as when you reappraise the rejection letter from a journal as an opportunity to improve the manuscript. Finally, response-focused strategies change the way one overtly expresses a motion on the face, body, and so on, such as when one abides by the British maxim to "keep a stiff upper lip" and limit the display of one's emotions.

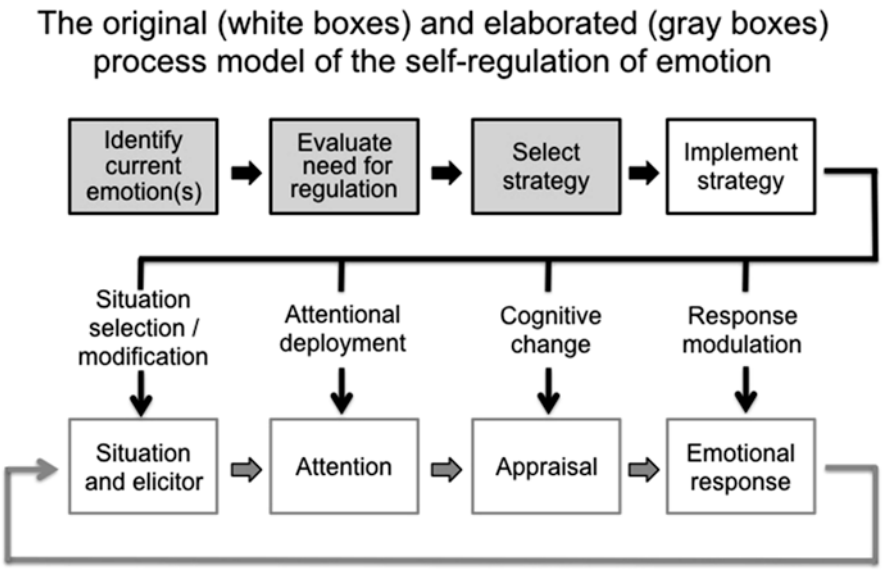

Fig. 3.2 White boxes: Elements of the original process model of emotion regulation (Gross, 1998), which described a system for classifying emotion regulation strategies in terms of the stage of an emotion generation sequence that they impact. This model spoke only to the strategies one could implement in a given situation. Gray boxes: Elements of an elaborated process model (Dore et al., 2016; Reeck et al., 2016; cf. Gross, 2015) specifying steps that logically precede the moment when one implements a strategy. One may identify the current emotional state, decide whether to regulate, and select a strategy

Proposing a Multilevel Model

Although the process model has been a powerful tool for organizing our understanding of the relationships between different types of regulatory strategies, it is silent about the neural mechanisms underlying them. To gain leverage on the nature of these mechanisms, the past decade has seen an enormous growth of functional magnetic resonance imaging (fMRI) research seeking to use patterns of brain activity to draw inferences about the psychological and neural mechanisms underlying specific strategies (Ochsner et al., 2012).

We were one of the first groups to take this approach (Ochsner, Bunge, Gross, & Gabrieli, 2002). When we began this research, late in the year 2000, virtually nothing was known about the neural systems supporting any of the emotion regulation strategies posited by the process model. We decided to start by focusing on a paradigm example of cognitive change—reappraisal—as well as attention-based strategies like distraction or selective attention. Drawing on prior work on "cold" forms of cognitive control, we proposed that strategies like reappraisal and attentional control might rely upon domain-general cognitive control systems localized in lateral prefrontal and inferior parietal cortices as well as posterior medial prefrontal

cortex (mPFC) and dorsal anterior cingulate cortex (dACC) (Botvinick, Braver, Barch, Carter, & Cohen, 2001; D'Esposito, Postle, Ballard, & Lease, 1999; Miller & Cohen, 2001). Effective regulation might depend on these systems effectively modulating activity in systems that generate emotional appraisals and the various components of an ensuing response. Our initial studies supported this prediction. And ever since, the lion's share of fMRI research on emotion regulation has continued to focus on reappraisal and attentional strategies. Four different meta-analyses showed that, to date, over 60 fMRI studies (Buhle et al., 2014; Kohn et al., 2014; Morawetz, Bode, Derntl, & Heekeren, 2017; O'Driscoll, Laing, & Mason, 2014) have supported our initial proposal that prefrontal regions implement processes like working memory to keep in mind regulatory goals and strategies, as well as selection processes necessary to either pick the right way to implement a given strategy and/or limit the pull of one's initial affective response. As lateral prefrontal regions implement these control processes, posterior medial frontal regions, including the dACC, are thought to monitor the extent to which reappraisal is desirable and successful, signaling the extent to which ongoing regulation is necessary. Together, these lateral and medial control systems are thought the change the way one attends to and interprets the meaning of affective stimuli whose value is computed by largely subcortical regions, such as the amygdala—which signals the presence of goal-relevant stimuli and can trigger initial affective responses to them—and the striatum, whose ventral portions are involved in computing expectancies about the reward value of stimuli (Helion, Krueger, & Ochsner, 2019; Ochsner et al., 2012). Figure 3.3 schematically illustrates these regions.

Notably, the model posits that prefrontal control and largely subcortical affect systems can interact in multiple ways, depending on the strategy in question and one's goals when using it (Helion et al., 2019; Ochsner et al., 2012). Reappraisal,

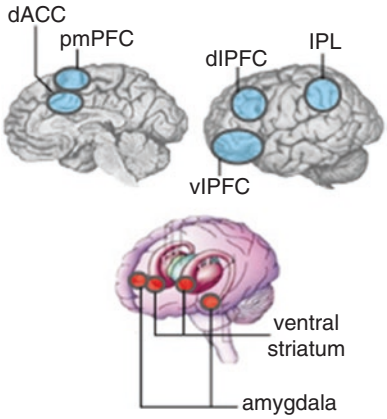

Fig. 3.3 Schematic representation of brain regions supporting reappraisal as suggested by meta-analyses. Control-related regions shown in blue, affect generation-related regions shown in pink. *dACC* dorsal anterior cingulate cortex, *pmPFC* posterior medial prefrontal cortex, *dlPFC* dorsal lateral prefrontal cortex, *vlPFC* ventral lateral prefrontal cortex, *IPC* inferior parietal cortex

for example, can be used to downregulate negative emotion by reinterpreting upsetting events in ways that lessen their emotional punch. But it also can be used to expand and embellish negative appraisals that make you feel much worse than you had initially (Ochsner, Ray, et al., 2004). This stands in contrast to other models of self-regulation that posit reciprocal and/or inhibitory relationships between cognitive control and emotion systems (Drevets & Raichle, 1998; Heatherton & Wagner, 2011; Lieberman et al., 2007; Metcalfe & Mischel, 1999). Such theories view cognition and emotion as generally antagonistic. As one comes to the fore, the other recedes. While this is surely the case some of the time—as when one uses reappraisal to downregulate emotion—we and others have documented numerous instances where cognitive control systems can be used to amplify or even wholly generate emotional responses, as when one imagines a seemingly innocuous stimulus, like the creak of a floorboard in a quiet house, might be an indication that something sinister is afoot. And as discussed elsewhere (Ochsner, 2013, 2014), control can be used in support of various other affective abilities as well. The key point is that cognitive control systems allow us to flexibly interpret and reinterpret all kinds of external sensory inputs and internal sensations, and depending on how we attend to and appraise these stimuli, different types of emotions will be produced to differing degrees (Ochsner, 2013, 2014).

Elaborating the Initial Model

While the work summarized above has helped flesh out a multilevel model of the self-regulation of emotion that connects behavior, psychological processes, and underlying brain systems, in the past few years, it has become increasingly clear that it may be the tip of the proverbial regulatory iceberg (Dore et al., 2016; Gross, 2015; Ochsner & Gross, 2014; Reeck et al., 2016).

Implementation The core idea is that the process model as initially formulated speaks only to the way in which individuals *implement* regulatory strategies and is silent about how one got to the point at which one is trying to regulate. What's more, extant laboratory techniques are not designed to test anything other than implementation. Indeed, the vast majority of them present a narrow range of aversive stimuli for which regulation might obviously be desirable, instruct/train participants how to regulate, and tell them when to regulate. In everyday life, however, all of these factors are underdetermined and may play key roles in determining whether regulation is successful. With these kinds of considerations in mind, the gray boxes in Fig. 3.2 outline three steps that we think may precede the act of implementing a given regulatory strategy (Dore et al., 2016; Reeck et al., 2016) and are described below.

Selection Prior to implementation, we believe that one must *select* a strategy from some set of alternatives that could be considered. How many strategies you consider will depend on your knowledge of the kinds of strategies that could be used, in

general, and the extent to which situational cues, your regulatory goals, or other factors bring them to mind. For instance, if you are in a heated argument with a friend and want to regulate your response, should you move away (changing the nature of the emotion-eliciting situation), try to change the topic of conversation to something less conflictual (using distraction, a form of attentional control), try to reinterpret the nature of the conflict or your friend's actions in a way that makes you and them feel less upset (an instance of cognitive reappraisal), or should you simply try to mask your facial and bodily expressions of anger so that your friend can't tell how upset you are (an instance of response modulation)?

In recognition of the potential importance of selection to the regulatory process, to date, this stage has seen the most behavioral research of all the expanded stages discussed here. In upward of half a dozen studies, Sheppes and colleagues have probed the *selection* stage by asking under what circumstances people decide to reappraise as compared to distract themselves in the face of unpleasant stimuli. Across younger and older participants and across typically developing and current or formerly clinical populations (e.g., remitted bipolar or a current borderline diagnosis), they have found that distraction is more often chosen for the most intensely aversive experiences whereas reappraisal is chosen for less intense aversive experiences (Hay, Sheppes, Gross, & Gruber, 2015; Sauer et al., 2016; Scheibe, Sheppes, & Staudinger, 2015; Sheppes et al., 2014; Sheppes & Levin, 2013; Sheppes, Scheibe, Suri, & Gross, 2011; Suri, Sheppes, & Gross, 2013).

Evaluation Prior to strategy selection, it is necessary to *evaluate* whether or not regulation is needed at all. In some circumstances, it may be wholly appropriate to experience even intensely negative emotions, such as when one is appropriately angry with an insult, experiences grief at a funeral, is afraid of a high-risk investment opportunity, or is faced with a situation or emotion that is too ambiguous or too intense to be regulated. In such circumstances, it might be wise to wait until your emotions calm down or the situation becomes clearer before thinking again about whether regulation is called for. What's more, attempting to downregulate your emotions may sometimes prove counterproductive. Recent research suggests, for example, that individuals high in self-control may take unwarranted risks in situations where they should heed their fears of failure or loss (Konnikova, 2013). Similarly, reappraisal may be most beneficial in situations that cannot be controlled, where other types of strategies might not be possible and rethinking the meaning of what is happening may be the best option. Iris Mauss and colleagues have found this to be true in lab situations that are less controllable as well as in everyday life situations faced by low SES individuals who may have less control over life stressors than do high SES individuals (Mauss et al., 2011; Troy, Ford, McRae, Zarolia, & Mauss, 2017; Troy, Shallcross, & Mauss, 2013).

Identification Finally, in order to decide whether or not your current emotional state needs to be regulated, there has to be some internal representation of that current state. Note that this representation may in many circumstances be conscious—as when you introspectively assess how you are feeling and realize you are anxious

and afraid prior to giving a very important talk—but in others the representation of your current state can be non-conscious, as when regulatory systems take as inputs the outputs of emotional response systems and implement regulatory actions outside your awareness. In such cases, regulation is guided by non-conscious goals and processes that may engage the lateral and dorsal medial prefrontal systems mentioned above (Lau & Passingham, 2007), but they may also depend on ventral medial prefrontal regions important for learning the affective values of stimuli and how those values change within different spatiotemporal contexts (see Braunstein et al., 2017 for more discussion).

Neuroscience Research on the Expanded Model

While behavioral studies have increasingly begun recognizing the potential importance of these additional regulatory steps, neuroscience research has yet to significantly take up their investigation. Below we offer what currently is known about the brain systems supporting each stage, including detailing our lab's initial forays into studying the evaluation and identification stages.

Selection Above we discussed the work by Sheppes and colleagues showing that distraction versus reappraisal is preferentially chosen more for high versus low intensity aversive experiences (Sheppes et al., 2011; Sheppes et al., 2014; Sheppes & Levin, 2013; Suri et al., 2013). While there are many possible reasons for this, one explanation may be that highly arousing and aversive experiences can trigger a series of both fast and slow stress-related responses. Fast responses include changes in the neurotransmitter profiles of prefrontal cortex, and slow responses include cortisol release that modulates energy metabolism and amygdala encoding (Arnsten, 2015; Peters, McEwen, & Friston, 2017; Sapolsky, 2015). Extant animal work with rodent models dovetails with recent functional imaging and stress studies in humans to suggest that, together, these responses may diminish prefrontal capacity in the face of acute stressors while at the same time enhancing amygdala responsivity (Maier, 2015; van Ast et al., 2016). Behaviorally, these neural effects can lead to riskier choices (Uy & Galvan, 2017), reduced model-based learning (Otto, Raio, Chiang, Phelps, & Daw, 2013), and reduced ability to reappraise stimuli eliciting conditioned fear responses (Raio, Orederu, Palazzolo, Shurick, & Phelps, 2013; Raio & Phelps, 2015). Over time, exposure to chronic stressors can rework cortical-subcortical pathways to make these changes long-lasting, resulting, for example, in greater amygdala responses in individuals exposed to chronic long-term stressors (Muscatell et al., 2015). Similar effects can be seen in individuals who faced a single severe stressor only 1 month prior (Reynaud et al., 2015).

To the extent that reappraisal depends critically on the integrity of prefrontal systems and their ability to communicate with the amygdala, choosing to reappraise in the face of a highly aversive situation may not always be optimal, especially when given the choice to distract oneself instead. It is worth noting, however, that

individual differences in stress reactivity and cognitive control capacity will loom large for these and all other stages of the expanded regulation model. For example, although stress can diminish prefrontally dependent working memory performance (Oei, Everaerd, Elzinga, van Well, & Bermond, 2006), individuals with greater working memory capacity (as measured by operation span) may be better able to resist the effects of stress on cognitive performance (Otto et al., 2013). Future work should ask whether individuals with greater cognitive control capacity may be more likely to choose strategies like reappraisal that depend on the kinds of prefrontal resources disrupted by stress.

Evaluation In our laboratory, we recently studied the *evaluation* stage using fMRI. We wanted to know what brain- and behavior-based variables would predict one's choice to reappraise as compared to just allowing the more natural response when faced with unpleasant events. To study this, we devised a two-part procedure (see Fig. 3.4). First, we presented participants with a set of neutral and moderate to high arousal aversive images, asking them to rate how they felt in response to each one. Whole-brain fMRI data were collected during this exposure phase. Then, in a second, choice phase that took place outside the scanner, participants were once again presented with all of these images and asked if they wanted to simply look at the image (and respond naturally) or regulate their response to the image using reappraisal. Based on these choices, we could bin the imaging data from initial presentation to differentiate activity for images to which participants chose to respond naturally versus reappraise. This allowed us to first identify activity in specific regions that predicted whether or not a given individual would subsequently regulate the response. To address this question, we focused on regions of interest (ROIs) in prefrontal cortex and amygdala that are involved in reappraisal, as identified in our 2014 meta-analysis (Buhle et al., 2014). We found that, when one first encountered an aversive image, activity in all of these regions predicted greater likelihood of an individual reappraising that image—and this finding generalized to predicting choices for similar novel aversive images as well. The fact that individuals showing prefrontal and amygdala activation were more likely to subsequently choose to reappraise raises the possibility that the amygdala response to an aversive image triggers prefrontal engagement, which in turn predicts a choice to regulate down the road. We tested this mediational relationship and found it to be significant (schematically illustrated in Fig. 3.5). We also compared relative strength with which brain-based (prefrontal and amygdala activity) and behavior-based (self-reported affect) variables could predict subsequent regulation choices. Notably, prefrontal activity was the single strongest predictor, and models that took into account both brain and behavior variables achieved high levels of accuracy for predicting which individuals are most likely to regulate when faced with aversive events.

We then turned to the question of whether or not we can predict the aversive images for which regulation was most likely to be chosen. Here, we again focused analyses on ROIs from our 2014 meta-analysis (Buhle et al., 2014), but this time performed a pattern expression analysis. This analysis asks to what extent, during the presentation of a given image, the whole brain pattern of activity is similar to the

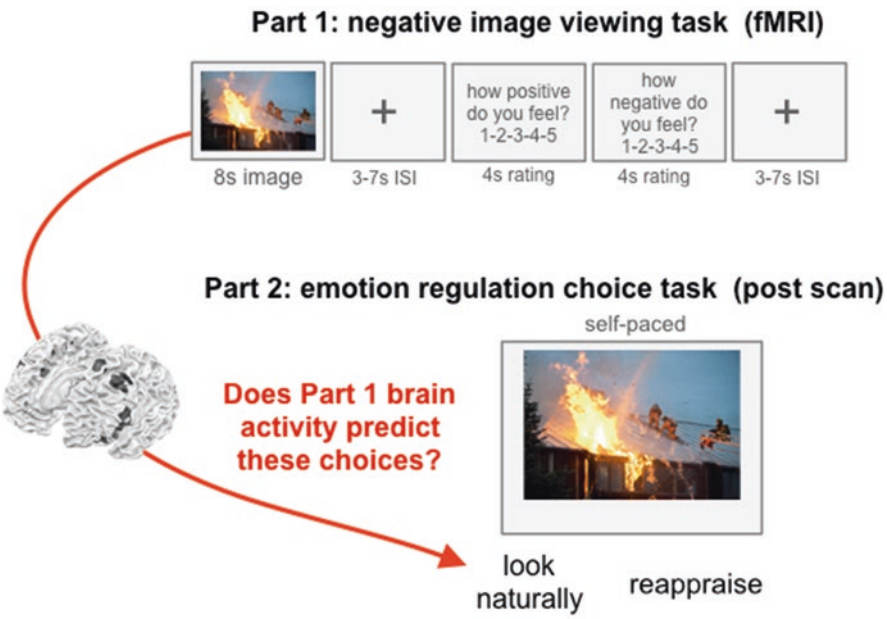

Fig. 3.4 Design of fMRI study exploring brain systems involved in evaluating the need to regulate aversive emotion (Dore, Morris, Burr, Picard, & Ochsner, 2017; Dore, Weber, & Ochsner, 2017)

whole brain reappraisal pattern from our meta-analysis. We found that the greater the expression of this reappraisal pattern in response to a given image, the more likely an individual was to later choose to reappraise their response to it. We then asked which variables—brain- or behavior-based—were the best predictors of choices to reappraise. We found that reappraisal pattern expression and average levels of activity in prefrontal and amygdala ROIs were the single best predictors and that models that took into account both brain and behavior variables again were the best predictors of choices to reappraise a given image. Together, these data suggest that when one first encounters an image, affect systems, like the amygdala, signal the presence of a goal-relevant stimuli (in this case, potential threats). That response recruits prefrontal activity to help interpret the meaning of the image. If one judges the stimulus to require regulation (in this case, to be sufficiently aversive), then one will be more likely to decide to reappraise.

Identification We have also begun investigating the identification stage to ask how it works and what are the consequences of introspectively identifying your emotions in different ways. We and others have previously shown that dorsal medial prefrontal cortex (dmPFC) is critically involved in attention to and awareness of one's internal emotional state (Lane, Fink, Chau, & Dolan, 1997; Ochsner, Knierim, et al., 2004; Phan et al., 2003; Taylor, Phan, Decker, & Liberzon, 2003), whereas regions of lateral PFC were important for selecting among competing alternative labels for those states (Lieberman et al., 2007; Satpute, Badre, & Ochsner, 2014; Satpute, Shu,

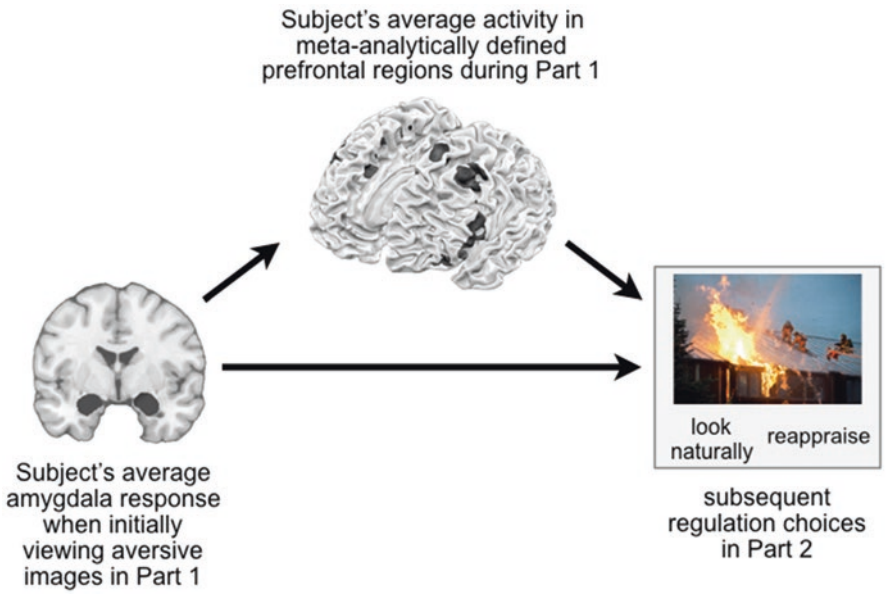

Fig. 3.5 A key result from Dore et al. (2017) and Dore, Weber, and Ochsner (2017)—using imaging data collected during initial Part 1 exposure to images (see Fig. 3.4), we could predict which individual participants would later (Part 2) choose to regulate emotion. Specifically, participants showing greater amygdala activation were more likely to engage prefrontal systems (presumably to help appraise the meaning of the images) and then later on choose to reappraise. Note that given that *stronger* rather than *weaker* amygdala responses predict subsequent regulation choices, it is unlikely that participants are already reappraising during initial Part 1 uninstructed exposure. Instead, PFC activity likely reflects cognitive processes related to evaluating image meaning (i.e., appraisal) whose engagement predicts the choice to regulate later on

Weber, Roy, & Ochsner, 2013). We have described the dorsomedial region as being important for a high or abstract level of representation for knowledge about mental states including emotional ones (Ochsner & Gross, 2014). The everyday language of emotion—I am happy, I am angry, I am sad, etc.—invokes a set of conceptual categories that can generalize across people and situations. We can use these terms to describe our own emotions, those of our friends, family, and so on, across a variety of circumstances. In this sense, being able to "recognize" our emotions has a lot in common with the recognition of objects in the world. To identify that an object sitting at the end of a conference table during a meeting is a cell phone rather than something else, we draw on high level, abstract knowledge of the form of objects that is viewpoint- and exemplar-irrelevant (Kosslyn, 1994). That is, it does not matter at what angle you view the phone and which specific phone it happens to be (an iPhone, Samsung smartphone, flip phone, etc.); in all cases, you know it is a phone. Identifying your emotions may work the same way: across viewpoints (i.e., emotion-arousing situations) and exemplars (i.e., the person experiencing the emotion, whether it is you or someone else), you can use high level, abstract knowledge about

mental states to identify the emotion in question. We believe that the dmPFC and associated regions (e.g. those that comprise a so-called mentalizing network; see Amodio & Frith, 2006; Zaki & Ochsner, 2012) more generally play key roles in the representation and use of this knowledge in everyday contexts where such information is useful, ranging from instances where you introspect about your own emotions to complex social interactions (Satpute et al., 2013, 2016; Zaki & Ochsner, 2012).

The fact that this knowledge takes the form of linguistically describable emotion categories—that we can think and talk about—turns out to have important and unexpected consequences. Imagine you are talking to a colleague about negative reviews of the manuscript you recently submitted to a top journal. As you recount the elements of the review, your emotions may swing from an initially neutral starting point to the depths of despair and back again. If your friend asks you to be specific about how you felt about a reviewer's request to conduct five new analyses, how would you respond? Might you note that you felt about a five on a seven-point scale, where one is neutral and seven is extremely negative? Or would you simply pick what seems like the most appropriate descriptor—in this case, angry? Everyday communication is heavily trafficked by terms like angry, happy, or sad used to describe our emotions. It is only in the world of the laboratory where we ask people to rate anything and everything on a seven-point scale. As such, when your negative emotion is quite strong, it might be easy to tell your friend you are angry. Conversely, when you are feeling calm, it might be easy to say you feel neutral. However, what about moments when you feel something in between—perhaps moderately but not extremely negative? Do you say you feel neutral or angry? And does it matter which one you pick?

We recently used a novel behavioral method combined with fMRI to ask how people make judgments about such liminal emotional states (Satpute et al., 2016). The method draws on the category boundary effect in perception research. As a hypothetical example, imagine sorting a set of balls of varying sizes into two bins; how do you do it? Very large balls could go into a large bin, whereas very small balls could go into a small bin. But what about the ball sized somewhere in between? It turns out that if you place them in the large bin, you come to perceive them as being larger than you had initially, whereas if you place them in the small bin, you come to perceive them as being smaller than you had initially. These types of effects are observed in various perceptual domains, ranging from vision to speech and audition (Anderson, Silverstein, Ritz, & Jones, 1977; Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Harnad, 1987).

Our task investigated category boundary effects in the perception of emotion using a variant of this procedure. Participants viewed images varying from neutral to moderately negative to highly negative. In the *continuous* condition, they rated their affective response by clicking anywhere that was appropriate along a graded scale ranging from neutral to bad. In the categorical condition, they had to choose which term—neutral or bad—best described their emotional response. Using psychophysical techniques, for each participant, we calculated a curve relating the

probability they responded neutral or bad versus the normative degree of negativity in the image based on prior norming samples of participants (or put another way, we plotted stimulus attributes vs. the ability to perceive their presence). This allowed us to determine how normatively negative an image had to be for a given participant's experience to cross a subjective "tipping point" for judging their own response to be negative. The critical question was to what extent being forced to choose a single categorical descriptor shifted this threshold as compared to being able to click anywhere along a continuous scale.

To make this concrete, consider a case analogous to the earlier example of sorting balls into large versus small bins. Imagine you are presented with a moderately negative image and are asked to rate your emotional response along a continuous scale. You might click somewhere near the midpoint of the continuously graded neutral-to-bad scale, indicating a moderately bad reaction. Now imagine that in the categorical condition, you must rate your reaction to this image by selecting either of two words—neural *or* bad—that best describes that response. When presented with these two cases in our experiment, we observed that some participants would pick the word "bad" to describe their reaction to moderately aversive stimuli, suggesting that when using categorical language to describe their emotions, they had a liberal threshold for judging whether a stimulus made them feel bad. Conversely, other participants tended to pick the word "neutral" to describe their reaction to moderately aversive stimuli, suggesting that when using categorical language to describe their emotions, they had a conservative threshold for judging whether a stimulus made them feel bad. Put another way, depending on whether your threshold for judging the negativity of your emotions became more liberal versus more conservative, you would lump those reactions into either the "bad" or "neutral" response category.

These behavioral data suggested that participants were actually experiencing the liminal, moderately aversive, boundary-level stimuli differently when forced to describe their emotional responses using linguistic categorical terms of the sort we use in everyday communication. Brain data backed this up. Participants with more liberal or lower thresholds for reporting negative responses also showed greater amygdala and insula activity, whereas participants with more conservative or higher thresholds for reporting negative responses showed weaker amygdala and insula activity (see left and center panels of Fig. 3.6). Notably, the extent of the shift was predicted by greater connectivity between amygdala and insula, with dorsal medial prefrontal regions thought to support access to linguistic category descriptors of affective states (right panels, Fig. 3.6; note also that although the topography of these regions looks slightly different when statistically thresholded at conventional levels, they do not meaningfully differ when directly compared). Together, these data highlight that simply introspecting about and reporting on certain kinds of emotional states can actually change them, leading them to be neurally represented—and perhaps amplifying or diminishing their experience—in a way consistent with the terms you use to describe those states (cf. Kircanski, Lieberman, & Craske, 2012).

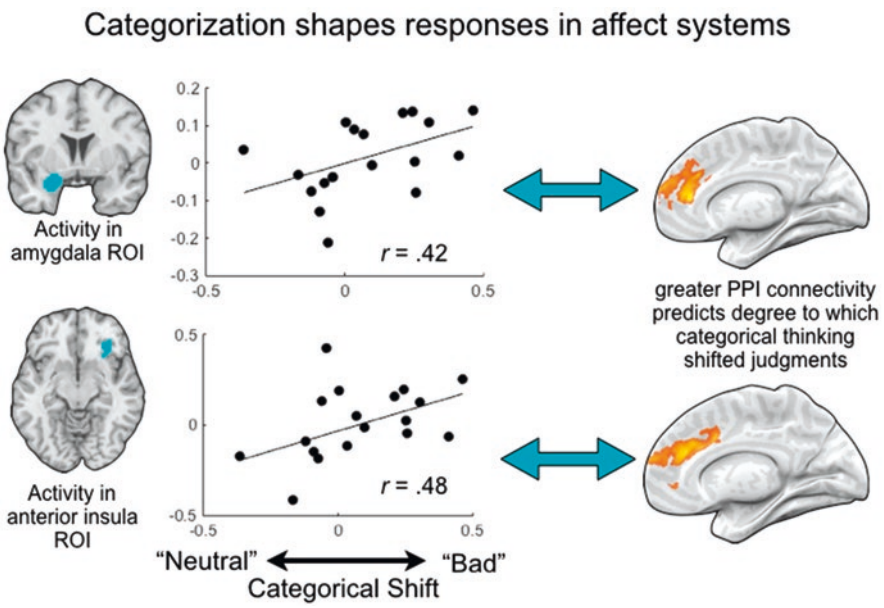

Fig. 3.6 Key results from study of identification stage of elaborated process model (Satpute et al., 2016). Individual participants vary in the extent to which categorical thinking shifts the threshold for reporting feeling neutral or bad when viewing moderately aversive images. The degree of this shift to report that more stimuli make you feel neutral versus bad correlates with lesser versus greater activity in key affective response systems (i.e., amygdala, insula) and was predicted by greater connectivity of these regions with dorsal medial prefrontal regions supporting access to linguistic category descriptors of affective states. See text for details

Evolving the Model of Self-Regulation to Account for the Social Regulation of Emotion

These elaborations to our initial model of the self-regulation of emotion broaden its scope and enable it to account for a much wider range of phenomena than the initial model could. Whereas the initial model was focused entirely on the implementation of strategies, the revised and expanded model describes three stages that come before implementation of a given strategy—identifying your emotion(s), evaluating the need to regulate your emotion(s), and if regulation is desired, selecting an appropriate strategy. As described above, new areas of research are growing up around these newly proposed stages. Over the next few years this work should help elucidate their psychological and neural bases—and perhaps more importantly, begin to elucidate how individual differences in our everyday emotions may arise from differences (between individuals or within an individual across time) in the way each of these stages operates. For example, an individual with anxiety might show low positive and elevated negative affect not just because they are unable to implement a particular kind of emotion regulation strategy, but rather (or also) because they have

a lower threshold for perceiving negative emotions, do not always identify situations where they should regulate, and/or have trouble selecting appropriate strategies even when they do deem that regulation is necessary. Likewise, other clinical populations—from substance users to individuals with mood or personality disorders—may also differ from the normative population in these ways. And children or older adults may similarly differ from young adult populations in the ways they identify their emotions, evaluate the need to regulate, and tend to select specific strategies.

But there is still another important aspect of our emotional lives and capacity for regulation on which the model described thus far is silent: the social context of emotion regulation and, in particular, the way in which one individual may actively regulate the emotions of another. The social regulation of emotion may be at least as common, if not more common, than the self-regulation of emotion. Indeed, countless times a day, parents must actively help their children respond emotionally to various challenges. Friends help each other respond to life's setbacks. Relationship partners provide regulatory support in times of need. Therapists assist their clients with current or long-running emotional struggles. And sometimes, social forms of regulation are undertaken with the intent not to help, but to disrupt regulation, as when competitors in sports or business attempt to emotionally disequilibrate their opponents in a game or negotiation.

The expanded model of the self-regulation of emotion in Fig. 3.2 can also account for social forms of regulation. If self-regulation involves using one's frontal lobe to regulate one's affective response systems, then social regulation might involve the use of your frontal lobe to regulate another person's affective response systems (*illustrated in* Fig. 3.7). In terms of the elaborated process model (Fig. 3.2), we can accommodate social regulation with two simple twists illustrated in Fig. 3.8. First, we can use the top row of boxes, starting with emotion identification and ending with strategy implementation, to describe the series of processing steps that take place in the mind and brain of an individual—designated the *regulator*—who is attempting to alter or shape the emotions of another individual. Second, we can use the bottom row of boxes, starting with the perception of an emotion-eliciting stimulus and ending with an emotional response, to describe the series of emotion-generating processing steps taking place in the mind and brain of that second individual who we designate the *target* of the first person's regulatory attempts (again, see Fig. 3.7).

To make this concrete, consider the example offered in the introduction of this chapter. Imagine that your move to a new university is experiencing some expected, but nonetheless significant, emotional turbulence as you to attempt to build your new lab and navigate the politics and bureaucracy of the new institution. As you experience and express your anxiety about one particular setback, your relationship partner perceives your emotional state and identifies it correctly, judges that this might be a moment where regulatory action could be helpful, and decides to try and improve the situation by taking you on a relaxing evening out.

In this way, our multilevel model of emotion regulation can account for both the self-regulation and the social regulation of emotion. Whereas the top row always describes processing steps engaged by a regulator, the bottom row describes the

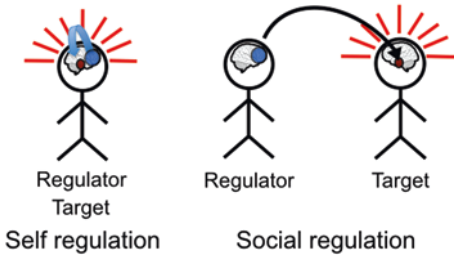

Fig. 3.7 Schematic of the relationship between the self-regulation and social-regulation of emotion: During self-regulation, you use your frontal lobe to regulate your affective response systems; during social regulation, you use your frontal lobe to regulate another person's affective response systems

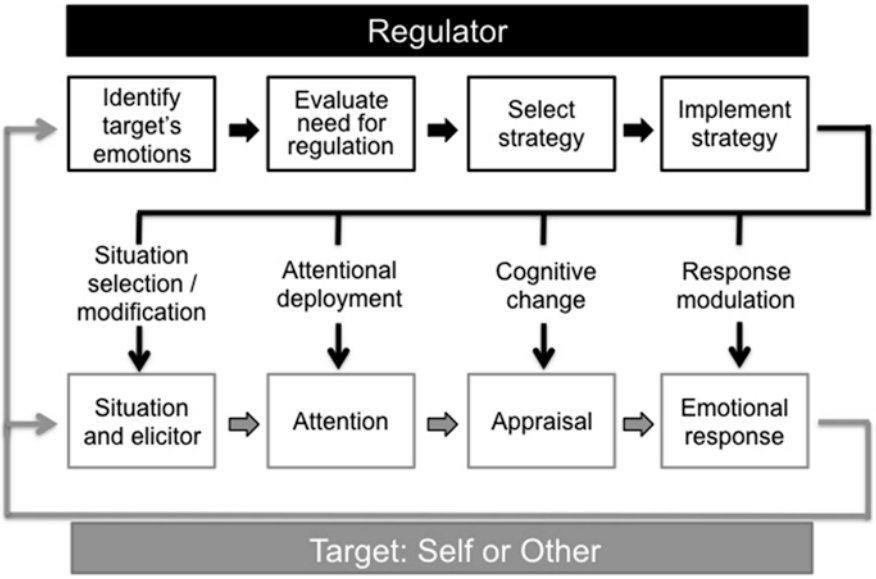

Fig. 3.8 A generalized model of emotion regulation that accounts for the self-regulation and the social regulation of emotion (cf. Dore et al., 2016; Reeck et al., 2016). See text for details and comparison to Fig. 3.2

processing steps generating the emotions of a regulatory target—whether that target is yourself (as in the case of self-regulation) or another person (as in the case of social regulation). In this way, what began as a model of the self-regulation of emotion—that we later adapted to the social regulation of emotion—can now be seen as a generalized emotion regulation model.

Conceptually, we have begun describing how this generalized model can organize our understanding of various kinds of behavioral and brain imaging data concerning the social regulation of emotion (Reeck et al., 2016). There have also

been other approaches to studying the social regulation of emotion that emphasize different factors, including the ecological and relational context of social regulation (Beckes & Coan, 2011) or a target's motivation for seeking out the assistance of a regulator (Zaki & Williams, 2013). We see the social application of our generalized models as complementary to these approaches.

Empirically, we have begun using our social application of the generalized model to begin exploring the psychological and neural underpinnings of various types of social emotion regulatory phenomena. We now turn to a few illustrative examples of this work.

Attempting to Identify Another's Emotions Can Result in the Self-Regulation of Emotion

The first example is a bridge from our earlier self-regulation research to our current interest in social regulation. It comes from a study asking how identifying another person's emotions—by simulating or empathizing with them—might have unexpected self-regulatory effects. The need to simulate and empathize with other people arises in virtually all of our relationships. And over time we learn to internally represent the people we know and the way they experience the world. Consider, for example, a friend recounting a close call with a New York City taxi cab where they were almost struck while crossing the street. If that friend is neurotic and reactive—like the popular conception of well-known New York resident Woody Allen—you might expect them to have felt a great deal of fear and anxiety. By contrast, if they are stoic and strong—like a character played by the Western movie actor John Wayne—then you might expect them to have kept their relative cool. We wanted to know whether the act of empathizing with the friend's response to an emotionally charged situation—an instance where you may need to identify their emotions prior to deciding whether it is appropriate to offer regulatory support—may have unexpected regulatory consequences all by itself.

To study this, we asked participants to take part in a study that ostensibly was about empathic accuracy (Gilead et al., 2016). In an initial behavioral session, they responded to a number of self-report questions about their personal preferences, tastes, and attributes. They were also asked to read what they were told were the responses of two prior participants. In reality, each set of responses was from a fictitious participant—one set having been pretested to come across as highly emotionally reactive, like the Woody Allen example above, whereas the other set was pretested to come across as strong and resilient, like the John Wayne example. In a second session that took place in an MRI scanner, participants were asked to complete an empathic accuracy test. On each of a series of trials, they would see a potentially emotionally charged photographic image and would be cued to subsequently rate either their own emotional response or what they believed would have been the emotional response of the (unbeknownst to the participants, fictitious) Woody- or Wayne-like individuals.

Behaviorally, we found that adopting the Woody versus Wayne perspectives resulted in larger versus smaller ratings of estimated negative affect for those targets—that is, relative to the amount reported for the participant's own reactions on trials where the images are experienced from one's own perspective. Neurally, a similar pattern was found for amygdala responses. Notably, the amygdala effects were observed in a set of voxels that were more active when negative images were viewed from one's own personal perspective, which provided initial evidence that simulating another person's perspective on an event may have the unintended effect of changing the way *you* are appraising the meaning of that event. Intriguingly—like the study on the self-identification of emotion described earlier—we found that functional connectivity between amygdala and dmPFC was positive when participants adopted the emotion-amplifying Woody perspective and connectivity was negative when they adopted the emotion-dampening Wayne perspective. Consistent with the idea that this dorsal medial prefrontal region is important for the high-level differentiation of mental states associated with each perspective, a multi-voxel pattern analysis of activity in this region showed significantly different patterns as a function of the perspective adopted on a given trial. Figure 3.9 illustrates these two results.

One problem with interpreting these results, however, is that amygdala responses can reflect a variety of processes, not just threat appraisals or aversive emotions. Current accounts of amygdala function suggest that it may have a more general neuromodulatory function, surveilling the environment for stimuli relevant to both your aversive *and* appetitive goals (Cunningham & Brosch, 2012; Todd, Cunningham, Anderson, & Thompson, 2012). As such, amygdala activation when

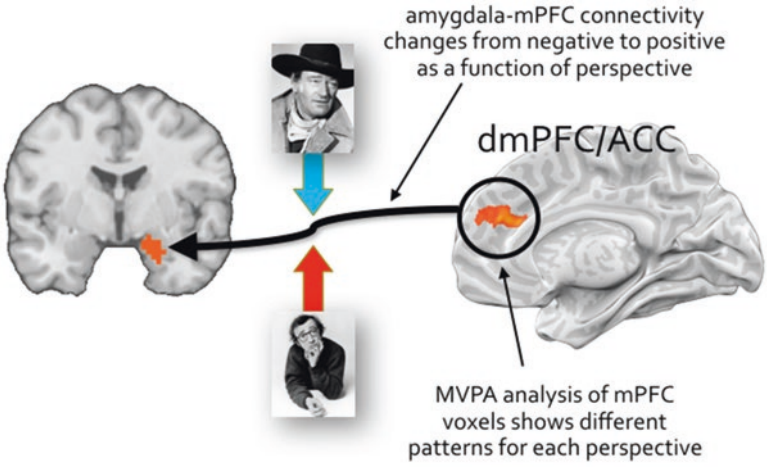

Fig. 3.9 In an initial study of how perspective taking serves as a form of social regulation (Gilead et al., 2016), we found that connectivity of amygdala with dmPFC was positive when participants adopted an emotion-amplifying perspective and connectivity was negative when adopting an emotion-dampening perspective. Multi-voxel pattern analysis of dmPFC activity showed significantly different patterns as a function of the perspective adopted

simulating another's perspectives could reflect changes in the way you are attending to and encoding goal-relevant information. While this may boil down to semantics in the sense that emotional appraisal maybe constituted of exactly these types of attentional and goal-directed processes, we nonetheless sought to provide a more stringent test of the idea that stimulating another's perspectives changes your own emotional experience. We then turned to an analytic technique developed by Luke Chang, Tor Wager, and colleagues at the University of Colorado Boulder (Chang, Gianaros, Manuck, Krishnan, & Wager, 2015). They identified a whole-brain pattern whose degree of expression predicts varying degrees of self-reported negative emotional experience. This pattern was derived from a large data set where participants view diverse images like the ones used in our study. We asked to what extent expression of this pattern varied as a function of the Woody versus Wayne perspective taken in our study. Critically, this pattern was expressed more strongly versus more weakly on trials where photos were viewed from the reactive Woody versus stoic Wayne perspectives. These data support the idea that simulating another person's emotional state—which we think may be the first step in a chain of events that could lead to the decision to help them regulate—may actually help regulate one's own emotional response.

Social Influence as an Example of the Social Regulatory Effects of Passively Identifying Another's Emotions

If actively simulating someone else's emotions may change the way you appraise and respond to an event, then an open question is whether and how passive exposure to another's emotions might also impact our emotions. Here, we started with the idea that any number of situations involve reacting to emotional events alongside other people. One common scenario is when we watch a movie in a crowded theater. If the moviegoers around us are laughing, we might be more likely to laugh as well. And if they are gasping in horror, our own fear might be heightened. Behavioral models of emotional contagion suggest that such effects should occur (Anderson, Monroy, & Keltner, 2017; Jordan, Rand, Arbesman, Fowler, & Christakis, 2013; Neumann & Strack, 2000), but little is known about the underlying neural mechanisms.

Observations such as these led us to ask how knowledge of other people's emotional responses to a shared event might have social regulatory effects. In this way, our model of the social regulation of emotion could help provide an account of the way in which social influences shape affective responding more generally. Neuroscience interest in social influence has increased over the past few years as evidenced by a handful of studies asking how knowledge of another's preferences for faces, music, and food is shaped by knowledge of group preferences for these stimuli. In general, these studies find that when you learn others have liked something either more or less than you do, subsequent tests demonstrate a corresponding shift in how much you like that stimulus—as well as in neural markers of subjective

liking, such as activity in ventral portions of the medial PFC and striatum (Izuma & Adolphs, 2013; Klucharev, Hytonen, Rijpkema, Smidts, & Fernandez, 2009; Klucharev, Munneke, Smidts, & Fernandez, 2011; Nook & Zaki, 2015; Zaki, Schirmer, & Mitchell, 2011).

To study this phenomenon in an emotion context, we developed a variant of the methods used to study social influence over subjective preferences. In an initial phase, participants viewed neutral, positive, and aversive photographic images and rated how good or bad they felt in response on a scale that ranged from very negative to very positive. A few seconds after making their own rating, they were shown what they were told was the average emotional response to that image recorded in a prior group of peer participants. In reality, however, this information was manipulated so as to equally often match or to be more toward the negative or more toward the positive end of the scale than was the participant's own response. In a subsequent second phase, participants viewed all of these images a second time and were asked to rate their current emotional responses to them. Instructions explained that we were simply interested in the way in which emotional responses may or may not change across time. Consistent with the prior research on subjective preferences (Izuma & Adolphs, 2013; Klucharev et al., 2009; Nook & Zaki, 2015), learning that your peers had responded to a given image more positively or more negatively than you did led you to have subsequent reactions that had shifted to be more similar to the peer response.

Insight into the neural mechanisms producing these effects came from an analysis of fMRI data collected at the moment participants learned that peers had responded with dissimilar versus similar emotions to their own. Activity in posterior medial frontal regions (e.g., dACC) was associated with response conflict, as well as dorsal and ventral lateral prefrontal regions associated with cognitive control. Notably, many (if not all) of these regions had been shown in a meta-analysis (Buhle et al., 2014) to be recruited when one self-regulates emotions via reappraisal. This suggested that simply knowing that others have responded to an event with emotions different than your own motivates reconsideration of your initial appraisal of the meaning of that event. Consistent with this idea, amygdala response to aversive images (but not other image types) became stronger after learning that peers had responded to these images more negatively than you had initially (Fig. 3.10). Curiously, we did not find that amygdala responses weakened when peers responded less negatively to an image than you had initially, and responses in other regions associated with appetitive responding (e.g., the ventral striatum) did not change as a function of influence-related changes in positive responses. While the selective nature of these findings is intriguing and in need of replication and extension, we are tempted to speculate that this pattern is consistent with a broader theme in behavioral research on emotion often summarized with the phrase "bad is stronger than good" (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001). The maxim is meant to convey that numerous studies demonstrate that negative emotions exert a stronger influence over behavior than do positive emotions. The present results fit this maxim insofar as a larger change in neural markers of appraisal (i.e., amygdala response) was found when participants learned that others believed an event was more emo-

tionally upsetting as compared to less upsetting than you did and that parallel effects were not observed for more positive reactions to aversive stimuli.

Implementing Social Reappraisals

A final example comes from the realization that social emotion regulation is not isolated to single dyadic relationships between relationship partners, parents and children, clients and therapists, and so on. In a world where we each inhabit multiple social roles, embedded in different social networks, with communication aided by social media, social regulation may be taking place in multiple relationships in parallel. Motivated by this realization, we planned to develop a means for studying the social regulation of emotion by capitalizing on digital platforms that facilitate multiple lines of communication between individuals. These plans became a reality in a collaboration with Rosalind Picard and her graduate student Rob Morris in the Affective Computing group of MIT's Media Lab.

For his dissertation research, Rob Morris devised an online platform known as Panoply (Morris, Schueller, & Picard, 2015). This platform allowed individuals to anonymously login and do two things: share short descriptions of stressors for which they might seek the support of others, and/or read other people's descriptions and provide supportive written responses. Each participant was free to choose which one they wanted to do whenever they logged into the site. Critically,

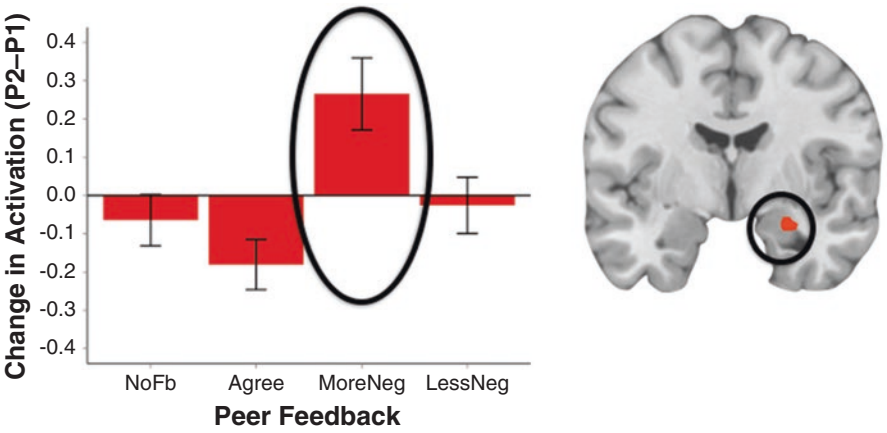

Fig. 3.10 Key result from study of social influence over affective responding (Martin, Weber, Koscik, Cunningham, & Ochsner, in press). Amygdala response to aversive images became stronger after learning that peers reported more negative responses to them. NoFb participants received no feedback about peer's emotion. *Agree* participants received feedback that peers had same emotional response to images. *MoreNeg, LessNeg* participants received feedback that peers had either more or less negative emotional response to images

however, in order to become a member of this platform, you had to first receive training in how to write brief summaries of stressful life events as well as how to provide effective written support to other people who posted descriptions of their own stressful life events. For writing responses to others, examples of three different kinds of regulatory strategies were provided: validating other people's feeling, debugging automatic negative thoughts (à la cognitive behavioral therapy), and reframing, which essentially was reappraisal in a social context. In an initial report on a group that participated in the Panoply environment for 3 weeks, Morris and colleagues reported an important finding: in general (i.e., without consideration of whether you wrote about a dilemma or provided supportive responses to others), participants felt happier and less depressed after having been in the environment (Morris et al., 2015).

These findings led us to ask *what aspect* of participating in the Panoply environment led to the mood benefits (Dore et al., 2017). In particular, we wondered whether or not the act of socially reappraising other people's unpleasant experiences might improve one's own reappraisal abilities. To address this question, we first asked what best predicted drops in depression: the number of stressful events a participant posted or the number of times a participant wrote supportive responses to other people's posts (note that because participants were free to either post, respond, or do both, analyses of the effects of each variable controlled for levels of the other variable)? Strikingly, providing support to others predicted one's own drop in depression.

Notably, the most common form of support offered in response to another's posts was reframing or reappraisal of the life events they described. This led us to ask a second question—did helping other people reframe their experiences change the way in which you reappraised your own life experiences? To test this hypothesis, we compared pre- vs. post-Panoply reports of the frequency with which participants reappraised in their daily lives collected using a common measure (the ERQ, see: Gross & John, 2003). This analysis showed that the frequency of reappraisal increased after having participated in the online environment, which motivated us to test a mediational model demonstrating that the extent to which writing supportive posts for others led you to feel less depressed depended on the extent to which it also increased the frequency of reappraisal in daily life. Or put another way, by helping others regulate their emotions, you may have more frequently reappraised your own emotional reactions, and that helped you feel less depressed. Again, these intriguing results may raise more questions than they answer, and future research is needed to explore the conditions under which personal benefits are derived from helping socially regulate other people's emotions. Numerous variables could ultimately prove important, ranging from the timing and frequency with which one helps others to the specific strategies used to the feedback a regulator receives from targets about the desirability and efficacy of their regulatory attempts (cf. Dore & Morris, 2018).

Looking Ahead: The Continued Evolution of Our Multilevel Model of Emotion Regulation

Studies of the brain systems supporting emotion regulation—when combined with careful behavioral analyses—can help us identify the psychological processes that connect behavior to brain. In so doing, they help us build a multilevel model of emotion regulation specifying the ways in which different classes of regulatory strategies rely on different sets of cognitive and affective processes that, in turn, arise from interactions among networks of brain systems.

This interdisciplinary, multilevel approach to studying emotion regulation has become so commonplace that it can be hard to remember it has been around for only a bit more than 10 years (Ochsner & Gross, 2005). As such, the day is still young and there is much research to come. In this final section of the chapter, we consider ways in which the model of regulation we propose will continue to evolve as well as its relationship to—and usefulness for—other related areas of research.

Evolution

While there are likely to be many ways in which the model will need to evolve—depending on the results of new studies conducted in the future—here we focus on two factors the model will need to increasingly consider.

From Groups to Individuals In many ways, the "holy grail" of all of psychological research is being able to provide an account of the behavior of a single individual. Indeed, we would all like to be able to specify how our theories make predictions for, and provide accounts of, the behavior of specific individuals. Unfortunately, of course, most so-called basic research is not designed to address this use. Instead, basic research is best suited for addressing questions and making predictions about the behavior of populations. As such, we make predictions for the group average, for processes that "in general" function in a certain way.

Ultimately, for models of emotion regulation to provide accounts that matter for our daily lives, the gulf between the population and the individual must be bridged. We suggested a means for building this bridge that involves conceptualizing every instance of emotion regulation as a person × situation × strategy interaction (Dore et al., 2016). Person level variables include one's genes, dispositional characteristics, knowledge, memories, and appraisal tendencies as shaped by the accumulated effects of his/her life history, from the prenatal environment to early life influences to current experiences. Situational variables include the specific emotion-evocative stimuli being encountered as well as the social, ecological, and temporal context of that encounter. Strategy variables include the specific means chosen to regulate a response.

The value of spelling out this three-way interaction is that it highlights the need to consider all of these variables when understanding whether the use of a specific emotion regulation strategy was or was not successful for a given person, in a given situation. On this view, strategies are neither universally useful nor universally pernicious. Whether they help or hurt depends on who is using them in a given situation. As mentioned above, work on the strategy *selection* stage of the model already suggests that these interactions are important and powerful because it has shown that reappraisal is not always useful for everyone in every circumstance. Likewise, children and older adults, or specific clinical populations, may not be as able as young adults to use certain classes of strategies, either because they lack knowledge of and experience with them or because they depend on brain systems that are immature and are undergoing age-related decline or whose function is impacted by some sort of clinical disorder (Helion et al., 2019; Silvers, Buhle, & Ochsner, 2014). Future work will need to manipulate all three variables—person, situation, and strategy—to examine their inter-relationships and inter-dependencies.

Learning Learning must play a key role in our regulatory lives, and yet relatively little is known about how this happens. In theory, learning plays a role in every stage of our generalized model: we learn how to identify our emotions or the emotions of others, how to evaluate a situation to decide if regulation is needed, the range of strategies we can select, and how to implement them.

And there are both lifespan and training-related aspects to each of these factors. For example, it is essential to know how the environment influences the development of different brain systems supporting a child's growing ability to learn how to identify, evaluate, select, and implement. And for older adults, we need to know how these abilities and their underlying systems change with age. We need to know this because children and older adults are vulnerable populations—if we can identify relative weak points in their regulatory abilities, we could design training regimes to strengthen them.

Although there is a steadily growing literature on emotion regulation in children and older adults (like most extant research), this work focuses primarily on the implementation stage (Helion et al., 2019). To date, this work has shown that both groups are less able to use certain forms of reappraisal to downregulate negative emotion, that regulation of some appetitive impulses may be effective, and that older adults may effectively use attentional and situation-focused strategies (Allard & Kensinger, 2014; Livingstone & Isaacowitz, 2015; Opitz, Rauch, Terry, & Urry, 2012; Silvers, Insel, et al., 2016; Silvers et al., 2012; Silvers, Shu, Hubbard, Weber, & Ochsner, 2015; Winecoff, Labar, Madden, Cabeza, & Huettel, 2010). But whether or not older adults differ from young adults for the other stages of the model remains to be seen. Work on clinical populations suggests that there may be deficits for some populations in the ability to implement specific forms of reappraisal to downregulation negative emotion, although results have been inconsistent (Denny et al., 2014; Dillon & Pizzagalli, 2013; Johnstone, van Reekum, Urry, Kalin, & Davidson, 2007; Kanske, Heissler, Schonfelder, & Wessa, 2012; Koenigsberg et al., 2009; Silvers, Hubbard, et al., 2016).

It is possible the identification, evaluation, and selection stages may reveal more consistent differences between clinical and typically developing populations.

Assuming we can identify deficits in the ability to engage specific regulatory processes, then knowing how to address these deficits with training becomes an essential question. A growing number of studies are asking how training can impact the ability to implement strategies for self-regulation. By and large these studies have taken one of two approaches. The first provides training in a specific strategy, typically attentional control or reappraisal (Denny, Inhoff, Zerubavel, Davachi, & Ochsner, 2015; Denny & Ochsner, 2014). The second trains cognitive control abilities like working memory or selective attention with the assumption that the processes underlying these abilities are domain general (Cohen, Henik, & Moyal, 2012; Cohen, Moyal, & Henik, 2015). As such, strengthening these processes via working memory training may provide some benefit to any other behavior that taps into the same domain general processes—like reappraisal. We recently reviewed both areas of research (Cohen & Ochsner, 2018) and concluded that there is much promise, but much more work to be done—including asking how training can improve the identification, evaluation, and selection stages of the regulatory cycle.

Connection to Other Areas of Research

It is a truism that the mappings from behavior to psychological process to underlying brain systems are not one-to-one (Poldrack & Farah, 2015). Put another way, single behaviors arise from the concerted actions of multiple underlying psychological processes, and each process may be supported by a network of interacting brain regions. This can be visualized by thinking about multiple pathways connecting the behavioral level of analysis embodied in Fig. 3.1, the psychological levels of analyses described in Figs. 3.2 and 3.8, and the kinds of brain systems described in Fig. 3.3. When thinking about emotion regulation, this means that the brain systems we discuss—lateral and medial PFC, amygdala, striatum, and so on—all participate in processes that contribute to multiple other behaviors. Given this, we do our best to characterize behavior-process-brain mappings in our model of emotion regulation in ways that make sense in the context of related research. In this way, our thinking about the model is informed and constrained by widespread findings.

This influence can be bidirectional—we can also think about how the model may inform the way we think about other phenomena. For example, the model, as currently constructed, provides a means for conceptualizing the way in which interactions between individuals shape their emotional states. If we make the (perhaps strong) assumption that emotions provide the core of meaning for individuals—after all, emotions tell us how and why things matter to us and provide guidance in how to respond appropriately (Osgood, Suci, & Tannenbaum, 1957)—then our model of emotion regulation could be seen as a starting point for formulating a more comprehensive model of socio-emotional behavior.

The model might be well-suited for this given that it specifies processing stages for self or other that reflect core topics in the study of emotion, self-regulation, and social behavior more generally. For example, the identification stage corresponds to the study of emotion perception and social cognition more generally. The evaluation and selection stages relate to the study of affective decision processes and social cognition to the extent that they involve assessments of the impact of regulation on a target's mental state. Evaluation, selection, and implementation also draw on selection and working memory processes studied under the aegis of executive/cognitive control. And the entire bottom sequence that specifies the steps that trigger the emotions in need of regulation corresponds to the study of emotion generation more generally.

The model could also be broadened to account for social and self-regulatory phenomena occurring beyond dyads. As noted earlier, we and others have studied the way in which regulation occurs in the context of online groups where multiple people interact in a pairwise fashion (Dore & Morris, 2018; Dore, Weber, & Ochsner, 2017; Morris et al., 2015). The model could be expanded, however, to accommodate multi-person interactions, where the emotional responses of multiple possible regulation targets are being simultaneously identified by multiple people, all of whom have to evaluate whether or not regulation is needed. This could take place in the context of interactions on Facebook or other social media platforms where individuals or groups broadcast their (often emotionally charged) experiences to multiple others who are free to decide whether and how to respond in a variety of ways. And it can happen in person-to-person contexts as well. Anyone who has been a parent at the birthday party of small children knows relevant situations quite well, as multiple children may become a bit too obstreperous, rowdy, or combative, and multiple adults are witnessing this and evaluating whether they need to step in and help regulate.

These examples also illustrate a final important aspect of the model as it is instantiated in social contexts. Social regulation is embedded in the context of relationships of all kinds (Clark, Armentano, Boothby, & Hirsch, 2017; Eisenberg et al., 2000; Impett et al., 2010; Kneeland, Dovidio, Joormann, & Clark, 2016). In the social media example, we may feel more free to offer regulatory support to people to whom we are close, such as friends or family. In the parent-child example, a parent may be more comfortable intervening with regulatory support for their own as compared to other people's children. And how an adult or child appraises the meaning of regulatory assistance from a friend or family member—as helpful and wanted, for example, as compared to disruptive and annoying—will determine its effectiveness. In general, differences in status, friendship, age, the motivation one has for regulating others—and the target's perception of that motivation—along with other variables will significantly determine the efficacy of social regulatory interactions (Reeck et al., 2016; Williams, Morelli, Ong, & Zaki, 2018; Zaki & Williams, 2013).

Conclusion

If the day is still young for the study of many aspects of emotion regulation, then it is good that there is evident excitement within the field for their study. As has been noted, the study of emotion regulation has grown exponentially over the past 15 years (Gross, 2015). I began this chapter by observing that the lion's share of this work has concerned the implementation of strategies for the self-regulation of emotion in contexts where experimenters tell participants when and how to regulate using a strategy in which they have received some degree of instruction. This chapter closes with the hope that the many other aspects of regulation discussed here (and others have discussed elsewhere, e.g., Gross, 2015; Zaki & Williams, 2013) increasingly become the field's new focus.

Acknowledgments Completion of this manuscript was supported by grants AG057202 and AG043463 from NIA, AA023653 from NIAAA, and MH090964 from NIMH.

References

Allard, E. S., & Kensinger, E. A. (2014). Age-related differences in functional connectivity during cognitive emotion regulation. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 69*, 852. https://doi.org/10.1093/geronb/gbu108

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7*(4), 268–277.

Anderson, C. L., Monroy, M., & Keltner, D. (2017). Emotion in the wilds of nature: The coherence and contagion of fear during threatening group-based outdoors experiences. *Emotion, 18*, 355. https://doi.org/10.1037/emo0000378

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review, 84*(5), 413.

Arnsten, A. F. (2015). Stress weakens prefrontal networks: Molecular insults to higher cognition. *Nature Neuroscience, 18*(10), 1376–1385. https://doi.org/10.1038/nn.4087

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323

Beckes, L., & Coan, J. A. (2011). Social baseline theory: The role of social proximity in emotion and economy of action. *Social and Personality Psychology Compass, 5*(12), 976–988. https://doi.org/10.1111/j.1751-9004.2011.00400.x

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*(3), 624–652.

Braunstein, L. M., Gross, J. J., & Ochsner, K. N. (2017). Explicit and implicit emotion regulation: A multi-level framework. *Social Cognitive and Affective Neuroscience, 12*(10), 1545–1557. https://doi.org/10.1093/scan/nsx096

Buhle, J. T., Silvers, J. A., Wager, T. D., Lopez, R., Onyemekwu, C., Kober, H., … Ochsner, K. N. (2014). Cognitive reappraisal of emotion: A meta-analysis of human neuroimaging studies. *Cerebral Cortex, 24*(11), 2981–2990. https://doi.org/10.1093/cercor/bht154

Chang, L. J., Gianaros, P. J., Manuck, S. B., Krishnan, A., & Wager, T. D. (2015). A sensitive and specific neural signature for picture-induced negative affect. *PLoS Biology, 13*(6), e1002180. https://doi.org/10.1371/journal.pbio.1002180

Clark, M. S., Armentano, L. A., Boothby, E. J., & Hirsch, J. L. (2017). Communal relational context (or lack thereof) shapes emotional lives. *Current Opinion in Psychology, 17*, 176–183. https://doi.org/10.1016/j.copsyc.2017.07.023

Cohen, N., Henik, A., & Moyal, N. (2012). Executive control attenuates emotional effects-for high reappraisers only? *Emotion, 12*(5), 970–979. https://doi.org/10.1037/a0026890

Cohen, N., Moyal, N., & Henik, A. (2015). Executive control suppresses pupillary responses to aversive stimuli. *Biological Psychology, 112*, 1–11. https://doi.org/10.1016/j.biopsycho.2015.09.006

Cohen, N., & Ochsner, K. N. (2018). The emerging science of emotion regulation training. *Current Opinion in Behavioral Sciences, 24*, 143–155.

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

D'Esposito, M., Postle, B. R., Ballard, D., & Lease, J. (1999). Maintenance versus manipulation of information held in working memory: An event-related fMRI study. *Brain and Cognition, 41*(1), 66–86.

Denny, B. T., Fan, J., Liu, X., Ochsner, K. N., Guerreri, S., Mayson, S. J., … Koenigsberg, H. W. (2014). Elevated amygdala activity during reappraisal anticipation predicts anxiety in avoidant personality disorder. *Journal of Affective Disorders, 172C*, 1–7. https://doi.org/10.1016/j.jad.2014.09.017

Denny, B. T., Inhoff, M. C., Zerubavel, N., Davachi, L., & Ochsner, K. N. (2015). Getting over it: Long-lasting effects of emotion regulation on amygdala response. *Psychological Science, 26*(9), 1377–1388. https://doi.org/10.1177/0956797615578863

Denny, B. T., & Ochsner, K. N. (2014). Behavioral effects of longitudinal training in cognitive reappraisal. *Emotion, 14*(2), 425–433. https://doi.org/10.1037/a0035276

Dillon, D. G., & Pizzagalli, D. A. (2013). Evidence of successful modulation of brain activation and subjective experience during reappraisal of negative emotion in unmedicated depression. *Psychiatry Research, 212*(2), 99–107. https://doi.org/10.1016/j.pscychresns.2013.01.001

Dore, B. P., & Morris, R. R. (2018). Linguistic synchrony predicts the immediate and lasting impact of text-based emotional support. *Psychological Science, 29*, 956797618779971. https://doi.org/10.1177/0956797618779971

Dore, B. P., Morris, R. R., Burr, D. A., Picard, R. W., & Ochsner, K. N. (2017). Helping others regulate emotion predicts increased regulation of one's own emotions and decreased symptoms of depression. *Personality and Social Psychology Bulletin, 43*(5), 729–739. https://doi.org/10.1177/0146167217695558

Dore, B. P., Silvers, J. A., & Ochsner, K. N. (2016). Towards a personalized science of emotion regulation. *Social and Personality Psychology Compass, 10*(4), 171–187.

Dore, B. P., Weber, J., & Ochsner, K. N. (2017). Neural predictors of decisions to cognitively control emotion. *The Journal of Neuroscience, 37*(10), 2580–2588. https://doi.org/10.1523/JNEUROSCI.2526-16.2016

Drevets, W. C., & Raichle, M. E. (1998). Reciprocal suppression of regional cerebral blood flow during emotional versus higher cognitive processes: Implications for interactions between emotion and cognition. *Cognition & Emotion, 12*(3), 353–385.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*(3968), 303–306.

Eisenberg, N., Guthrie, I. K., Fabes, R. A., Shepard, S., Losoya, S., Murphy, B. C., … Reiser, M. (2000). Prediction of elementary school children's externalizing problem behaviors from attentional and behavioral regulation and negative emotionality. *Child Development, 71*(5), 1367–1382.

Gilead, M., Boccagno, C., Silverman, M., Hassin, R. R., Weber, J., & Ochsner, K. N. (2016). Self-regulation via neural simulation. *Proceedings of the National Academy of Sciences of the United States of America, 113*(36), 10037–10042. https://doi.org/10.1073/pnas.1600159113

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*, 271–299.

Gross, J. J. (2015). The extended process model of emotion regulation: Elaborations, applications, and future directions. *Psychological Inquiry, 26*(1), 130–137. https://doi.org/10.1080/1047840X.2015.989751

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology, 85*, 348–362.

Harnad, S. (1987). *Categorical perception: The groundwork of cognition*. New York, NY: Cambridge University Press.

Hay, A. C., Sheppes, G., Gross, J. J., & Gruber, J. (2015). Choosing how to feel: Emotion regulation choice in bipolar disorder. *Emotion, 15*(2), 139–145. https://doi.org/10.1037/emo0000024

Heatherton, T. F., & Wagner, D. D. (2011). Cognitive neuroscience of self-regulation failure. *Trends in Cognitive Sciences, 15*(3), 132–139. https://doi.org/10.1016/j.tics.2010.12.005

Helion, C., Krueger, S., & Ochsner, K. N. (2019). Emotion regulation across the lifespan. In J. Grafman & M. D'Esposito (Eds.), *The Handbook of Clinical Neurology: The Frontal Lobes (3rd ed)* (Vol. 163, pp. 257–280). New York: Elsevier; US.

Impett, E. A., Gordon, A. M., Kogan, A., Oveis, C., Gable, S. L., & Keltner, D. (2010). Moving toward more perfect unions: Daily and long-term consequences of approach and avoidance goals in romantic relationships. *Journal of Personality and Social Psychology, 99*(6), 948–963. https://doi.org/10.1037/a0020271

Izuma, K., & Adolphs, R. (2013). Social manipulation of preference in the human brain. *Neuron, 78*(3), 563–573. https://doi.org/10.1016/j.neuron.2013.03.023

Johnstone, T., van Reekum, C. M., Urry, H. L., Kalin, N. H., & Davidson, R. J. (2007). Failure to regulate: Counterproductive recruitment of top-down prefrontal-subcortical circuitry in major depression. *The Journal of Neuroscience, 27*(33), 8877–8884.

Jordan, J. J., Rand, D. G., Arbesman, S., Fowler, J. H., & Christakis, N. A. (2013). Contagion of cooperation in static and fluid social networks. *PLoS One, 8*(6), e66199. https://doi.org/10.1371/journal.pone.0066199

Kanske, P., Heissler, J., Schonfelder, S., & Wessa, M. (2012). Neural correlates of emotion regulation deficits in remitted depression: The influence of regulation strategy, habitual regulation use, and emotional valence. *NeuroImage, 61*(3), 686–693. https://doi.org/10.1016/j.neuroimage.2012.03.089

Kircanski, K., Lieberman, M. D., & Craske, M. G. (2012). Feelings into words: Contributions of language to exposure therapy. *Psychological Science, 23*(10), 1086–1091. https://doi.org/10.1177/0956797612443830

Klucharev, V., Hytonen, K., Rijpkema, M., Smidts, A., & Fernandez, G. (2009). Reinforcement learning signal predicts social conformity. *Neuron, 61*(1), 140–151. https://doi.org/10.1016/j.neuron.2008.11.027

Klucharev, V., Munneke, M. A., Smidts, A., & Fernandez, G. (2011). Downregulation of the posterior medial frontal cortex prevents social conformity. *The Journal of Neuroscience, 31*(33), 11934–11940. https://doi.org/10.1523/JNEUROSCI.1869-11.2011

Kneeland, E. T., Dovidio, J. F., Joormann, J., & Clark, M. S. (2016). Emotion malleability beliefs, emotion regulation, and psychopathology: Integrating affective and clinical science. *Clinical Psychology Review, 45*, 81–88. https://doi.org/10.1016/j.cpr.2016.03.008

Koenigsberg, H. W., Siever, L. J., Lee, H., Pizzarello, S., New, A. S., Goodman, M., … Prohovnik, I. (2009). Neural correlates of emotion processing in borderline personality disorder. *Psychiatry Research, 172*(3), 192–199. https://doi.org/10.1016/j.pscychresns.2008.07.010

Kohn, N., Eickhoff, S. B., Scheller, M., Laird, A. R., Fox, P. T., & Habel, U. (2014). Neural network of cognitive emotion regulation—An ALE meta-analysis and MACM analysis. *NeuroImage, 87*, 345–355. https://doi.org/10.1016/j.neuroimage.2013.11.001

Konnikova, M. (2013). *The limits of self-control: Self-control, illusory control, and risky financial decision making*. Psychology, Columbia University, Columbia University Academic Commons.

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.

Lane, R. D., Fink, G. R., Chau, P. M., & Dolan, R. J. (1997). Neural activation during selective attention to subjective emotional responses. *Neuroreport, 8*(18), 3969–3972.

Lau, H. C., & Passingham, R. E. (2007). Unconscious activation of the cognitive control system in the human prefrontal cortex. *The Journal of Neuroscience, 27*(21), 5805–5811. https://doi.org/10.1523/JNEUROSCI.4335-06.2007

Lieberman, M. D., Eisenberger, N. I., Crockett, M. J., Tom, S. M., Pfeifer, J. H., & Way, B. M. (2007). Putting feelings into words: Affect labeling disrupts amygdala activity in response to affective stimuli. *Psychological Science, 18*(5), 421–428.

Livingstone, K. M., & Isaacowitz, D. M. (2015). Situation selection and modification for emotion regulation in younger and older adults. *Social Psychological and Personality Science, 6*(8), 904–910. https://doi.org/10.1177/1948550615593148

Maier, S. F. (2015). Behavioral control blunts reactions to contemporaneous and future adverse events: Medial prefrontal cortex plasticity and a corticostriatal network. *Neurobiology of Stress, 1*, 12–22. https://doi.org/10.1016/j.ynstr.2014.09.003

Martin, R. E., Silvers, J. A., Hardi, F., Stephano, T., Helion, C., Insel, C., Franz, P. J., Ninova, E., Lander, J. P., Mischel, W., Casey, B. J. & Ochsner, K. N. (in press). Longitudinal changes in brain structures related to appetitive reactivity and regulation across development. *Developmental Cognitive Neuroscience*.

Mauss, I. B., Shallcross, A. J., Troy, A. S., John, O. P., Ferrer, E., Wilhelm, F. H., & Gross, J. J. (2011). Don't hide your happiness! Positive emotion dissociation, social connectedness, and psychological functioning. *Journal of Personality and Social Psychology, 100*(4), 738–748. https://doi.org/10.1037/a0022410

Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological Review, 106*(1), 3–19.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*, 167–202.

Morawetz, C., Bode, S., Derntl, B., & Heekeren, H. R. (2017). The effect of strategies, goals and stimulus material on the neural mechanisms of emotion regulation: A meta-analysis of fMRI studies. *Neuroscience and Biobehavioral Reviews, 72*, 111–128. https://doi.org/10.1016/j.neubiorev.2016.11.014

Morris, R. R., Schueller, S. M., & Picard, R. W. (2015). Efficacy of a web-based, crowdsourced peer-to-peer cognitive reappraisal platform for depression: Randomized controlled trial. *Journal of Medical Internet Research, 17*(3), e72. https://doi.org/10.2196/jmir.4167

Muscatell, K. A., Dedovic, K., Slavich, G. M., Jarcho, M. R., Breen, E. C., Bower, J. E., … Eisenberger, N. I. (2015). Greater amygdala activity and dorsomedial prefrontal-amygdala coupling are associated with enhanced inflammatory responses to stress. *Brain, Behavior, and Immunity, 43*, 46–53. https://doi.org/10.1016/j.bbi.2014.06.201

Neumann, R., & Strack, F. (2000). "Mood contagion": The automatic transfer of mood between persons. *Journal of Personality and Social Psychology, 79*(2), 211–223.

Nook, E. C., & Zaki, J. (2015). Social norms shift behavioral and neural responses to foods. *Journal of Cognitive Neuroscience, 27*(7), 1412–1426. https://doi.org/10.1162/jocn_a_00795

O'Driscoll, C., Laing, J., & Mason, O. (2014). Cognitive emotion regulation strategies, alexithymia and dissociation in schizophrenia, a review and meta-analysis. *Clinical Psychology Review, 34*(6), 482–495. https://doi.org/10.1016/j.cpr.2014.07.002

Ochsner, K. N. (2013). The role of control in emotion, emotion regulation and empathy. In D. Hermans, B. Rime, & B. Mesquita (Eds.), *Changing emotions* (pp. 157–165). New York, NY: Psychology Press.

Ochsner, K. N. (2014). What is the role of control in emotional life? In M. S. Gazzaniga & G. R. Mangun (Eds.), *The cognitive neurosciences* (5th ed., pp. 719–730). Cambridge, MA: MIT Press.

Ochsner, K. N., Bunge, S. A., Gross, J. J., & Gabrieli, J. D. (2002). Rethinking feelings: An FMRI study of the cognitive regulation of emotion. *Journal of Cognitive Neuroscience, 14*(8), 1215–1229.

Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences, 9*(5), 242–249.

Ochsner, K. N., & Gross, J. J. (2014). The neural bases of emotion and emotion regulation: A valuation perspective. In J. J. Gross & R. H. Thompson (Eds.), *Handbook of emotion regulation* (Vol. 2, 2nd ed., pp. 23–42). New York, NY: Guilford Press.

Ochsner, K. N., Knierim, K., Ludlow, D., Hanelin, J., Ramachandran, T., & Mackey, S. (2004). Reflecting upon feelings: An fMRI study of neural systems supporting the attribution of emotion to self and other. *Journal of Cognitive Neuroscience, 16*(10), 1746–1772.

Ochsner, K. N., Ray, R. D., Cooper, J. C., Robertson, E. R., Chopra, S., Gabrieli, J. D. E., & Gross, J. J. (2004). For better or for worse: Neural systems supporting the cognitive down- and up-regulation of negative emotion. *NeuroImage, 23*(2), 483–499.

Ochsner, K. N., Silvers, J. A., & Buhle, J. T. (2012). Functional imaging studies of emotion regulation: A synthetic review and evolving model of the cognitive control of emotion. *Annals of the New York Academy of Sciences, 1251*, E1–E24. https://doi.org/10.1111/j.1749-6632.2012.06751.x

Oei, N. Y., Everaerd, W. T., Elzinga, B. M., van Well, S., & Bermond, B. (2006). Psychosocial stress impairs working memory at high loads: An association with cortisol levels and memory retrieval. *Stress, 9*(3), 133–141. https://doi.org/10.1080/10253890600965773

Opitz, P. C., Rauch, L. C., Terry, D. P., & Urry, H. L. (2012). Prefrontal mediation of age differences in cognitive reappraisal. *Neurobiology of Aging, 33*(4), 645–655. https://doi.org/10.1016/j.neurobiolaging.2010.06.004

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Oxford, England: University of Illinois Press.

Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences of the United States of America, 110*(52), 20941–20946. https://doi.org/10.1073/pnas.1312011110

Peters, A., McEwen, B. S., & Friston, K. (2017). Uncertainty and stress: Why it causes diseases and how it is mastered by the brain. *Progress in Neurobiology, 156*, 164–188. https://doi.org/10.1016/j.pneurobio.2017.05.004

Phan, K. L., Taylor, S. F., Welsh, R. C., Decker, L. R., Noll, D. C., Nichols, T. E., … Liberzon, I. (2003). Activation of the medial prefrontal cortex and extended amygdala by individual ratings of emotional arousal: A fMRI study. *Biological Psychiatry, 53*(3), 211–215.

Poldrack, R. A., & Farah, M. J. (2015). Progress and challenges in probing the human brain. *Nature, 526*(7573), 371–379. https://doi.org/10.1038/nature15692

Raio, C. M., Orederu, T. A., Palazzolo, L., Shurick, A. A., & Phelps, E. A. (2013). Cognitive emotion regulation fails the stress test. *Proceedings of the National Academy of Sciences of the United States of America, 110*(37), 15139–15144. https://doi.org/10.1073/pnas.1305706110

Raio, C. M., & Phelps, E. A. (2015). The influence of acute stress on the regulation of conditioned fear. *Neurobiology of Stress, 1*, 134–146. https://doi.org/10.1016/j.ynstr.2014.11.004

Reeck, C., Ames, D. R., & Ochsner, K. N. (2016). The social regulation of emotion: An integrative, cross-disciplinary model. *Trends in Cognitive Sciences, 20*(1), 47–63. https://doi.org/10.1016/j.tics.2015.09.003

Reynaud, E., Guedj, E., Trousselard, M., El Khoury-Malhame, M., Zendjidjian, X., Fakra, E., … Khalfa, S. (2015). Acute stress disorder modifies cerebral activity of amygdala and prefrontal cortex. *Cognitive Neuroscience, 6*(1), 39–43. https://doi.org/10.1080/17588928.2014.996212

Sapolsky, R. M. (2015). Stress and the brain: Individual variability and the inverted-U. *Nature Neuroscience, 18*(10), 1344–1346. https://doi.org/10.1038/nn.4109

Satpute, A. B., Badre, D., & Ochsner, K. N. (2014). Distinct regions of prefrontal cortex are associated with the controlled retrieval and selection of social information. *Cerebral Cortex, 24*(5), 1269–1277. https://doi.org/10.1093/cercor/bhs408

Satpute, A. B., Nook, E. C., Narayanan, S., Shu, J., Weber, J., & Ochsner, K. N. (2016). Emotions in "black and white" or shades of gray? How we think about emotion shapes our perception

and neural representation of emotion. *Psychological Science, 27*(11), 1428–1442. https://doi.org/10.1177/0956797616661555

Satpute, A. B., Shu, J., Weber, J., Roy, M., & Ochsner, K. N. (2013). The functional neural architecture of self-reports of affective experience. *Biological Psychiatry, 73*(7), 631–638. https://doi.org/10.1016/j.biopsych.2012.10.001

Sauer, C., Sheppes, G., Lackner, H. K., Arens, E. A., Tarrasch, R., & Barnow, S. (2016). Emotion regulation choice in female patients with borderline personality disorder: Findings from self-reports and experimental measures. *Psychiatry Research, 242*, 375–384. https://doi.org/10.1016/j.psychres.2016.04.113

Scheibe, S., Sheppes, G., & Staudinger, U. M. (2015). Distract or reappraise? Age-related differences in emotion-regulation choice. *Emotion, 15*(6), 677–681. https://doi.org/10.1037/a0039246

Sheppes, G., & Levin, Z. (2013). Emotion regulation choice: Selecting between cognitive regulation strategies to control emotion. *Frontiers in Human Neuroscience, 7*, 179. https://doi.org/10.3389/fnhum.2013.00179

Sheppes, G., Scheibe, S., Suri, G., & Gross, J. J. (2011). Emotion-regulation choice. *Psychological Science, 22*(11), 1391–1396. https://doi.org/10.1177/0956797611418350

Sheppes, G., Scheibe, S., Suri, G., Radu, P., Blechert, J., & Gross, J. J. (2014). Emotion regulation choice: A conceptual framework and supporting evidence. *Journal of Experimental Psychology. General, 143*(1), 163–181. https://doi.org/10.1037/a0030831

Silvers, J. A., Buhle, J. T., & Ochsner, K. N. (2014). The neuroscience of emotion regulation: Basic mechanisms and their role in development, aging and psychopathology. In K. N. Ochsner & S. M. Kosslyn (Eds.), *The Oxford handbook of cognitive neuroscience: Vol. 2. The cutting edges* (pp. 58–73). New York, NY: Oxford University Press.

Silvers, J. A., Hubbard, A. D., Biggs, E., Shu, J., Fertuck, E., Chaudhury, S., … Stanley, B. (2016). Affective lability and difficulties with regulation are differentially associated with amygdala and prefrontal response in women with Borderline Personality Disorder. *Psychiatry Research, 254*, 74–82. https://doi.org/10.1016/j.pscychresns.2016.06.009

Silvers, J. A., Insel, C., Powers, A., Franz, P., Helion, C., Martin, R. E., … Ochsner, K. N. (2016). vlPFC-vmPFC-amygdala interactions underlie age-related differences in cognitive regulation of emotion. *Cerebral Cortex, 25*, 128–137. https://doi.org/10.1093/cercor/bhw073

Silvers, J. A., McRae, K., Gabrieli, J. D., Gross, J. J., Remy, K. A., & Ochsner, K. N. (2012). Age-related differences in emotional reactivity, regulation, and rejection sensitivity in adolescence. *Emotion, 12*(6), 1235–1247. https://doi.org/10.1037/a0028297

Silvers, J. A., Shu, J., Hubbard, A. D., Weber, J., & Ochsner, K. N. (2015). Concurrent and lasting effects of emotion regulation on amygdala response in adolescence and young adulthood. *Developmental Science, 18*(5), 771–784. https://doi.org/10.1111/desc.12260

Suri, G., Sheppes, G., & Gross, J. J. (2013). Predicting affective choice. *Journal of Experimental Psychology. General, 142*(3), 627–632. https://doi.org/10.1037/a0029900

Taylor, S. F., Phan, K. L., Decker, L. R., & Liberzon, I. (2003). Subjective rating of emotionally salient stimuli modulates neural activity. *NeuroImage, 18*(3), 650–659.

Todd, R. M., Cunningham, W. A., Anderson, A. K., & Thompson, E. (2012). Affect-biased attention as emotion regulation. *Trends in Cognitive Sciences, 16*(7), 365–372. https://doi.org/10.1016/j.tics.2012.06.003

Troy, A. S., Ford, B. Q., McRae, K., Zarolia, P., & Mauss, I. B. (2017). Change the things you can: Emotion regulation is more beneficial for people from lower than from higher socioeconomic status. *Emotion, 17*(1), 141–154. https://doi.org/10.1037/emo0000210

Troy, A. S., Shallcross, A. J., & Mauss, I. B. (2013). A person-by-situation approach to emotion regulation: Cognitive reappraisal can either help or hurt, depending on the context. *Psychological Science, 24*(12), 2505–2514. https://doi.org/10.1177/0956797613496434

Uy, J. P., & Galvan, A. (2017). Acute stress increases risky decisions and dampens prefrontal activation among adolescent boys. *NeuroImage, 146*, 679–689. https://doi.org/10.1016/j.neuroimage.2016.08.067

van Ast, V. A., Spicer, J., Smith, E. E., Schmer-Galunder, S., Liberzon, I., Abelson, J. L., & Wager, T. D. (2016). Brain mechanisms of social threat effects on working memory. *Cerebral Cortex, 26*(2), 544–556. https://doi.org/10.1093/cercor/bhu206

Williams, W. C., Morelli, S. A., Ong, D. C., & Zaki, J. (2018). Interpersonal emotion regulation: Implications for affiliation, perceived support, relationships, and well-being. *Journal of Personality and Social Psychology, 115*(2), 224–254. https://doi.org/10.1037/pspi0000132

Winecoff, A., Labar, K. S., Madden, D. J., Cabeza, R., & Huettel, S. A. (2010). Cognitive and neural contributors to emotion regulation in aging. *Social Cognitive and Affective Neuroscience, 6*(2), 165–176. https://doi.org/10.1093/scan/nsq030

Zaki, J., & Ochsner, K. (2012). The neuroscience of empathy: Progress, pitfalls and promise. *Nature Neuroscience, 15*(5), 675–680. https://doi.org/10.1038/nn.3085

Zaki, J., Schirmer, J., & Mitchell, J. P. (2011). Social influence modulates the neural computation of value. *Psychological Science, 22*(7), 894–900. https://doi.org/10.1177/0956797611411057

Zaki, J., & Williams, W. C. (2013). Interpersonal emotion regulation. *Emotion, 13*(5), 803–810. https://doi.org/10.1037/a0033839

Chapter 4
Bringing Together Cognitive and Genetic Approaches to the Understanding of Stress Vulnerability and Psychological Well-Being

Elaine Fox and Robert Keers

Stress vulnerability refers to a basic susceptibility to developing mental health disorders, and much of the research in this field has focused on the highly comorbid conditions of depression and anxiety. Multiple factors are undoubtedly involved in the etiology of depression and anxiety as well as in determining resilience to these disorders. However, different lines of research have identified the two broad categories of genetic vulnerability (e.g., Dunn et al., 2015) and cognitive vulnerability (e.g., Reilly, Ciesla, Felton, Weitlauf, & Anderson, 2012) as constituting particularly important risk factors for the development of these disorders. While these two research literatures are both extensive, the investigation of genetic and cognitive factors together is rare, and it has been argued recently that these disparate fields could be fruitfully combined to develop a deeper understanding of psychopathology and psychological well-being (Fox & Beevers, 2016).

In this chapter, we aim to provide an overview of both cognitive and genetic approaches to mental health in order to encourage greater collaboration between those who typically take a cognitive approach to experimental psychopathology and those working in molecular genetics, who typically take a more biological approach. Because of the extensive nature of both fields, our review is necessarily selective. Nevertheless, we hope that this chapter will encourage greater interdisciplinary collaboration between geneticists and cognitive clinical scientists in a quest to deepen our understanding of psychopathology and psychological well-being. We believe that there are many opportunities for researchers to combine forces to help us move

E. Fox (✉)
Department of Experiment Psychology, University of Oxford, New Radcliffe House, Radcliffe Observatory Quarter, Oxford, UK
e-mail: Elaine.fox@psy.ox.ac.uk

R. Keers
School of Biological and Chemical Sciences, Queen Mary University of London, London, UK
e-mail: r.keers@qmul.ac.uk

M. Neta, I. J. Haas (eds.), *Emotion in the Mind and Body*, Nebraska Symposium on Motivation 66, https://doi.org/10.1007/978-3-030-27473-3_4

closer to understanding why some individuals appear to be particularly vulnerable to emotional disorders while others seem to be more resilient to the onset of disorder.

An Integrated Model of Cognitive Biases, Genes, and Emotional Disorders

Expanding on the CogBIAS hypothesis outlined by Fox and Beevers (2016), we present a broad theoretical framework outlining how different genetic factors and cognitive biases may work together across development in the etiology of emotional disorders and well-being (see Fig. 4.1).

Our central hypothesis is that cognitive biases may lie on a causal pathway between genetic influences and psychopathology. Both quantitative and molecular genetic studies have reported genetic effects on emotional disorders and well-being. These genetic variants are likely to include (1) those that have a main effect on emotional disorders or well-being, (2) those that alter sensitivity to stress (stress-sensitivity variants), (3) those that influence response to positive environmental influences (vantage-sensitivity variants), and (4) those that increase plasticity more

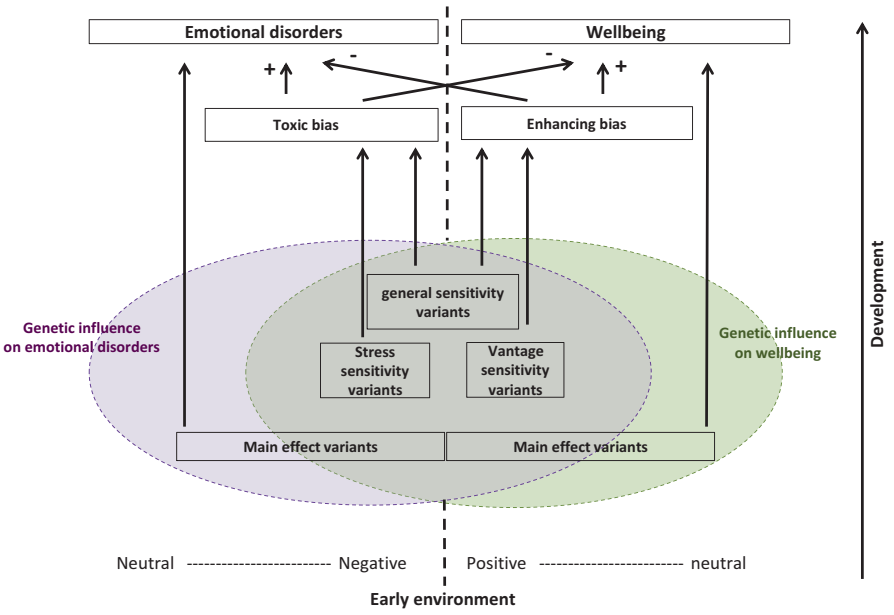

Fig. 4.1 The CogBIAS model. A developmental model of how common genetic variants and cognitive biases influence psychological wellbeing and psychopathology

generally, affecting responses to both positive and negative environmental influences (general-sensitivity variants).

The proposed effects of each of these variants on cognitive biases, emotional disorders, and well-being are delineated in the context of the early environment, which is presented along a neutral-negative and neutral-positive axis. Genetic variants that are purported to have a main effect on emotional disorders or well-being are represented by means of a direct pathway leading from genetic risk to the outcome of interest. These factors are hypothesized to be unaffected by early environmental influences and not mediated by cognitive biases. Genetic variants that increase sensitivity to negative environments (stress-sensitivity variants) lead to a negative cognitive bias but only in the context of early adversity such as bullying or child abuse. In contrast, genetic variants that increase sensitivity to positive environments (vantage-sensitivity variants) lead to an enhancing cognitive bias but only in the context of a positive early environment such as a supportive and enriched family environment. Finally, genetic variants that increase sensitivity to both negative and positive environments (general-sensitivity variants) may lead to either toxic or enhancing biases again depending on the early environmental context.

The development of toxic or enhancing cognitive biases across development leaves people more open and sensitive to negative or positive environmental influences, resulting in different life trajectories that tip individuals toward either well-being or mental ill health, respectively. In the remainder of this chapter, we provide a selective overview of research on both cognitive and genetic influences on emotional vulnerability in the context of our theoretical framework. We then consider some important challenges for developing translational research programs that incorporate these two important influences on human development.

Cognitive Approaches

Psychological science underwent a transformative "cognitive revolution" in the 1950s and 1960s that altered fundamentally the nature of the discipline. In his classic book *Cognitive Psychology*, Neisser (1967) described *cognition* as "all processes by which the sensory input is transformed, reduced, elaborated, stored, recovered, and used" (p. 4). He went on to say that "cognition is involved in everything a human being might possibly do; that every psychological phenomenon is a cognitive phenomenon" (p. 4). It is no surprise then that throughout the 1960s and 1970s, the cognitive approach swept through to every corner of psychological science. Cognitive models were proposed to explain almost all aspects of human behavior, and our understanding of emotion and emotional disorders was no exception.

Cognitive Approaches to Emotional Disorders: Clinical Perspectives

Cognitive Vulnerability Models of Anxiety and Depression

Aaron Beck (1967), in the very same year as Neisser's seminal book on cognitive psychology was published, presented a compelling case that emotional disorders, such as depression and anxiety, were sustained by systematic negative distortions in a cognitive triad of thinking about the world, the self, and the future. This cognitive approach shattered long-held assumptions that emotional disorders were primarily caused by fundamental biological dysfunctions and paved the way for new avenues of research and therapeutic approaches. Beck's model generated cognitive behavioral therapy (CBT), which has become one of the most successful and widely implemented psychological interventions for a wide range of emotional disorders (Hofmann, Asnaani, Vonk, Sawyer, & Fang, 2012), and is considered by some to be the current "gold standard" of psychotherapy (David, Cristea, & Hofmann, 2018).

More recently, there have been efforts to incorporate neurobiological findings into the cognitive approach (Clark & Beck, 2010). There is some evidence, for instance, that the efficacy of CBT might be associated with a simultaneous reduction of activation in amygdalohippocampal subcortical regions—known to be involved in the generation of negative emotional states—alongside increased activation of prefrontal cortical regions involved in the cognitive control of negative emotion (see Clark & Beck, 2010). While Beck's cognitive model has continued to be updated and modified throughout the years, its essence remains that various aspects of cognition are responsible for the maintenance of depression and anxiety.

The most important cognitive elements highlighted by Beck (1967) are what he called *schemas*, which refer to hypothetical cognitive structures that can be inferred from persistent themes in a person's thoughts and images. Someone suffering from depression, for instance, may have a repetitive and persistent thought such as "I am a failure," whereas someone with an anxiety disorder may have persistent thoughts such as "I am in danger." The central idea is that such negative beliefs about the self, the world, and the future become rigid, resistant to reason, and very difficult to change. Schemas are thought to be automatic in the sense that they are easily activated by a wide range of life events or internal thoughts and, once activated, tend to dominate the entire cognitive system resulting in *biased information processing*.

This biased information processing denotes the preferential encoding and retrieval of any information that is congruent with a current mood state. The proposal is that mood-congruent cognitive biases typically magnify negative and self-referential material in the case of depression and drive selective processing of information relating to threat, danger, and helplessness in the case of anxiety. The outcome of this biased information processing system, according to Beck's model, is the subjective experience of schema-congruent *negative automatic thoughts*, which include a number of common cognitive distortions such as drawing conclusions from insufficient evidence (*arbitrary inference*) or drawing conclusions from

just one aspect of a situation (*selective abstraction*) to name just two. These negative automatic thoughts are suggested to flow from the primary negative schemas that a person might hold. When looking for a new job, for instance, a depressed person who holds a schematic belief such as "I'm a loser" might have a range of negative automatic thoughts such as "I'll probably screw this up," "there's no point in even trying," and "I'm never going to get ahead." Beck used the descriptions of the symptoms that patients gave and collated these into a 21-item self-report instrument. The Beck Depression Inventory, which was updated in 1996, is widely used to provide a subjective assessment of depression severity in both basic research and clinical research (Beck, Steer, & Brown, 1996).

An important feature of Beck's model is the developmental aspect, which assumes that schemas are formed by early life experiences such as being bullied, criticized, or abused. Once these schemas are formed and consolidated, they can culminate in the production of a range of persistent negative automatic thoughts that, in turn, have pervasive effects on cognition, behavior, and emotions. A simplified schematic of Beck's model is illustrated in Fig. 4.2.

The primary purpose of Beck's cognitive model was to gain a better understanding of emotional disorders and to guide treatment approaches, and in that it has been very successful. The idea of cognitive structures (schemas) that (1) are laid down early in life, (2) are easily activated by trigger stimuli, (3) are often infused with affective valence, (4) bias information processing, and (5) have pervasive effects on behavior, thoughts, and feelings makes intuitive sense and has been theoretically useful in many areas of psychology beyond clinical science (e.g., Devine, 1989).

Beck's cognitive theory has led to the development of other cognitive vulnerability models such as the *hopelessness theory of depression* (Abramson, Alloy, & Metalsky, 1989). This theory proposes that an important cognitive cause of depressive symptoms is the general expectation that undesirable outcomes will occur while desirable outcomes will not occur and that there is little one can do about this situation. Just as in Beck's (1967) model, negative life events can become trigger points for people to develop depression. In Abramson et al.'s (1989) approach, however, three types of inference about negative situations are thought to contribute to the development of hopelessness and, in turn, depression. Specifically, depression is likely to follow when negative life events are:

1. Attributed to stable and global causes;
2. Perceived as likely to lead to other negative consequences;
3. Perceived as implying that the person is worthless or deficient in some way.

If a student fails an exam, for instance, she might make the series of attributions that this (a) is because she is not intelligent enough, (b) is likely to prevent her from graduating and hence not be able to go to university, and (c) means that she is worthless and useless. According to hopelessness theory, it is this chain of negative inferences that leads to depression. In contrast, somebody without this cognitive vulnerability in the same situation might make the attribution that (a) she did not study hard enough; (b) she now must do far more work in order to do well in future examinations; and (c) this failure has no implications for self-worth. As with Beck's

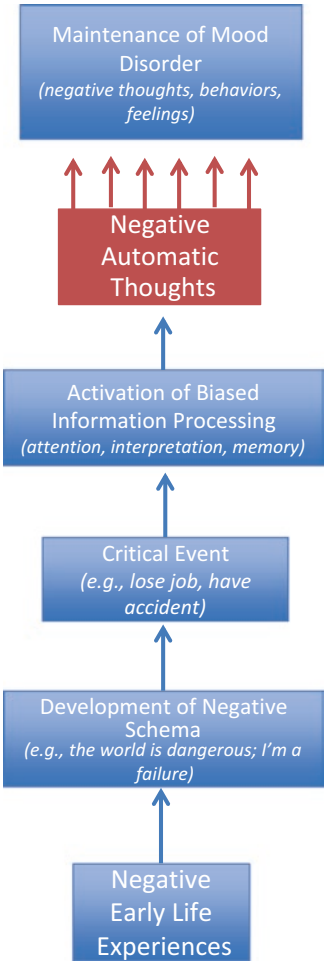

Fig. 4.2 A schematic outline of Beck's model of emotional disorders

model, the interaction of a cognitive vulnerability with a negative life event is central. Importantly, in the absence of a negative situation, people with a depressogenic inferential style are not considered to be at any higher risk of depression.

Testing Cognitive Vulnerability Models

A powerful method of testing a *cognitive vulnerability* hypothesis is to utilize what has been called a behavioral high-risk design. This is a method in which participants who do not currently have the disorder of interest (e.g., depression), but who are hypothesized to be at either high or low risk on the basis of their status on a measure of cognitive vulnerability to the disorder, are recruited and studied over a specified period of time in a longitudinal study (e.g., Alloy, Abramson, Whitehouse, & Hogan, 2006).

Using this approach, Alloy et al. (2006) recruited 347 currently non-depressed university students for the Temple-Wisconsin Cognitive Vulnerability to Depression Project who were judged to be at either high (n = 172) or low (n = 175) risk for depression on the basis of their cognitive style as assessed by the Cognitive Style Questionnaire (CSQ; Alloy et al., 2000) and the Dysfunctional Attitudes Scale (DAS; Weissman & Beck, 1978). Over a 2.5-year period, the onset of major depression was found to be about seven times higher in those with high cognitive risk relative to those individuals with low cognitive risk for depression (Alloy et al., 2006). Thus, consistent with a cognitive vulnerability hypothesis, non-depressed individuals who demonstrated negative inferential cognitive styles alongside dysfunctional attitudes had a much higher likelihood of experiencing prospective onsets of depression. This was an important finding as it provided early empirical evidence that a negative cognitive style can indeed confer vulnerability to clinically diagnosed depression rather than just to mild levels of nonclinical depressive symptomatology (Coyne & Gotlib, 1983).

In summary, a large body of work within clinical psychology has utilized a cognitive perspective and, consistent with our CogBIAS model, has demonstrated that cognitive functions are central to the development and maintenance of affective disorders such as depression and anxiety.

Cognitive Approaches to Emotional Disorders: Experimental Psychopathology Perspectives

A Network-Based Model

Coming from a completely different perspective, Bower (1981, 1987) developed a network theory of emotion to develop a deeper understanding of the relations among cognitive and affective processes. Unlike clinical theorists (e.g., Beck, 1967) whose starting point was self-reports from patients with anxiety and depressive disorders, Bower (1981) drew instead upon fundamental research and theoretical developments in experimental cognitive psychology. Specifically, he utilized the associative network theory of human memory (Anderson & Bower, 1973), which is essentially a semantic network model in which events are represented in memory as "nodes" that are linked via associated pathways. The central idea is that a node stores information and concepts and that the links between these nodes represent the strength of the association between different concepts. Thus, when two things are frequently encountered or thought about together—horse and cart, say—then when one comes to mind, the other is also likely to come to mind. The hypothesis is that once a specific node or concept is activated, there is a degree of "spreading activation" to other nodes in the network so that more closely associated concepts are more strongly activated, while concepts that are very distant in the network will not be activated at all.

The insight proposed by Bower (1981) was that specific differentiated emotions such as sadness, happiness, or disgust, for instance, are also stored in memory as specific nodes, just like any other semantic concepts, and that each of these nodes is linked to a variety of events and concepts that are associated with that emotion by means of associative pathways (see Fig. 4.3).

Bower's model was developed specifically to account for a body of empirical work that indicated evidence for strong mood-congruency effects in memory. These experiments had shown *state-dependent memory effects* such that memory recall was enhanced when there was a match between a person's mood state when they were *encoding* material and their mood state when they were subsequently *retrieving* that material. To demonstrate this effect, Bower (1981) and his colleagues induced happy and sad mood states (via hypnosis) on two different occasions. They then gave participants a list of unrelated words to remember during each mood state and asked people to recall as many words as they could from one of the lists. The results indicated that memory was typically better when there was a congruency between mood at encoding and mood at recall. Words encoded when in a *sad* mood, for instance, were better recalled when again in a *sad* mood as opposed to when in a *happy* mood state. Bower (1981) went on to demonstrate a number of other effects such that current mood state was an important predictor of the types of autobiographical memories (pleasant or unpleasant) that were likely to be recalled. When a person was in a happy mood state, for instance, they were more likely to bring pleasant memories to mind, whereas when they were in a sad mood state, unpleasant memories were more likely to be remembered.

Based on these and related findings, Bower developed his network model of emotion, which suggested that emotion (or more accurately mood states—see

Fig. 4.3 A fragment of an associative network surrounding an emotion node in memory

Fox, 2008) should influence a range of cognitive processes such as the interpretation of ambiguity and enhancing the salience of mood-congruent stimuli. Specifically, the prediction from Bower's model is that the current mood state should activate a range of associated nodes within a network so that if a sad mood state is present, for example, then ambiguous social situations would be interpreted in a negative way, selective attention would highlight negative relative to positive information, and negative events would be more easily recalled. The prediction of such pervasive effects of mood state on attention, interpretation, and memory is very similar, of course, to those derived from Beck's schema theory albeit based on very different theoretical foundations.

Of particular interest is the fact that Bower's work moved from a purely subjective assessment of biased cognition—by means of questionnaire measures—to a behavioral measure of biased memory. The assumption was that those in a depressed mood state would be more likely to show a greater differential between recall of negative relative to benign or neutral events, thus providing researchers with a behavioral memory-based indictor of mood state, a kind of cognitive marker of mood state. This approach heralded a new line of research in which affective scientists—as they are now called—began to look toward experimental cognitive psychology to identify behavioral tasks that would allow for an objective assessment of biases in attention, interpretation, and memory.

The Emergence of Experimental Psychopathology

Much of the early (and indeed current) clinical research in psychopathology was based on subjective reports of those experiencing conditions such as anxiety and depression. This approach is essential, of course, as psychiatric diagnosis is still based primarily on subjective report and there is no better way of finding out how someone is feeling than asking them in a systematic way via a structured interview or a psychometrically validated questionnaire. However, there are a variety of problems with self-report measures, not least of which is the impossibility of reporting on phenomena of which a person is not aware. The development of information-processing approaches provided researchers with experimental tasks and techniques that allowed for assessment of cognitive processes that study participants or clinical patients might not be aware of. Take the traditional Stroop task (Stroop, 1935) as an example. When people are asked to read out the color that a list of color words are printed in, naming times are much slower when the color of the ink and the word itself are incongruent (e.g., RED printed in blue) compared to when they are congruent (e.g., RED printed in red). While participants are often not aware of this delay—typically called Stroop interference—it can provide a subtle measure of attention allocation. Experimental psychology has developed a large range of tasks allowing researchers to investigate all sorts of processes relating to attention, interpretation, and memory, and in the 1980s groups of clinical scientists began to modify these tasks in order to address cognitive processes that were relevant for clinical phenomena at an implicit level (e.g., Williams, Watts, MacLeod, & Mathews, 1988).

This combination of behavioral tasks with subjective assessment provided a powerful approach and led to a burgeoning of research on the nature of the relations among cognitive and emotional processes, especially as they related to emotional disorders, establishing a new field of experimental psychopathology. In a comprehensive and elegant review of this growing literature, Williams et al. (1988) presented a novel information-processing model of emotional disorders that built upon earlier approaches (e.g., Beck, Emery, & Greenberg, 1985; Bower, 1987) and also moved the field onto a more experimental setting.

In contrast to the pervasive mood-congruent cognitive biases across all aspects of information processing, attention, interpretation, and memory predicted by previous theories (Beck, 1967; Beck et al., 1985; Bower, 1981, 1987), Williams et al. (1988) proposed that the impact of a mood state on a specific cognitive process might actually be dependent on the nature of the mood that was experienced during the cognitive process. For instance, there was evidence indicating that the effects of anxiety influenced selective attentional processes, with less evidence for an impact on selective recall. Depressed mood, in contrast, seemed to be associated more with selective recall of negative material, with little evidence that selective attentional processes prioritized negative material.

To account for these findings that the nature of the bias observed—in attention, interpretation, or memory—might be dependent on the specific mood experienced (e.g., anxious or depressed mood states), Williams et al. (1988) drew heavily on a theoretical distinction that was common within cognitive psychology at the time between *automatic* and *strategic* processing (Schneider & Shiffrin, 1977). The idea was that automatic processes are fast, operate in parallel, are not constrained by capacity limitations, and occur without intent or awareness. In contrast, strategic—or what are often called "controlled"—processes (Hasher & Zacks, 1979) are relatively slow, typically serial in nature, intentional, and are capacity-limited. The insight achieved was that fundamental research in implicit memory experiments might provide a clue as to the nature of information processing biases that occur in emotional disorders or by mood states more generally. Williams et al. (1988) highlighted some studies on memory that had revealed a distinction between what is now widely known as explicit and implicit memory (Craik & Tulving, 1975). Explicit memory occurs when memory requires conscious recollection, whereas implicit memory occurs when performance is facilitated by prior exposure without deliberate or conscious remembering of a study episode. Graf and Mandler (1984), for instance, presented participants with words to be remembered and varied how the words were encoded. On some occasions, the structural features of the word (e.g., How many *vowels* does a word contain?) were emphasized, while on other occasions, the semantic features of the word (e.g., Is this an unpleasant word?) were emphasized. The latter task is assumed to induce a more elaborate form of processing. Memory performance was subsequently assessed by requiring people to explicitly recall as many words as they could remember. In contrast, implicit memory was assessed by a series of word-stem completions (e.g., FOR---, forbid, forget, forest), in which some of the possible solutions (e.g., forest) had been presented previously, while others had not. Explicit recall and recognition were better when elaborative

processing had occurred at encoding, but of more interest was the finding that implicit memory (as assessed by the word-stem task) was unaffected by the depth of processing. In other words, just being exposed to a word led to better performance on a subsequent implicit task, and semantic elaboration at encoding made little difference for performance. In contrast, semantic processing at encoding made a large difference for performance on explicit recall and recognition tasks.

To account for their results, Graf and Mandler (1984) made a distinction between two types of processes that operate on mental representations: *integration* and *elaboration*. Integration refers to the simultaneous activation of many different aspects of a single schema and makes a word or concept become more accessible so that it comes to mind more easily. In contrast, elaboration refers to the activation of several schemas in the presence of a particular mental event and leads to further associations developing among a number of schemata. Williams et al. (1988) suggested that the type of tasks used to assess biased attention in anxiety tapped into integrative processes, whereas the type of tasks used to assess memory bias in depression were dependent upon elaborative processes.

What this means is that the evidence that anxiety was associated with biased attention to a greater extent than memory bias might reflect the fact that anxiety was associated with disruption in integrative processes. Depressed mood on the other hand was considered to be associated with elaborative processes so that mood-congruent material would become more retrievable but not necessarily more detectable in attentional or implicit processing tasks. Subsequently, several empirical findings have been reported, and while some are consistent with the theoretical distinctions made by Williams et al. (1988), many other findings are not (See Mathews & MacLeod, 2005; Williams, Watts, MacLeod, & Mathews, 1997, for reviews). To give just one example, Russo and colleagues found no evidence for anxiety-related biases in implicit memory (Russo, Fox, Bellinger, & Nguyen-Van-Tam, 2001; Russo, Fox, & Bowles, 1999). In contrast, explicit recall bias was found for threat-related material in high trait-anxious participants but only when shallow (i.e., integrative) processing at encoding was encouraged and not when semantic (i.e., elaborative) encoding took place (Russo et al., 2001). Russo et al. (2001) suggested that this anxiety-related mood-congruent recall bias was likely due to enhanced attentional/encoding processing bias toward threat-related material that did influence explicit processes. This hypothesis is supported by a recent meta-analysis, in which the magnitude of an anxiety-related recall bias following shallow processing tasks was associated with attentional biases toward threat (d = 0.71; Herrera, Montorio, Cabrera, & Botella, 2017).

While early experimental psychopathology researchers were, of course, aware of and interested in top-down processes such as attentional control, the field became dominated by studies on automatic processing and how this might lead to fundamental biases in processing information that, in turn, might lead to the development and maintenance of anxiety and depression. A number of theories have directly addressed ways in which goal-directed cognitive mechanisms and automatic processes, which are often stimulus-driven, might operate together to underpin anxiety and depression. Attentional control theory (ACT; Eysenck, Derakshan, Santos,

& Calvo, 2007) posits that there are two attentional systems: a goal-directed system influenced by expectation and current goals and a stimulus-driven attentional system that responds to sudden or salient stimuli in the environment (Corbetta, Kincade, & Shulman, 2002). This theory proposes that anxiety increases the influence of the stimulus-driven system while decreasing the influence of the goal-directed system. This happens because the automatic processing of threat-related stimuli captures and holds the stimulus-driven system (e.g., Fox, Russo, Bowles, & Dutton, 2001) so that the influence of goal-directed processes becomes disrupted. It is further assumed that anxiety reduces attentional control directly, especially in the presence of threat-related stimuli, so that cognitive resources are very likely to be diverted on any task that involves the inhibition or the shifting of attention.

Substantial evidence supports many elements of ACT (e.g., Eysenck & Derakshan, 2011), and this approach highlights the importance of taking multiple cognitive functions (e.g., goal-directed and stimulus-driven) into account when investigating anxiety-related performance on a given cognitive task. For example, ACT suggests that what might look like a pure stimulus-driven effect (e.g., delayed disengagement from a threatening stimulus; Fox et al., 2001) may actually be due to an impairment in attentional control and/or the capacity to shift attention from one task to the other. Mogg and Bradley (2016, 2018) have also proposed that anxiety and cognitive biases—especially attention biases—are both caused by multiple cognitive processes. In setting out a broad framework to examine the nature of cognitive processing in anxiety, they present evidence that threat-related biases in anxiety are influenced by a diverse range of factors that include top-down processes that give priority to goal-relevant stimuli as well as bottom-up processes that prioritize task-irrelevant threat-related stimuli. This model, as with the ACT model, is designed primarily to explain situations in which threat detection itself is not the primary goal. A primary role of biases in attention toward or away from threat, they suggest, is to support the current motivational priorities of the individual. This means that a bias may have many functions including the capacity to rapidly detect and react to threat (orient to threat), to support the elaborative processing of negative stimuli (maintain attention on threat), to minimize subjective discomfort (avoid threat), and to support task performance (suppress biases in attention that distract attention from the main task). The important point here is that unlike previous models that proposed that an automatic bias to orient attention toward threat plays a key role in anxiety (e.g., Williams et al., 1988), the cognitive motivational approach makes no assumption that vigilance for threat necessarily plays a critical role in the maintenance and development of anxiety. Instead, attentional biases are driven by a diverse range of factors that are all in the service of the motivational goals of the individual (Mogg & Bradley, 2016). Figure 4.4 shows a simplified version of the Mogg and Bradley (2018) framework indicating that both attentional biases and anxiety symptoms emerge from a dynamic interplay between bottom-up automatic salience detection processes and top-down goal-directed control processes. While this is an interesting approach that is consistent with much of the literature, it is open to the criticism that it is non-falsifiable.

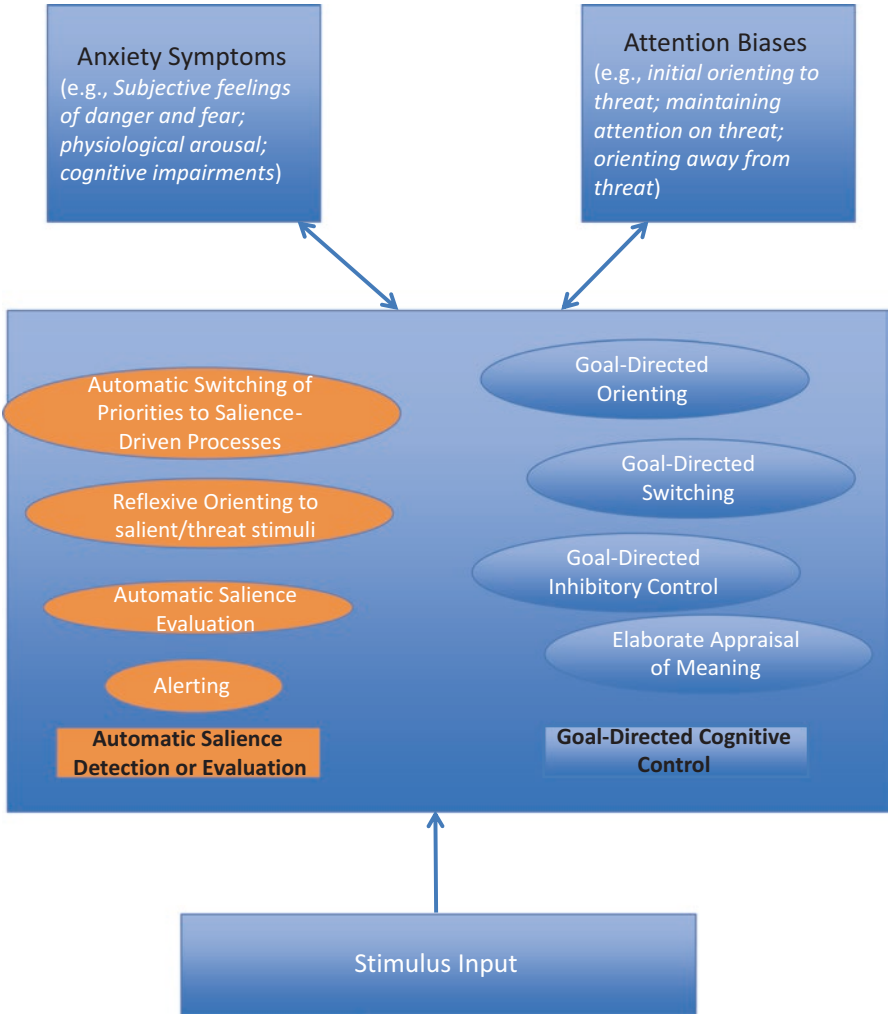

Fig. 4.4 A schematic outline of Mogg and Bradley's (2018) cognitive motivational model of psychopathology

Information-processing models of emotional vulnerability have been very influential, but it is becoming clear that the evidence for an association between threat-related biases in attention and anxiety is more mixed than we might expect (e.g., Van Bockstaele et al., 2014). Different studies show evidence for every pattern of bias: vigilance toward threat, delay in disengaging from threat, or avoidance of threat (Barry, Vervliet, & Hermans, 2015), and a recent meta-analysis found no association at all between negative attention bias (as assessed by the dot probe task) and clinical anxiety (Kruijt, Parsons, & Fox, 2018).

This mixed pattern of association, along with no association at all, is easier to explain within a cognitive motivational framework (Mogg & Bradley, 2016, 2018) than an information-processing framework (Williams et al., 1988). This is because anxiety-related biases in attention are assumed to reflect the current motivational state of the individual that, in turn, is influenced by the details of the current situation along with various individual differences (e.g., trait anxiety). Biases are assumed to be highly dynamic and to vary from threat vigilance to threat avoidance (Mogg & Bradley, 2018). A challenge for cognitive motivational approaches moving forward will be to develop a clear set of hypotheses as to the particular conditions under which specific biases should or should not be observed.

Another challenge for this field is to ensure that the methods used to assess biases in cognitive processing are valid and reliable as inconsistencies in the pattern of observed results may be due, in part, to difficulties and lack of reliability with assessment. The following section provides a brief overview of some common methods used to assess selective processing biases in the domains of attention, interpretation, and memory.

Methods to Assess Emotion-Related Cognitive Biases

Beginning in the early 1980s, a wide range of cognitive behavioral tasks were developed in order to assess selective processing biases so that questions related to the nature of relations among emotional and cognitive processes could be addressed in an objective way. A selective sample of just some of these tasks is presented in the next section to give a flavor of the types of tasks that have been used to assess biased cognition.

Selective Attention

A variety of assessment techniques have been used to assess emotion-related attentional biases, and just a few are mentioned here (see Fox, Derakshan, & Standage, 2011 for a more extensive review). The study of selective attention began with the hearing modality—using dichotic listening experiments, in which people had to attend to the sound stream coming through headphones of one ear while ignoring the content in the other headphone (e.g., Cherry, 1953). Today, the study of selective attention almost exclusively uses the visual modality. Perhaps one of the most common methods to assess selective attention is the color-naming task known as the "Stroop" task mentioned briefly earlier. In the *emotional* Stroop paradigm, words that differ in valence are presented in different colors, and again the task is to indicate the color in which the word is printed. Emotionally specific threatening (e.g., rape, attack), or more general negative/threat (e.g., cancer, failure), positive (e.g., sunshine, happy) or neutral (e.g., chair, water) words are matched for frequency and familiarity, and the time to indicate the response to the *color* of the ink is recorded.

Response times are typically slower for negative relative to positive or neutral words (Pratto & John, 1991), and highly anxious adults and children (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van IJzendoorn, 2007; $k = 70$, $d = 0.49$, 95% CI = [0.43, 0.56]) and just children (Dudeney, Sharpe, & Hunt, 2015; $k = 10$, $d = 0.44$, 95% CI = [0.19, 0.70]) take longer to respond to the color of threat-related relative to affectively neutral words compared to low-anxious participants. This anxiety-related Stroop interference for threat-related words has been widely interpreted as reflecting an anxiety-related selective processing of threat.

The fact that both task-irrelevant (word meaning) and task-relevant (word color) attributes occur together at the focus of attention in the Stroop task means that the spatial distribution of selective attention cannot be assessed. This means that it is impossible to determine whether slower reaction times on negative and threat words are due to a selective capturing of attention away from the primary task or to a more general slowing for negative stimuli (see Fox, 1993). Both mechanisms would, of course, result in slowed reaction times to trials containing threat-related or negative words. Moreover, we can never be completely clear as to whether the slowing on threat-related trials might be due to early perceptual/attentional processes or to response biases that occur much later in information processing (Fox, 1993; Williams et al., 1988, 1997).

Building upon earlier work with the dichotic listening task as well as studies showing that spatially separate visual items can produce Stroop-like interference effects (see Eriksen & Eriksen, 1974); MacLeod, Mathews, and Tata (1986) designed an attentional-probe task in an attempt to find a direct way to measure how attention is distributed in a visual scene. In the original task, a pair of words was presented on a computer screen—located about 3 cm apart, one above the other—for 500 ms, and participants had to read aloud the word in the upper location. Occasionally a small dot (the probe) appeared in either the upper or the lower location, and when this happened, participants had to press a handheld button to record reaction time. Attention bias is calculated by subtracting mean response times when probes appear in the location previously occupied by a threat-related word (e.g., cancer) from mean reaction times on trials in which the probe replaces a neutral word (e.g., corner). To illustrate, a numerically positive attention bias index (e.g., +20 ms) indicates a bias *toward* threat-related words, while a negative numerical index (e.g., −20 ms) indicates a bias *away* from threat-relevant material. MacLeod et al. (1986) found that reaction times for patients diagnosed with generalized anxiety disorder (GAD) were indeed faster when the probe appeared in a location recently occupied by a threat-related word (e.g., rape) relative to when the location had been occupied by a neutral word, which was interpreted as reflecting a bias in attention toward threat. This speeding did not occur in a control group of non-anxious participants or in a group of patients diagnosed with major depressive disorder (MDD). The authors concluded that that anxious, but not depressed, participants do indeed shift their attention toward emotionally threatening stimuli in their visual environment.

The attentional-probe task was considered to have several advantages over other measures of selective processing in that it required a neutral response (button

pressing) to a neutral stimulus (dot probe) and therefore could not be due to a response bias toward emotionally valenced stimuli. Moreover, because the outcome of interest is a speeding of reaction time, the results cannot be due to a general slowing down of response in the presence of threatening material as is often observed in anxious populations. Hundreds of studies have subsequently been conducted using many variations of the original attentional-probe paradigm (see Mathews & MacLeod, 2005, for a review). Contemporary studies typically require participants to keep their eyes focused in the middle of a word pair, rather than reading aloud—i.e., attending to—the upper word and consist of the simultaneous presentation of pairs of words, pictures, or faces, followed by a probe to be categorized (e.g.,: or ..), rather than simply identified.

A large literature (with a variety of techniques of presenting the attentional-probe task) has shown mixed results with early evidence for anxiety-related threat biases using the attentional-probe task in a meta-analysis combining adult and child populations (Bar-Haim et al., 2007; $k = 35$, $d = 0.37$, 95% CI = [0.28, 0.46]). However, more recent meta-analyses considering only studies with children (Dudeney et al., 2015; $K = 28$, $d = 0.18$, 95% CI = [−0.006, 0.37]) as well as baseline attentional bias in clinically anxious adult patients who were enrolled in trials to assess the impact of an attentional bias modification procedure found no evidence of any threat-related bias at all (Kruijt et al., 2018; $k = 13$, $d = 0.05$, 95% CI = [−0.6, 4.3]).

These mixed results raise serious questions about the stability of anxiety-related biases in attention as measured by the attentional-probe task even in clinical populations. This may indicate that contrary to cognitive theory, the evidence for anxiety-related biases in attention is actually more mixed than is often acknowledged. Some theoretical and methodological factors may, of course, explain the inconsistency of results with the attentional-probe task. One possibility is that different mechanisms may be at play in the attentional-probe, which is effectively a static method attempting to capture a snapshot of what is, of course, an inherently dynamic process at a specific point in time—typically across a 500-ms period (Mogg & Bradley, 2016; Zvielli, Bernstein, & Koster, 2014). For instance, in a sample of people diagnosed with spider phobia, attention bias toward spider-related photographs (highly threatening images for this population) was found with presentation times of 200 ms with no evidence of bias at longer exposure times of 500 ms or 2000 ms (Mogg & Bradley, 2006). This pattern of results suggests that an early vigilance for threat in spider phobia may not be maintained over time. It is not clear why this might be, but it may reflect a regulatory mechanism designed to reduce anxiety by avoiding the source of threat.

It is also possible that patterns of attentional allocation toward threat and orientation of attention away from threat effectively cancel out each other out in attentional-probe studies (Fox, Zougkou, Ashwin, & Cahill, 2015; Mogg & Bradley, 2016). For instance, in a sample of 127 people self-reporting high levels of spider fear, it was found that 44 demonstrated a bias toward spider-related images (defined as a reaction time difference of 25 ms or more), 36 showed a bias away from threat, and the remaining 47 showed no bias in either direction (Fox et al., 2015). Overall, this sample showed no bias on the attentional-probe, and yet the impact of an attentional

bias modification procedure in terms of reducing spider fear was strongly influenced by the nature of the initial bias at baseline. Only those with an initial bias toward threat showed any benefit from the intervention. Thus, pre-existing levels of selective processing biases may be an important factor to consider in assessing the impact of intervention studies designed to reduce biases and hence improve clinical symptoms.

This short overview highlights some of the complexity in interpreting results from the attentional-probe task and is consistent with a growing concern about the statistical reliability of the attentional-probe task, at least when reaction times are used as the dependent measure. When reliability (e.g., split-half reliability) has been assessed—which is rare for most behavioral tasks—it has typically been very low ranging from −0.12 to 0.68 (Parsons, Kruijt, & Fox, 2018; Price et al., 2015; Staugaard, 2009). It is not clear whether these results can be attributed to a genuine absence of biased attention in anxiety or whether the attentional-probe task itself is simply not a reliable measure of biased attention. Reliability measures for the attentional-probe task have been found to be somewhat better when eye-tracking indices are used as the dependent measure (c. 0.32; Price et al., 2015). Likewise, when ERP measures are used as the dependent measure, reliability estimates ranging from 0.52 to 0.79 have been reported in attentional-probe studies (Kappenman, Farrens, Luck, & Proudfit, 2014; Moran, Jendrusina, & Moser, 2013) and the N2pc component has been shown to have a greater degree of split-half reliability when directly compared with reaction time measures (Reutter, Hewig, Wieser, & Osinsky, 2017).

On theoretical grounds, one of us has argued previously that the attentional-probe task is unable to differentiate between two potential mechanisms that might lead to biased attention: an enhanced tendency to orient toward threat on the one hand or a delay in disengaging attention from threat once it is noticed on the other (Fox et al., 2001). This limitation is due to the fact that the allocation of attention is typically measured at one point in time (e.g., 500 ms) in the attentional-probe task. This means that participants may shift their attention back and forward between the two images so that any resulting bias may be due to the *holding* of attention by threat-related stimuli rather than an initial allocation of attention to those stimuli. To address this hypothesis, Fox et al. (2001) modified the spatial cueing task that was initially developed by Posner, Snyder, and Davidson (1980) and conducted several studies with participants who varied in terms of their self-reported anxiety and depression. The original cueing task presented three boxes on a computer screen, one centered at fixation and two located on both sides of the central box (left and right). Either of the peripheral boxes could be cued by a brief flickering, which was followed by the presentation of a target that had to be detected in either the valid (cued) or invalid (un-cued) box. Detection of targets is typically faster in the cued relative to the un-cued locations, which is taken to reflect the automatic reflexive orientation of attention to the cue. When the target appears in the un-cued box, a disengagement of attention is required from the cued location. Attention then has to shift to the un-cued location in order to process the target, and this explains the slower detection latencies on invalid trials.

Fox et al. (2001) modified this task to make an emotional version by replacing the light flicker cue with a threatening, positive, or neutral word or a photograph of happy, neutral, or angry facial expressions as cues. They reasoned that detection latencies on *un-cued* trials would provide an indication of differences in the speed of attentional disengagement from threatening, positive, or neutral stimuli. Across several experiments, it was found that high-anxiety participants were indeed slower on un-cued trials when the cue was threat related (e.g., a threat-related word or image of an angry facial expression), and this was interpreted as reflecting an anxiety-related delay in disengaging from threat. A similar pattern of results was subsequently found with a wide range of stimuli (e.g., affective pictures, faces, words) in other studies for participants varying in both anxiety and depression (e.g., Derryberry & Reed, 2002; Koster, Crombez, Verschuere, & De Houwer, 2004). These findings support the hypothesis that attentional biases toward negative, especially threat-related, stimuli may often reflect a difficulty in disengaging attention from negative material rather than—or in addition to—enhanced attentional orientation toward threat (Fox et al., 2001; see Mogg, Holmes, Garner, & Bradley, 2008, for an alternative interpretation).

Selective Interpretation

Those who are prone to anxiety and depression are especially likely to interpret ambiguity in a consistently negative way, and this is true for both adults (Hertel & Mathews, 2011; Mathews & MacLeod, 2005) and children (Lau & Waters, 2017). As with measures of selectivity in attention, a large number of techniques have been used to assess biases in interpretation, and these have been extensively reviewed by Schoth and Liossi (2017). A selection of some of the more common measures is presented here.

The homophone task presents spoken homophones that have both threatening and neutral meanings (e.g., pane/pain; wore/war; dye/die), and participants simply have to write down what they hear. Those prone to anxiety (Mathews & MacLeod, 2005) and depression (Mogg, Bradbury, & Bradley, 2006) are more likely to write down the negative spellings relative to those with lower levels of anxiety or depression. Similarly, homographs (e.g., beat, growth, arms) can be presented as a prime word prior to either a nonword or a real word that is related to either the threatening or neutral meaning of the homograph. Richards and French (1992), for instance, found that anxious individuals were faster to categorize a letter string as a "word" in a lexical decision task when it was related to the more negative meaning of a prior homograph (e.g., *weapon* was categorized faster than *legs* following the prime *arms*). Several studies have shown biased interpretation related to elevated anxiety, depression, and chronic pain using this task (see Schoth & Liossi, 2017).

In the scrambled sentences task, participants are presented with a mixed sequence of words (scrambled sentence) such as *the*, *know, is, future, going, bleak, to, bright, be, I* that they are asked to resolve as quickly as possible (Wenzlaff, 1993). These sentences can be resolved in either a positive "I know the future is going to be

bright" or a negative "I know the future is going to be bleak" way, providing an indication of bias. This task is straightforward, although can be difficult for participants, and the task is not particularly easy to administer or score. Some researchers have used sentences or short scenarios, in which reading times of sentences *following* an ambiguous sentence are taken as an index of the degree to which the initial sentence was interpreted in either a negative or a positive direction. For instance, MacLeod and Cohen (1993) presented students with sentences such as "The two men completed the service and filled in the hole." These ambiguous sentences were followed by others that were either consistent with a threatening ("The funeral was soon finished") or a benign ("The repair was soon finished") meaning. Using this task, it has been found that anxious individuals are faster to read follow-on sentences that are consistent with the more threatening interpretation of the ambiguous sentences supporting the argument that anxiety is characterized by mood-congruent interpretations when reading a sentence (Calvo, Eysenck, & Estevez, 1994).

Another paradigm developed by Eysenck, Mogg, May, Richards, and Mathews (1991) requires participants to listen to sentences that could be interpreted in either a neutral or a threatening way that are later followed by alternative versions that resolve the interpretation in one way or another. For example, a sentence such as "The doctor examined little Emily's growth" could be later followed by sentences in a recognition memory task such as "The doctor measured little Emily's cancer/height." Later versions of this task have replaced the recognition test with a simpler rating scale (e.g., Holmes & Mathews, 2005), and participants are sometimes encouraged to use mental imagery to simulate the scenarios (e.g., Holmes, Lang, & Shah, 2009). An ambiguous scenarios task has been developed specifically to assess interpretation bias for depressed mood (Berna, Lang, Goodwin, & Holmes, 2011). Using similar tasks with adolescents, it has been found that those with anxiety and depression are more likely to endorse threatening/negative interpretations more than benign/neutral interpretations relative to non-anxious/depressed control participants (e.g., Haller, Raeder, Scerif, Kadosh, & Lau, 2016). Given the mixed findings for anxiety and attention biases as discussed previously, it is possible that these biases may be driven primarily by depression.

The Adolescent Interpretation and Belief Questionnaire (AIBQ) is a self-assessment instrument that is useful for studies with adolescent populations and provides indices of biased interpretation for both social and nonsocial situations. Developed by Miers, Blote, Bogels, and Westenberg (2008), the AIBQ requires participants to read ten ambiguous scenarios that they should imagine are happening to them. For example, "you've invited a group of classmates to your birthday party, but a few have not yet said that they're coming." Participants are then provided with a list of three thoughts that might typically arise in response to the situation and are asked to indicate how likely it is that this thought would pop into their head on a five-point Likert scale. In the above example, the thoughts are "they don't know yet if they can come or not" (neutral interpretation), "they don't want to come because they don't like me" (negative interpretation), and "they're definitely coming; they don't need to tell me that" (positive interpretation). Some of the examples are nonsocial (e.g., your new watch does not work). Therefore, there are four outcome

measures: positive interpretation (social), positive interpretation (nonsocial), negative interpretation (social), and negative interpretation (nonsocial).

Selective Memory

Autobiographical memory refers to memory for personal events that shape our emotional lives. Decades of research have shown that depression is associated with dysfunction in how we recall these personal memories (Kohler et al., 2015). The most widely used assessment of autobiographical memory in depression is the Autobiographical Memory Test (AMT; Williams & Broadbent, 1986), which asks study participants to recollect a specific memory in response to a presented word cue within a specified time period (e.g., when you were between the ages of 25 and 30 years). The reported memories are then classified according to a range of factors including their valence, content, and specificity. Two clear findings have emerged in the literature. First, higher levels of self-reported and diagnosed depression are associated with a bias toward favoring negative experiences relative to positive or benign experiences. Second, depression has been associated with what have been called "over-general" memories. For instance, in response to the prompt, "recall a time that you were happy" a nondepressed person is likely to give a highly specific example such as the day they got married, whereas a depressed person is more likely to say something much more general such as "summer holidays." These over-general memories have been shown to be relatively good predictors of higher depression symptoms over time (Sumner, Griffith, & Mineka, 2010). A recent study, for instance, showed that higher specificity in recall of positive—but not negative—autobiographical events lowered vulnerability to depression in adolescents over a 1-year timeframe (Askelund, Schweizer, Goodyer, & van Harmelen, 2019).

A problem with autobiographical memory research is that we have, of course, no way of knowing the incidence of positive and negative life events a person has experienced, and therefore it is difficult to determine whether results represent a genuine bias or not. For this reason, cognitive and emotion researchers often use consistent lists of words or pictures so that they can be sure that if more negative than positive words are remembered, this represents a genuine bias. Several recall and recognition tasks have been used to assess anxiety- and depression-related explicit memory bias. The typical procedure is to present a study participant with a list of words that vary in valence (negative, positive, neutral) but that are carefully matched across valence categories for word frequency, familiarity, pronounceability, and other factors that may affect recall (see Rubin & Friendly, 1986, for a comprehensive list). It is important to include some neutral words at the start and at the end of the list to control for primacy and recency effects (Russo et al., 2001). How the words are encoded can make a difference to subsequent recall—studies have asked people to read aloud the initial words, count the number of syllables, or make some type of semantic judgment. Participants are usually not told that they will be asked to remember the words later, and then, typically following a distractor task, participants are asked to recall as many of the words that they can. Rather than free recall,

participants can be presented with mixed lists of the words they have seen at encoding along with an equal number of matched words that they have not seen, and they are then asked to recognize those words that were presented before. Memory is typically much better under these conditions, and it is often more difficult to find evidence for bias under these conditions (Russo et al., 2001).

The self-referential encoding task (SRET; Hammen & Zupan, 1984) was designed to reflect an individual's underlying negative cognitive schema (Beck, 1967) and asks participants to judge whether a series of adjectives that can have a negative, positive, or neutral valence describes them or not. This encoding task is followed by an incidental free-recall task. Findings from both adult and child samples demonstrate that depression is associated with a greater degree of self-referential encoding biases in terms of the endorsement of more negative and fewer positive adjectives as being self-descriptive as well as the recall of more negative relative to positive adjectives, regardless of whether they were endorsed or not (Auerbach, Stanton, Proudfit, & Pizzagalli, 2015; Phillips, Hine, & Thorsteinsson, 2010).

Measures of implicit memory are designed to assess the influence of a past experience on the performance of a cognitive task that is seemingly unrelated to the previous experience. For example, a list of words is given at encoding, and then instead of being asked to recall those words, following a filler task, participants are instead presented with a list of word stems (e.g., c a _ _ _ _) and asked to complete them with the first word that comes to mind. In general, word stems are more likely to be completed by words that were presented at encoding (e.g., cancel) than by words not presented earlier (e.g., cancer). Early studies reported evidence for a threat-related implicit memory bias in anxiety but not in depression (e.g., Williams et al., 1997), while subsequent research has typically failed to find consistent evidence for a larger implicit memory bias for threat-related words in anxious individuals (e.g., Russo et al., 1999).

Genetic Approaches

Genetic Approaches to Emotional Disorders and Well-being

Quantitative Genetics

Alongside the "cognitive revolution," the fledgling field of quantitative genetics was beginning to investigate genetic influences on emotional disorders. By the 1960s it was well-established that almost all mental illnesses, including emotional disorders, ran in families. Acknowledging that this familial resemblance may reflect either shared genes or shared environmental influences, quantitative genetic studies subsequently set about disentangling genetic from environmental effects using samples of twins.

Twin studies compare within-pair concordance or within-pair correlations in identical (MZ) twins (who share 100% of their DNA) and nonidentical (DZ) twins

(who share, on average, 50%). A greater within-pair correlation in MZ twins, relative to DZ twins, indicates a genetic influence, while a within-pair correlation that is similar for MZ and DZ twins suggests a role of the shared environment. Twin studies also allow us to estimate the role of the non-shared environment (environmental factors that are unique to each member of the pair) which is indexed by the degree of discordance within MZ twin pairs. The results of several decades of twin studies suggest that genetic factors play a significant role in the etiology of both clinical and subclinical presentations of emotional disorders as well as positive outcomes, such as subjective well-being. While estimates vary between samples, between 30 and 40% of the variance in depression, anxiety, and subjective well-being is explained by genes, with the remaining variance explained by non-shared, rather than shared, environmental influences (Bartels, 2015; Hettema, Neale, & Kendler, 2001; Sullivan, Neale, & Kendler, 2000).

In bivariate extensions to the twin model, the concordance within twin pairs is compared within and across different traits. This allows researchers to estimate the extent to which genetic and environmental influences overlap for different outcomes. For instance, using these bivariate models, twin studies have shown a substantial genetic overlap between a wide variety of behavioral traits and psychiatric disorders. This genetic overlap is particularly striking for anxiety and depression where the genetic correlation is between 0.84 and 1 (Kendler, 1996). Interestingly, these findings also extend to positive outcomes with the genetic correlation between subjective well-being and emotional disorders ranging from −0.64 to −0.76 (Bartels, 2015; Kendler, Myers, & Keyes, 2011). While the genetic overlap between disorders and traits is a consistent finding across quantitative genetic studies, it appears that environmental influences are outcome-specific. This means that while generalist genes explain the shared risk of emotional disorders and well-being, trait-specific environmental factors explain how this genetic risk manifests itself (Kendler, 1996; Kendler & Eaves, 1986; Kendler, Heath, Martin, & Eaves, 1987). For instance, Eley and Stevenson (2000) showed that while loss events explained the development of depression, threat events were associated with symptoms of anxiety.

Molecular Genetic Studies

Quantitative genetic studies allow us to estimate the relative role of genetic and shared and non-shared environmental influences on a trait and can also be used to explore the extent to which these influences are shared between different outcomes. However, conventional epidemiological studies are required to identify the specific genetic or environmental factors that underlie this risk. An extensive literature has implicated multiple environmental factors including parenting styles, stressful life events (Brown & Harris, 1978), and childhood maltreatment (Nanni, Uher, & Danese, 2012) that likely explain the non-shared environment component of emotional disorders and well-being. Identifying the specific genetic variants that explain the genetic component remains a significant challenge.

Early molecular genetic studies of emotional disorders relied on what was called a candidate gene approach. In this approach, a gene is selected a priori based on its proposed involvement in the biology of a disease or trait. Variants within that gene are then tested for association with the outcome of interest. These approaches saw some early success in genes implicated in neurotransmission, neurogenesis, cell signaling, and the hypothalamic-pituitary-adrenal (HPA) axis (Lopez-Leon et al., 2008). Nevertheless, the majority of these findings have failed to replicate, suggesting that the search for genetic variants associated with emotional disorders should be extended beyond this small list of candidate genes.

Unlike the candidate gene approach, more recent genome-wide association studies (GWAS) are hypothesis-free and include up to a million variants across the genome. Although initially hampered by small sample sizes, GWAS of emotional disorders and well-being have begun to make progress identifying a handful of genetic variants that survive correction for multiple testing. Individually, the effects of these variants are vanishingly small (explaining less than 0.1% of the variance) and explain very little of the heritability observed in quantitative genetic studies (Okbay et al., 2016; Wray et al., 2018).

However, new techniques that simultaneously consider the aggregate effects of all genotyped variants are now beginning to close the gap between quantitative genetic estimates of heritability and molecular genetic estimates. One of these approaches—polygenic scoring—allows the effects of multiple variants to be summarized in a single score. Specifically, alleles associated with a trait in a discovery sample at a given p value threshold are selected in a target sample, and a score (the sum of these alleles weighted by their effect size) is then created for each individual. The effects of this score on the phenotype are determined using linear or logistic regression which includes an estimate of the variance explained (Wray et al., 2014). Findings from polygenic scoring studies and more recent polygenic approaches suggest that, on aggregate, common genetic variants do explain a substantial proportion of the heritability observed in twin models (Okbay et al., 2016; Wray et al., 2018). These approaches also confirm the results from bivariate twin studies. That is, the genetic overlap between depression, anxiety, and subjective well-being is substantial with genetic correlations of between 0.3 and 0.9 (Okbay et al., 2016).

Genes Environment Interaction in Emotional Disorders

Twin and adoption studies have established that both genes and the environment contribute to psychopathology and well-being. However, these factors do not operate in isolation. Rather, complex traits such as emotional disorders are likely the result of interplay between genetic and environmental influences. In one form of this interplay (gene-environment interaction, GxE), genes are proposed to affect an individual's sensitivity to environmental factors. Several twin studies have

provided evidence that the effects of the environment do indeed vary by genetic background. For example, in a large, longitudinal sample of female twins, Kendler et al. (1995) showed that for individuals with a low genetic risk of major depression, the occurrence of severe stressful life events increased the probability of a depressive episode from 0.5 to 6.3%. However, for those with the highest genetic risk, the same severe events increased the probability from 1.1 to 14.6%. These findings have been confirmed in both depression and anxiety by more sophisticated models which test the degree to which genetic effects are moderated by a measured environmental factor (Eaves, Silberg, & Erkanli, 2003; Lau & Eley, 2008b; Silberg, Rutter, Neale, & Eaves, 2001). In each of these studies, genetic effects appear to increase in adverse environments, indicating environmental control of genetic effects (Silberg et al., 2001). Quantitative genetic studies have shown that these findings also extend beyond severe stressors to daily negative and positive events. For example, using an experience sampling approach in a sample of twins, genetic factors have been shown to influence stress sensitivity (the relationship between stressful daily events and subsequent negative effect; Jacobs et al., 2006). A later study of the same sample suggested that, in line with previous studies, these genetic factors overlapped with those for major depression (Wichers et al., 2007). Specifically, the effects of negative daily events were significantly greater in those individuals with the highest genetic risk of depression, even in the absence of current depression symptoms or a personal history of the disorder. These studies have some similarity with cognitive high-risk studies in which those with a high degree of cognitive vulnerability have been shown to be more reactive to negative life events (Abramson et al., 1989).

Genetic studies using candidate gene approaches have also come to similar conclusions. In a seminal study in 2003, for instance, it was found that the 5-HTTLPR (a putatively functional genetic variant in the gene encoding the serotonin transporter) moderated the effects of adversity on the development of major depression and symptoms of depression (Caspi et al., 2003). Specifically, individuals with one or more short (S) alleles at this locus were more sensitive to the depressogenic effects of stressful life events, or childhood maltreatment, than those homozygous for the alternative long (L) allele. Similar findings have been reported for this locus across multiple studies of emotional disorders including anxiety (Kendler, Gardner, & Lichtenstein, 2008) and the effects of less severe daily negative stressors on negative mood (Gunthert et al., 2007). Other candidate GxE studies in both depression and anxiety have also found evidence for significant moderating effects of some candidate genes including brain-derived neurotrophic factor (BDNF; Bukh et al., 2009; Chen, Li, & McGue, 2012; Hosang, Shiles, Tansey, McGuffin, & Uher, 2014), dopamine receptor D2 (DRD2; Elovainio et al., 2007), corticotrophin-releasing hormone receptor 1 (CRHR1; Liu et al., 2013; Polanczyk et al., 2009), catechol-O-methyltransferase (COMT; Mandelli et al., 2007), and FK506 binding protein 5 (FKBP5; Zimmermann et al., 2011).

Diathesis Stress Models

The differential effects of the environment by genetic factors were originally conceptualized in diathesis-stress models, where genotypes were considered to predispose individuals to the negative effects of adversity. The study by Caspi et al. (2003) discussed above provides a good example of diathesis-stress in which individuals carrying specific risk alleles (e.g., the short allele of the 5-HTTLPR) were found to be more likely to develop depressive symptoms when exposed to environmental adversity. In other words, a heightened genetic risk is considered to confer a higher risk of developing psychopathology when adverse conditions are experienced. The hypothesized relationship between genetic and environmental risk is illustrated in Fig. 4.5a.

Differential Susceptibility Models

An alternative model for gene-environment interaction based on evolutionary genetic theory is the differential susceptibly hypothesis. In this model, genetic variants are not considered risk factors but instead act as "susceptibility" or "sensitivity" factors that moderate the effects of both negative *and* positive environments on outcome: for better *and* for worse (Belsky, Bakermans-Kranenburg, & van Ijzendoorn, 2007). In other words, the same genetic factors that may render individuals more vulnerable to adversity may also make them more likely to benefit from optimal and supportive environments (Belsky et al., 2007). The hypothesized relationship between genetic and environmental factors as envisaged by the differential susceptibility hypothesis is illustrated in Fig. 4.5b.

Consistent with differential susceptibility, studies suggest that individuals with one or more copies of the S allele of the 5-HTTLPR are not only at a greater risk of mood disorders following adversity but also benefit more from the protective effects of *positive* environmental influences such as maternal warmth (Sulik et al., 2012), social support, and interventions, including family support and the efficacy of CBT in children with anxiety. In addition to the 5-HTTLPR, differential susceptibility has been demonstrated for a number of genetic variants previously implicated in GxE (Belsky et al., 2009). These findings have paved the way for "experimental" GxE studies in which the environmental component (usually an intervention) is randomly assigned to individuals. This approach vastly increases the power of GxE studies by ensuring that, unlike observational studies, environmental conditions are standardized across participants and that equal numbers of individuals are exposed to the environment of interest. Moreover, as participants do not play a role in the selection of their environments, this approach also tackles any confounding effects of gene-environment correlation. The increased power of the experimental GxE approach was supported by a recent meta-analysis (van Ijzendoorn & Bakermans-Kranenburg, 2015). A more recent extension of the differential susceptibility

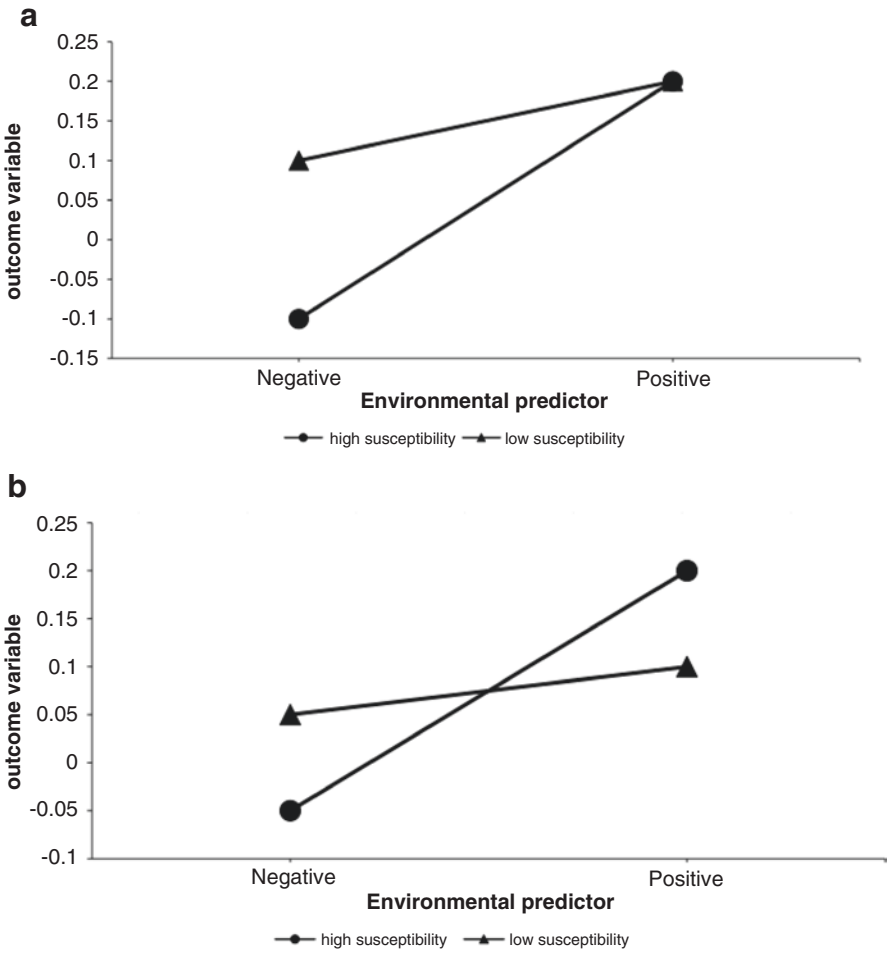

Fig. 4.5 A schematic outline of the (**a**) diathesis stress model and (**b**) differential susceptibility model of gene by environment interactions in the development of psychological well-being

hypothesis includes the concept of vantage sensitivity, where genotypes work to increase responsivity to positive influences (Pluess & Belsky, 2013). Much of the evidence for vantage sensitivity comes from genotypes implicated in differential susceptibility. However, there is some evidence to suggest that for specific genotypes, in specific contexts, effects on sensitivity to positivity are larger than the effects of sensitivity to adversity (Pluess & Belsky, 2013). Indeed, some studies have shown that genetic effects act exclusively on sensitivity to positive rather than negative influences (Pluess & Belsky, 2013).

Polygenic Gene-Environment Interaction

While providing several promising findings and new avenues for research, candidate-gene GxE findings, just as studies that look for main effects of candidate genes, often fail to replicate, even in high-quality studies with very similar methodologies (e.g. Fergusson, Horwood, Miller, & Kennedy, 2011). This has led to several meta-analyses that do not provide support for the initial findings (Munafo, Durrant, Lewis, & Flint, 2009; Risch et al., 2009). Multiple reasons for non-replication have been suggested including differences in the measurement of outcome and environments, the statistical approach used, and the developmental sensitivity of effects (Uher & McGuffin, 2010). It has also been suggested that these findings are simply the result of type 1 error and publication bias (Duncan & Keller, 2011).

The mixed findings within GxE research indicate that there may be some benefit in moving from a candidate-gene approach to a genome-wide, polygenic approach. In a polygenic approach, GxE studies use previous GWAS findings to quantify genetic risk of a disorder in target sample and then test the effects of this genetic risk at different levels of the environment. For example, Peyrot et al. (2014) developed a polygenic risk score based on a GWAS of depression and tested whether these effects differed in those who also experienced childhood maltreatment. In this study, childhood maltreatment had significantly greater effect in those with high vs low polygenic risk for major depression. Nevertheless, a later study reported contradictory findings, in which the effects of maltreatment were greatest in those with the lowest genetic risk (Mullins et al., 2016). A more recent meta-analysis combined the above findings with results from several further cohorts. This study provided little support for Peyrot et al.'s (2014) initial findings and showed no evidence for an interaction between a negative environment and polygenic risk on the development of major depressive disorder (Peyrot et al., 2018).

While the cause of these discrepant findings remains unclear, a recent study suggests that they may reflect inadequate measurement of environmental effects. Using a large health and retirement sample, Domingue, Liu, Okbay, and Belsky (2017) investigated polygenic GxE by assessing within-person changes in depression symptoms following the death of a spouse. The authors reported that while depression symptoms increased for all individuals experiencing bereavement, these effects were significantly greater for those with a high vs low polygenic score for major depression. Moreover, a polygenic score for subjective well-being, generated from a GWAS study of nearly 300,000 participants (Okbay et al., 2016), appeared to protect against the depressogenic effects of bereavement. This study suggests that the success of polygenic GxE studies may rely on more careful within-person studies of gene-environment interaction using externally validated environmental measures. This approach is not only more robust to reporter biases but also limits any confounding effects of gene-environment correlation, where exposure to a given environment may be influenced by an individual's behavior.

By using a polygenic score using genetic variants with a main effect on depression, the above GxE studies implicitly adopt a diathesis-stress approach to emotional disorders. That is, they test whether the effects of environmental adversity are

exacerbated by genetic risk. However, this approach is at odds with the differential susceptibility hypothesis, where genetic variants that increase sensitivity are proposed to have no main effect on psychopathology. Variants involved in differential susceptibility can therefore only be captured using a polygenic score of environmental sensitivity, rather than risk of a disorder. We recently presented a novel way to develop such a score using monozygotic twins. Since monozygotic twins are genetically identical, within-pair differences in an outcome is assumed to be the result of non-shared environmental factors. This means that genotypes that render MZ twins more sensitive to environmental influences (for better and for worse) should be associated with greater within-pair differences. Using this approach, Keers et al. (2016) conducted a GWAS of within-pair differences in emotional symptoms and used the results to construct a polygenic environmental sensitivity score. In a separate sample, this score moderated the effects of positive as well as negative parenting on psychopathology in a manner that was consistent with differential susceptibility. Specifically, individuals with a higher polygenic environmental sensitivity score were more sensitive to the negative effects of negative parenting but also benefited more from the protective effects of positive parenting. The polygenic score of sensitivity also predicted responses to a psychological intervention in separate sample of children with anxiety disorders. In these analyses, children with a high polygenic score of environmental sensitivity responded significantly better to high-intensity individual CBT than a less intensive course of guided self-help. In contrast, children with a low polygenic score responded equally well to both forms of treatment (Keers et al., 2016). In agreement with prior theoretical considerations (Belsky et al., 2007), the polygenic environmental sensitivity score did not predict psychopathology directly in either sample, suggesting that genetic determinants of environmental sensitivity may be distinct from those that directly increase the risk of a disorder. The increasing strength of GxE which occurred with the inclusion of more genetic variants in the polygenic environmental sensitivity score also suggests that genetic sensitivity to the environment is highly polygenic and distributed over a large number of genetic variants (Keers et al., 2016). This early finding suggests that detection of GxE will require genomic and polygenic methods that explicitly test sensitivity to environment and most of them will not be detected with either candidate gene GxE approaches or as a secondary investigation of genetic variants directly associated with a disorder.

In summary, quantitative and molecular genetic approaches suggest that emotional disorders and psychological well-being are complex traits caused by environmental factors and multiple genetic variants that each has a small effect. While genetic variants may have direct effects on outcomes, they also work by modifying response to the environment. Theoretical models of GxE including diathesis stress, differential susceptibility, and vantage sensitivity suggest that genetic effects on emotional disorders and psychological well-being may include genotypes that influence sensitivity to adversity (i.e., diathesis stress), sensitivity to protective factors such as social support (i.e., vantage resistance), or plasticity more generally (i.e., differential susceptibility).

Integrating Cognitive and Genetic Models of Emotional Disorders

Genes and Cognitive Vulnerabilities

Both Beck (1967) and Abramson et al. (1989) suggested that the "cognitive vulnerabilities" central to their models of their depression are the result of early childhood adversity. Multiple empirical studies support this hypothesis. Physically harsh parenting, for example, has been associated with schemas of guilt and shame in adolescents (Stuewig & McCloskey, 2005), while negative parenting style was found to predict adolescents' self-worth and attributional style 3 years later in a large prospective study (Garber & Flynn, 2001). These findings also extend into adulthood (Gibb et al., 2001). However, a growing literature suggests that in addition to environmental adversity, genetic factors are also likely to play a role in the development of cognitive vulnerabilities. In one of the first studies of its kind, Schulman, Keith, and Seligman (1993) reported a substantial role of genetic factors in individual differences in attributional style in adults. Similar, albeit more modest, findings have also been reported in studies of adolescents where genes explained 35% of the variance in attributional style in 15-year-olds (Lau, Rijsdijk, & Eley, 2006) and 44% of the variance at 2-year follow-up of the same sample (Lau & Eley, 2008a). In support of this, studies that use more reliable measures of attention bias based on event-related potentials (ERPs) rather than reaction times report an even larger role of genes. For example, genetic influences have been shown to explain up to 55% of the variance in P300 in adolescents (Anokhin, Golosheykin, Grant, & Heath, 2010) and adults (Weinberg, Venables, Proudfit, & Patrick, 2015).

Interestingly, findings for child samples are considerably more mixed. Moderate genetic contributions have been found for labeling various threatening facial expressions, including fear, sadness, and disgust (Trouton, Spinath, & Plomin, 2002), while responses to ambiguous words and ambiguous scenarios task also showed moderate heritabilities in 8- and 10-year-old children (Eley et al., 2008). However, little to no genetic influence on attributional style (Lau, Belli, Gregory, Napolitano, & Eley, 2012) or on attentional bias measured using the attentional-probe task (Brown et al., 2013) was found in this age group. There was also little evidence for a role of genes in explaining individual differences in interpersonal cognitions in 8-year-old children (Gregory et al., 2007), while genetic effects did emerge at a later follow-up of the same sample at age 10 (Lau, Belli, Gregory, & Eley, 2014). This increasing role of genetic factors in cognitive biases across development mirrors findings from other complex traits, including emotional disorders, where genetic effects appear to increase with age (Bergen, Gardner, & Kendler, 2007). In other words, genes may have little influence in childhood but increasing effects in adolescence when new genetic influences come online (Kendler et al., 2008; Scourfield et al., 2003). Indeed, it has been suggested that the maturation of cognitive biases during adolescence (and the subsequent effect of these biases on emotional

symptoms) may explain the increases in heritability observed for emotional disorders (Lau et al., 2014).

Taken together these findings suggest that in contrast to the models proposed by both Beck (1967) and Abramson et al. (1989), biases in cognition are more likely to be the result of both genetic and environmental influences. Of particular interest, this pattern of results also indicates that genetic effects may be developmentally sensitive such that genetic effects on specific biases may only begin to emerge during adolescence.

Genes, Cognitive Biases, and Emotional Disorders

In addition to providing evidence that cognitive biases are influenced by genetic factors, quantitative genetic studies have also provided insight into the interplay between genetic and cognitive factors in the etiology of emotional disorders. Using a bivariate extension of the twin model, Eley et al. (2008) reported a genetic correlation of 0.65 for interpretation biases and depression in a sample of 8-year-old children. This suggests a considerable degree of overlap in the genes that cause interpretation biases and those that cause depressive symptoms. Similar findings were reported in a sample of 15-year-old adolescents, in which the genetic correlation between depression symptoms and attributional style was −0.47 (Lau et al., 2006). These effects were further explored in a longitudinal follow-up of the same sample, with further data collected at age 17. Applying a cross-lagged model to these data, Lau and Eley (2008a) showed that attributional style at age 15 predicted depression symptoms at age 17. There was evidence for a reciprocal relationship, with earlier depression symptoms also predicting later attributional style. Nevertheless, even when these effects were taken into account, there remained a significant relationship between earlier attributional style and later depression, which appeared to be explained by a substantial genetic component. These findings suggest that attributional style lies on a causal pathway between genes and depression by increasing sensitivity to adversity. Lau and Eley (2008a, 2008b) provided further evidence to support this hypothesis by showing that the genetic correlation between attributional style and depression was greater in those individuals reporting more stressful life events. The pattern seems to be that genes might lead to individual differences in attributional style, and those negative attributional styles then subsequently lead to depression, but only in the context of stress.

In support of this hypothesis, the same genetic variants that have been implicated in molecular GxE studies of depression and anxiety in epidemiological samples have also been shown to be associated with cognitive biases. For example, multiple studies have shown that individuals homozygous for the S allele of the 5-HTTLPR genotype display greater biases toward emotional stimuli when compared to L homozygotes—a finding that has been supported by a large meta-analysis (Pergamin-Hight, Bakermans-Kranenburg, van Ijzendoorn, & Bar-Haim, 2012). Studies also report that the L allele is associated with positive cognitive biases (Fox, Ridgewell,

& Ashwin, 2009). Similar findings have been reported between further variants implicated in GxE and cognitive biases in genes including BDNF (Beevers, Wells, & McGeary, 2009), FKBP5 (Cristóbal-Narváez et al., 2016; Fani et al., 2013), COMT (Gong et al., 2013; Herrmann et al., 2009), and DRD2 (Gong et al., 2013).

Gene Environment Interplay and Cognitive Biases

In addition to a direct effect of genes, there is also evidence for GxE in the development of cognitive biases. For example, it has been reported that the 5-HTTLPR moderates the effects of childhood physical abuse on attentional biases for angry faces (Johnson, Gibb, & McGeary, 2010). Specifically, the relationship between childhood physical abuse and negative attention bias was found to be stronger for S allele carriers relative to those homozygous for the L allele. Similarly, carriers of the S allele report higher levels of rumination but only in the context of recent life stress (Canli et al., 2006) or childhood emotional abuse (Antypa & Van der Does, 2010). There is further evidence that these GxE effects on cognitive biases reflect differential susceptibility, influencing response to both positive and negative environmental influences. Using a sample of healthy volunteers, for instance, Fox, Zougkou, Ridgewell, and Garner (2011) used an attention bias modification (ABM) task designed to induce either a bias toward negative or toward positive affective pictures in different groups of participants. It was found that S allele carriers of the 5-HTTLPR developed stronger biases toward both negative and positive affective pictures when compared to individuals homozygous for the L allele. This suggests that in line with the differential susceptibility hypothesis, S allele carriers may be more sensitive to both positive and negative environmental influences and therefore acquire positive or negative biases more easily than L carriers depending on the environmental context.

Challenges and Future Directions

Both quantitative and molecular genetic studies provide preliminary support for our expanded CogBIAS model in showing interactions between genetic and cognitive factors in the development of emotional vulnerability. However, existing research is hampered by (1) the reliability of some cognitive bias measures, (2) a focus in most studies on assessing just one bias (e.g., attention or interpretation or memory) rather than addressing multiple biases in the same study, (3) a reliance on cross-sectional data, and (4) a focus on a single gene or a small set of genes in the majority of studies rather than taking a whole-genome approach. Larger-scale studies utilizing both polygenic and poly-bias scores, preferably in longitudinal designs that assess both negative and positive environmental influences, are required to move the field forward.

Improving the Reliability of Measures of Cognitive Biases

Indicators of statistical reliability (e.g., test-retest reliability or internal consistency) have not been typically assessed or reported in behavioral measures of cognitive functions (Hedges, Powell, & Sumner 2018) as is standard in subjective psychometric measures of mood states. This has been a particular problem with the attentional-probe task, which typically shows very low reliability (Parsons et al., 2018; Price et al., 2015). Low or absent measures of reliability in behavioral tasks raises particular problems when attempting to integrate genetic and cognitive approaches to emotional vulnerability and well-being. For instance, we have been attempting to incorporate some behavioral measures of cognitive biases into large-scale GWAS studies for a number of years now, but unless reliability is close to 0.8 or above, statistical geneticists will typically not approve of the incorporation of these measures into GWAS studies. Hence, most GWAS studies rely on indicators such as diagnosis- or questionnaire-based measure of anxiety or neuroticism, which have a reliability of above 0.8 Cognitive biases reflect distorted ways in which sensory information is perceived, processed, and remembered. While some traction on these processes can be gained by subjective reports, it is likely that online behavioral measures using reaction time, eye tracking, or measures of neural activity will provide data that is closer to the endophenotype of interest (e.g., propensity to develop anxiety or depression). Given the likely role that cognitive processing biases play in the pathway from genetic vulnerability to psychopathology or well-being, there is an urgent need for researchers to develop behavioral measures of these fundamental processing biases that have a high degree of statistical reliability. While much attention has focused on the reliability of the attentional-probe task, it is worth noting that because reliability is rarely measured for behavioral tasks, it is unknown as to whether the reliability of measures of interpretation or memory bias is any better.

Combining Cognitive Biases

The majority of studies have investigated cognitive biases in interpretation, attention, or memory separately so that little is known about how different cognitive biases work together to maintain psychopathology. An exception is the *combined cognitive bias hypothesis*, in which it was proposed that the combined effects of cognitive biases may have a greater impact on sustaining a disorder than if individual biases were to work in isolation (Hirsch et al. 2006). The interrelationships among multiple cognitive biases that may operate simultaneously and/or in succession could influence each other in a number of ways. Attention bias in the initial encoding phase of information processing, for instance, might influence subsequent biases such as memory bias (e.g., Russo et al., 2001). Alternatively, these biases may operate simultaneously but independently from each other (Everaert et al.

2012; Hirsch et al., 2006). Another causal pathway proposed by the combined cognitive bias hypothesis (Fox & Beevers, 2016) and incorporated in our extension of this model is that cognitive biases may have bidirectional effects, in which a range of reciprocal relationships could exist between the different biases. Thus, a bias in attention may influence subsequent interpretation processes which, in turn, may bias ongoing attentional processing.

Examining cognitive biases together rather than in isolation is likely to lead to a more comprehensive understanding of the cognitive processes that underpin psychopathology (Everaert et al., 2012; Everaert, Duyck, & Koster, 2014; Hertel et al. 2008; Hirsch et al., 2006; Klein, de Voogd, Wiers, & Salemink, 2017). We are taking this approach in a longitudinal study testing the CogBIAS hypothesis in which we are following 500 adolescents for a 5-year period to investigate the influence of a variety of cognitive biases in a *poly-bias score* in addition to examining a variety of polygenic scores (Booth, Songco, Parsons, et al., 2017). Utilizing multiple cognitive biases as well as multiple genetic variants is an important focus for future research to deepen our understanding of the components and determinants of emotional vulnerability and psychological well-being.

Demonstrating Causality Between Genes, Cognitive Biases, Psychopathology, and Psychological Well-being

We have hypothesized that genetic correlations between cognitive biases and mental health indices could suggest that cognitive biases lie on a causal pathway between genes and disorders. However, there are several other plausible explanations for such a correlation. One possibility is that cognitive biases may be a *consequence* of emotional disorders or both cognitive biases and emotional disorders may be caused by a higher-order trait under genetic control, such as neuroticism. Establishing a causal pathway from genes to cognitive biases and subsequent emotional disorders and well-being ultimately requires a multi-wave longitudinal design in which cognitive biases, emotional disorders, and well-being are measured at each time point. Applying cross-lagged mediation models would allow for investigation of the extent to which cognitive biases mediate the effects of genes on emotional disorders and well-being or whether emotional disorders and well-being mediate the effects of genes on cognitive biases.

A complementary approach to establishing causality in this context would be to manipulate cognitive biases and test these effects on subsequent emotional symptoms and well-being. While effects on symptoms of anxiety or depression are mixed, a large number of studies report that cognitive bias modification techniques (e.g., Fox, Derakshan, & Standage, 2011; Fox, Zougkou, et al., 2011) have the potential to alter cognitive biases. If indeed cognitive biases do mediate the relationship between genetic risk and emotional disorders, we might expect that genes would explain less of the variance in emotional disorders and well-being following successful reduction of negative biases following an intervention such as cognitive bias modification.

Moving Beyond Candidate Gene Studies

While quantitative genetic studies provide solid support for our CogBIAS model, evidence from molecular genetics remains limited to a handful of candidate genes previously implicated in emotional disorders and gene-environment interaction. Given the limitations of the candidate gene approach, and the fact that complex traits such as cognitive biases are influenced by many thousands of gene variants of very small effects, we argue that a robust test of the CogBIAS model requires genome-wide data. Given an adequately sized sample, polygenic methods could then be applied to test the aggregate effects of common genetic variants from across the genome on cognitive biases (the SNP-based heritability). In addition, these methods could be extended to investigate the genetic correlation between cognitive biases and emotional disorders and wellbeing.

A genome-wide test of the relationship between stress sensitivity, vantage sensitivity, and differential susceptibility variants and cognitive biases is, of course, challenging, although potentially highly informative. Stress-sensitivity and vantage-sensitivity variants are likely to show at least marginal main effects in GWAS of emotional disorders and well-being, respectively. Stratified analysis of such samples could be used to prioritize genetic variants that show greater effects in negative or positive environments, respectively, to create stress-sensitivity and vantage-sensitivity polygenic scores. These new polygenic scores could then be tested for their association with cognitive biases and emotional disorders and well-being. Identifying and testing the effects of differential susceptibility variants from GWAS data represent a further challenge. However, recent studies suggest that it is possible to produce environmental-sensitivity polygenic scores that moderate responses to both negative and positive environmental influences (e.g., Keers et al., 2016). Testing these scores in relation to cognitive biases and emotional disorders and well-being may allow us to explore the involvement of differential susceptibility variants in the generation and effects of cognitive biases.

Conclusion

Our expanded CogBIAS hypothesis provides a theoretical framework to explain how genetic, environmental, and cognitive factors interact in the development of emotional disorders and well-being. While our model is supported by existing findings from experimental psychology, quantitative, and molecular genetics, further studies are required to fully test the CogBIAS hypothesis. These studies should focus on the combined effects of reliably measured cognitive biases and move beyond a candidate-gene to a genome-wide, polygenic approach. Future studies should also apply longitudinal, developmentally sensitive designs to elucidate causal pathways between genes, environments, cognitive biases, and disorders.

Unraveling these pathways is a crucial next step for understanding the etiology and maintenance of emotional disorders and may also provide novel targets for prevention and intervention.

References

Abramson, L. Y., Alloy, L. B., & Metalsky, G. I. (1989). Hopelessness depression - A theory-based subtype of depression. *Psychological Review, 96*(2), 358–372. https://doi.org/10.1037/0033-295X.96.2.358

Alloy, L. B., Abramson, L. Y., Hogan, M. E., Whitehouse, W. G., Rose, D. T., Robinson, M. S., … Lapkin, J. B. (2000). The temple-Wisconsin cognitive vulnerability to depression project: Lifetime history of axis I psychopathology in individuals at high and low cognitive risk for depression. *Journal of Abnormal Psychology, 109*(3), 403–418. https://doi.org/10.1037//0021-843X.109.3.403

Alloy, L. B., Abramson, L. Y., Whitehouse, W. G., & Hogan, M. E. (2006). Prospective incidence of first onsets and recurrences of depression in individuals at high and low cognitive risk for depression. *Journal of Abnormal Psychology, 115*(1), 145–156. https://doi.org/10.1037/0021-843X.115.1.145|10.1037/0021-843X.115.145

Anderson, J., & Bower, G. H. (1973). *Human associative memory*. Washington, DC: Winston.

Anokhin, A. P., Golosheykin, S., Grant, J. D., & Heath, A. C. (2010). Developmental and genetic influences on prefrontal function in adolescents: A longitudinal twin study of WCST performance. *Neuroscience Letters, 472*(2), 119–122. https://doi.org/10.1016/j.neulet.2010.01.067

Antypa, N., & van der Does, A.J.W. (2010). Serotonin transporter gene, childhood emotional abuse and cognitive vulnerability to depression. *Genes Brain & Behavior*, *9*, 615–620.

Askelund, A. D., Schweizer, S., Goodyer, I. M., & van Harmelen, A.-L. (2019). Positive memory specificity is associated with reduced vulnerability to depression. *Nature Human Behaviour, 3*, 265. https://doi.org/10.1038/s41562-018-0504-3

Auerbach, R. P., Stanton, C. H., Proudfit, G. H., & Pizzagalli, D. A. (2015). Self-referential processing in depressed adolescents: A high-density event-related potential study. *Journal of Abnormal Psychology, 124*(2), 233–245. https://doi.org/10.1037/abn0000023

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M., & van IJzendoorn, M. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin, 133*(1), 1–24. https://doi.org/10.1037/0033-2909.133.1.1

Barry, T. J., Vervliet, B., & Hermans, D. (2015). An integrative review of attention biases and their contribution to treatment for anxiety disorders. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00968

Bartels, M. (2015). Genetics of wellbeing and its components satisfaction with life, happiness, and quality of life: A review and meta-analysis of heritability studies. *Behavior Genetics, 45*(2), 137–156. https://doi.org/10.1007/s10519-015-9713-y

Beck, A. (1967). *Depression: Clinical, experimental, and theoretical aspects*. New York, NY: Hoeber.

Beck, A., Emery, G., & Greenberg, R. (1985). *Anxiety disorders and phobias: A cognitive perspective*. New York, NY: Basic Books.

Beck, A., Steer, R., & Brown, G. (1996). *Manual for the Beck depression inventory-II*. San Antonio, TX: Psychological Corporation.

Beevers, C. G., Wells, T. T., & McGeary, J. E. (2009). The BDNF Val66Met polymorphism is associated with rumination in healthy adults. *Emotion, 9*(4), 579.

Belsky, J., Bakermans-Kranenburg, M. J., & van Ijzendoorn, M. H. (2007). For better and for worse: Differential susceptibility to environmental influences. *Current Directions in Psychological Science, 16*(6), 300–304. https://doi.org/10.1111/j.1467-8721.2007.00525.x

Belsky, J., Jonassaint, C., Pluess, M., Stanton, M., Brummett, B., & Williams, R. (2009). Vulnerability genes or plasticity genes? *Molecular Psychiatry, 14*, 746–754.

Bergen, S. E., Gardner, C. O., & Kendler, K. S. (2007). A meta-analysis of age-related changes in heritability of behavioral phenotypes over adolescence and young adulthood. *Behavior Genetics, 37*(6), 738–739.

Berna, C., Lang, T. J., Goodwin, G. M., & Holmes, E. A. (2011). Developing a measure of interpretation bias for depressed mood: An ambiguous scenarios test. *Personality and Individual Differences, 51*(3), 349–354. https://doi.org/10.1016/j.paid.2011.04.005

Booth, C., Songco., A., Parsons, S., Heathcote, L., Vincent, J., Keers, R., & Fox, E. (2017). The CogBIAS longitudinal study protocol: Cognitive and genetic factors influencing psychological functioning in adolescence. *BMC Psychology, 5*: 41.

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. (1987). Commentary on mood and memory. *Behaviour Research and Therapy, 25*(6), 443–455. https://doi.org/10.1016/0005-7967(87)90052-0

Brown, G. W., & Harris, T. O. (1978). *Social origins of depression: A study of psychiatric disorder in women*. New York, NY: Free Press.

Brown, H. M., McAdams, T. A., Lester, K. J., Goodman, R., Clark, D. M., & Eley, T. C. (2013). Attentional threat avoidance and familial risk are independently associated with childhood anxiety disorders. *Journal of Child Psychology and Psychiatry, 54*(6), 678–685. https://doi.org/10.1111/jcpp.12024

Bukh, J. D., Bock, C., Vinberg, M., Werge, T., Gether, U., & Kessing, L. V. (2009). Interaction between genetic polymorphisms and stressful life events in first episode depression. *Journal of Affective Disorders, 119*(1), 107–115.

Calvo, M. G., Eysenck, M. W., & Estevez, A. (1994). Ego-threat interpretive bias in test anxiety - Online inferences. *Cognition & Emotion, 8*(2), 127–146. https://doi.org/10.1080/02699939408408932

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., … Braithwaite, A. (2003). Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science, 301*(5631), 386–389.

Canli, T., Qiu, M., Omura, K., et al (2006). Neural correlates of epigenesis. *Proceedings of the National Academy of Sciences, 103 (43)*, 16033–16038.

Chen, J., Li, X., & McGue, M. (2012). Interacting effect of BDNF Val66Met polymorphism and stressful life events on adolescent depression. *Genes, Brain and Behavior, 11*(8), 958–965.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and two ears. *The Journal of the Acoustical Society of America, 25*, 5.

Clark, D. A., & Beck, A. T. (2010). Cognitive theory and therapy of anxiety and depression: Convergence with neurobiological findings. *Trends in Cognitive Sciences, 14*(9), 418–424. https://doi.org/10.1016/j.tics.2010.06.007

Corbetta, M., Kincade, J. M., & Shulman, G. L. (2002). Neural systems for visual orienting and their relationships to spatial working memory. *Journal of Cognitive Neuroscience, 14*(3), 508–523. https://doi.org/10.1162/089892902317362029

Coyne, J. C., & Gotlib, I. H. (1983). The role of cognition in depression - A critical appraisal. *Psychological Bulletin, 94*(3), 472–505. https://doi.org/10.1037//0033-2909.94.3.472

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

Cristóbal-Narváez, P., Sheinbaum, T., Rosa, A., Ballespí, S., de Castro-Catala, M., Peña, E., … Barrantes-Vidal, N. (2016). The interaction between childhood bullying and the FKBP5 gene on psychotic-like experiences and stress reactivity in real life. *PLoS One, 11*(7), e0158809.

David, D., Cristea, I., & Hofmann, S. G. (2018). Why cognitive behavioral therapy is the current gold standard of psychotherapy. *Frontiers in Psychiatry, 9*. https://doi.org/10.3389/fpsyt.2018.00004

Derryberry, D., & Reed, M. A. (2002). Anxiety-related attentional biases and their regulation by attentional control. *Journal of Abnormal Psychology, 111*(2), 225–236. https://doi.org/10.1037/0021-843X.111.2.225

Devine, P. G. (1989). Stereotypes and prejudice - Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*(1), 5–18. https://doi.org/10.1037/0022-3514.56.1.5

Domingue, B. W., Liu, H. X., Okbay, A., & Belsky, D. W. (2017). Genetic heterogeneity in depressive symptoms following the death of a spouse: Polygenic score analysis of the US health and retirement study. *American Journal of Psychiatry, 174*(10), 963–970. https://doi.org/10.1176/appi.ajp.2017.16111209

Dudeney, J., Sharpe, L., & Hunt, C. (2015). Attentional bias towards threatening stimuli in children with anxiety: A meta-analysis. *Clinical Psychology Review, 40*, 66–75. https://doi.org/10.1016/j.cpr.2015.05.007

Duncan, L. E., & Keller, M. C. (2011). A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *American Journal of Psychiatry, 168*(10), 1041–1049. https://doi.org/10.1176/appi.ajp.2011.11020191

Dunn, E. C., Brown, R. C., Dai, Y., Rosand, J., Nugent, N. R., Amstadter, A. B., & Smoller, J. W. (2015). Genetic determinants of depression: Recent findings and future directions. *Harvard Review of Psychiatry, 23*(1), 1–18. https://doi.org/10.1097/HRP.0000000000000054

Eaves, L., Silberg, J., & Erkanli, A. (2003). Resolving multiple epigenetic pathways to adolescent depression. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 44*(7), 1006–1014. https://doi.org/10.1111/1469-7610.00185

Eley, T. C., Gregory, A. M., Lau, J. Y. F., McGuffin, P., Napolitano, M., Rijsdijk, F. V., & Clark, D. M. (2008). In the face of uncertainty: A twin study of ambiguous information, anxiety and depression in children. *Journal of Abnormal Child Psychology, 36*(1), 55–65. https://doi.org/10.1007/s10802-007-9159-7

Eley, T. C., & Stevenson, J. (2000). Specific life events and chronic experiences differentially associated with depression and anxiety in young twins. *Journal of Abnormal Child Psychology, 28*(4), 383–394. https://doi.org/10.1023/a:1005173127117

Elovainio, M., Jokela, M., Kivimäki, M., Pulkki-Råback, L., Lehtimäki, T., Airla, N., & Keltikangas-Järvinen, L. (2007). Genetic variants in the DRD2 gene moderate the relationship between stressful life events and depressive symptoms in adults: Cardiovascular risk in young Finns study. *Psychosomatic Medicine, 69*(5), 391–395.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*(1), 143–149. https://doi.org/10.3758/BF03203267

Everaert, J., Koster, E. H. W., & Derakshan, N. (2012). The combined cognitive bias hypothesis in depression. *Clinical Psychology Review, 32*, 413–424.

Everaert, J., Duyck, W., & Koster, E. H. W. (2014). Attention, interpretation, and memory biases in subclinical depression: A proof-of principle test of the combined cognitive biases hypothesis. Emotion, 14, 331–340.

Eysenck, M. W., & Derakshan, N. (2011). New perspectives in attentional control theory. *Personality and Individual Differences, 50*(7), 955–960. https://doi.org/10.1016/j.paid.2010.08.019

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion, 7*(2), 336–353. https://doi.org/10.1037/1528-3542.7.2.336

Eysenck, M. W., Mogg, K., May, J., Richards, A., & Mathews, A. (1991). Bias in interpretation of ambiguous sentences related to threat in anxiety. *Journal of Abnormal Psychology, 100*(2), 144–150. https://doi.org/10.1037/0021-843X.100.2.144

Fani, N., Gutman, D., Tone, E. B., Almli, L., Mercer, K. B., Davis, J., … Dinov, I. D. (2013). FKBP5 and attention bias for threat: Associations with hippocampal function and shape. *JAMA Psychiatry, 70*(4), 392–400.

Fergusson, D. M., Horwood, L. J., Miller, A. L., & Kennedy, M. A. (2011). Life stress, 5-HTTLPR and mental disorder: Findings from a 30-year longitudinal study. *The British Journal of Psychiatry, 198*(2), 129–135. https://doi.org/10.1192/bjp.bp.110.085993

Fox, E. (1993). Attentional bias in anxiety - Selective or not. *Behaviour Research and Therapy, 31*(5), 487–493. https://doi.org/10.1016/0005-7967(93)90129-I

Fox, E., & Beevers, C. G. (2016). Differential sensitivity to the environment: Contribution of cognitive biases and genes to psychological wellbeing. *Molecular Psychiatry, 21*(12), 1657.

Fox, E., Derakshan, N., & Standage, H. (2011). The assessment of human attention. In K. C. Klauer, C. Stahl, & A. Voss (Eds.), *Cognitive methods in social psychology* (pp. 15–47). New York, NY: Guilford Press.

Fox, E. (2008). Emotion Science: Neuroscientific and Cognitive Approaches to Understanding Human Emotions. Palgrave Macmillan.

Fox, E., Ridgewell, A., & Ashwin, C. (2009). Looking on the bright side: Biased attention and the human serotonin transporter gene. *Proceedings of the Royal Society of London B: Biological Sciences, 276*(1663), 1747–1751.

Fox, E., Russo, R., Bowles, R., & Dutton, K. (2001). Do threatening stimuli draw or hold visual attention in subclinical anxiety? *Journal of Experimental Psychology: General, 130*(4), 681–700. https://doi.org/10.1037/0096-3445.130.4.681

Fox, E., Zougkou, K., Ashwin, C., & Cahill, S. (2015). Investigating the efficacy of attention bias modification in reducing high spider fear: The role of individual differences in initial bias. *Journal of Behavior Therapy and Experimental Psychiatry, 49*, 84–93. https://doi.org/10.1016/j.jbtep.2015.05.001

Fox, E., Zougkou, K., Ridgewell, A., & Garner, K. (2011). The serotonin transporter gene alters sensitivity to attention bias modification: Evidence for a plasticity gene. *Biological Psychiatry, 70*(11), 1049–1054. https://doi.org/10.1016/j.biopsych.2011.07.004

Garber, J., & Flynn, C. (2001). Predictors of depressive cognitions in young adolescents. *Cognitive Therapy and Research, 25*(4), 353–376. https://doi.org/10.1023/A:1005530402239

Gibb, B. E., Alloy, L. B., Abramson, L. Y., Rose, D. T., Whitehouse, W. G., Donovan, P., … Tierney, S. (2001). History of childhood maltreatment, negative cognitive styles, and episodes of depression in adulthood. *Cognitive Therapy and Research, 25*(4), 425–446. https://doi.org/10.1023/A:1005586519986

Gong, P., Shen, G., Li, S., Zhang, G., Fang, H., Lei, L., … Zhang, F. (2013). Genetic variations in COMT and DRD2 modulate attentional bias for affective facial expressions. *PLoS One, 8*(12), e81446.

Graf, P., & Mandler, G. (1984). Activation makes words more accessible, but not necessarily more retrievable. *Journal of Verbal Learning and Verbal Behavior, 23*(5), 553–568. https://doi.org/10.1016/S0022-5371(84)90346-3

Gregory, A. M., Rijsdijk, F. V., Lau, J. Y. F., Napolitano, M., McGuffin, P., & Eley, T. C. (2007). Genetic and environmental influences on interpersonal cognitions and associations with depressive symptoms in 8-year-old twins. *Journal of Abnormal Psychology, 116*(4), 762–775. https://doi.org/10.1037/0021-843X.116.4.762

Gunthert, K. C., Conner, T. S., Armeli, S., Tennen, H., Covault, J., & Kranzler, H. R. (2007). Serotonin transporter gene polymorphism (5-HTTLPR) and anxiety reactivity in daily life: A daily process approach to gene-environment interaction. *Psychosomatic Medicine, 69*(8), 762–768.

Haller, S. P. W., Raeder, S. M., Scerif, G., Kadosh, K. C., & Lau, J. Y. F. (2016). Measuring online interpretations and attributions of social situations: Links with adolescent social anxiety. *Journal of Behavior Therapy and Experimental Psychiatry, 50*, 250–256. https://doi.org/10.1016/j.jbtep.2015.09.009

Hammen, C., & Zupan, B. A. (1984). Self-schemas, depression and the processing of personal information in children. *Journal of Experimental Child Psychology, 37*(3), 598–608. https://doi.org/10.1016/0022-0965(84)90079-1

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108*(3), 356–388. https://doi.org/10.1037/0096-3445.108.3.356

Hedges, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*, 1166–1186.

Herrera, S., Montorio, I., Cabrera, I., & Botella, J. (2017). Memory bias for threatening information related to anxiety: An updated meta-analytic review. *Journal of Cognitive Psychology, 29*(7), 832–854. https://doi.org/10.1080/20445911.2017.1319374

Hertel, P. T., Brozovich, F., Joormann, J., & Gotlib, I. H. (2008). Biases in interpretation and memory in generalized social phobia. *Journal of Abnormal Psychology, 117*(2), 278–288.

Herrmann, M. J., Würflein, H., Schreppel, T., Koehler, S., Mühlberger, A., Reif, A., … Lesch, K.-P. (2009). Catechol-O-methyltransferase Val 158 Met genotype affects neural correlates of aversive stimuli processing. *Cognitive, Affective, & Behavioral Neuroscience, 9*(2), 168–172.

Hertel, P. T., & Mathews, A. (2011). Cognitive bias modification: Past perspectives, current findings, and future applications. *Perspectives on Psychological Science, 6*(6), 521–536. https://doi.org/10.1177/1745691611421205

Hettema, J. M., Neale, M. C., & Kendler, K. S. (2001). A review and meta-analysis of the genetic epidemiology of anxiety disorders. *American Journal of Psychiatry, 158*(10), 1568–1578. https://doi.org/10.1176/appi.ajp.158.10.1568

Hirsch, C. R., Clark, D. M., & Matthews, A. (2006). Imagery and interpretations in social phobia: Support for the combined cognitive biases hypothesis. *Behavior Therapy, 37*(3), 223–236.

Hofmann, S. G., Asnaani, A., Vonk, I. J. J., Sawyer, A. T., & Fang, A. (2012). The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy and Research, 36*(5), 427–440. https://doi.org/10.1007/s10608-012-9476-1

Holmes, E. A., Lang, T. J., & Shah, D. M. (2009). Developing interpretation Bias modification as a "cognitive vaccine" for depressed mood: Imagining positive events makes you feel better than thinking about them verbally. *Journal of Abnormal Psychology, 118*(1), 76–88. https://doi.org/10.1037/a0012590

Holmes, E. A., & Mathews, A. (2005). Mental imagery and emotion: A special relationship? *Emotion, 5*(4), 489–497. https://doi.org/10.1037/1528-3542.5.4.489

Hosang, G. M., Shiles, C., Tansey, K. E., McGuffin, P., & Uher, R. (2014). Interaction between stress and the BDNF Val66Met polymorphism in depression: A systematic review and meta-analysis. *BMC Medicine, 12*(1), 7.

Jacobs, N., Rijsdijk, F., Derom, C., Vlietinck, R., Delespaul, P., Van Os, J., & Myin-Germeys, I. (2006). Genes making one feel blue in the flow of daily life: A momentary assessment study of gene-stress interaction. *Psychosomatic Medicine, 68*(2), 201–206. https://doi.org/10.1097/01.psy.0000204919.15727.43

Johnson, A. L., Gibb, B. E., & McGeary, J. (2010). Reports of childhood physical abuse, 5-HTTLPR genotype, and women's attentional biases for angry faces. *Cognitive Therapy and Research, 34*(4), 380–387. https://doi.org/10.1007/s10608-009-9269-3

Kappenman, E. S., Farrens, J. L., Luck, S. J., & Proudfit, G. H. (2014). Behavioral and ERR measures of attentional bias to threat in the dot-probe task: Poor reliability and lack of correlation with anxiety. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.01368

Keers, R., Coleman, J. R., Lester, K. J., Roberts, S., Breen, G., Thastum, M., … Eley, T. C. (2016). A genome-wide test of the differential susceptibility hypothesis reveals a genetic predictor of differential response to psychological treatments for child anxiety disorders. *Psychotherapy and Psychosomatics, 85*(3), 146–158. https://doi.org/10.1159/000444023

Kendler, K. S. (1996). Major depression and generalised anxiety disorder - Same genes, (partly) different environments - revisited. *British Journal of Psychiatry, 168*, 68–75.

Kendler, K. S., & Eaves, L. J. (1986). Models for the joint effect of genotype and environment on liability to psychiatric illness. *The American Journal of Psychiatry, 143*, 279–289.

Kendler, K. S., Gardner, C. O., & Lichtenstein, P. (2008). A developmental twin study of symptoms of anxiety and depression: Evidence for genetic innovation and attenuation. *Psychological Medicine, 38*(11), 1567–1575. https://doi.org/10.1017/s003329170800384x

Kendler, K. S., Heath, A. C., Martin, N. G., & Eaves, L. J. (1987). Symptoms of anxiety and symptoms of depression - Same genes, different environments. *Archives of General Psychiatry, 44*(5), 451–457.

Kendler, K. S., Kessler, R. C., Walters, E. E., Maclean, C., Neale, M. C., Heath, A. C., & Eaves, L. J. (1995). Stressful life events, genetic liability, and onset of an episode of major depression in women. *American Journal of Psychiatry, 152*(6), 833–842.

Kendler, K. S., Myers, J. M., & Keyes, C. L. M. (2011). The relationship between the genetic and environmental influences on common externalizing psychopathology and mental wellbeing. *Twin Research and Human Genetics, 14*(6), 516–523. https://doi.org/10.1375/twin.14.6.516

Kohler, C. A., Carvalho, A. F., Alves, G. S., McIntyre, R. S., Hyphantis, T. N., & Cammarota, M. (2015). Autobiographical memory disturbances in depression: A novel therapeutic target? *Neural Plasticity*. https://doi.org/10.1155/2015/759139

Koster, E. H. W., Crombez, G., Verschuere, B., & De Houwer, J. (2004). Selective attention to threat in the dot probe paradigm: Differentiating vigilance and difficulty to disengage. *Behaviour Research and Therapy, 42*(10), 1183–1192. https://doi.org/10.1016/j.brat.2003.08.001

Klein, A. M., de Voogd, L., Wiers, R. W., Salemink, E. (2017). Biases in attention and interpretation in adolescents with varying levels of anxiety and depression. *Cognition and Emotion, 3*, 1–9.

Kruijt, A. W., Parsons, S., & Fox, E. (2018). No evidence for attention bias towards threat in clinical anxiety and PTSD: A meta-analysis of baseline dot-probe bias in attention bias modification RCTs. *PsyArXiv Preprints*.

Lau, J. Y. F., Belli, S. D., Gregory, A. M., Napolitano, M., & Eley, T. C. (2012). The role of children's negative attributions on depressive symptoms: An inherited characteristic or a product of the early environment? *Developmental Science, 15*(4), 569–578. https://doi.org/10.1111/j.1467-7687.2012.01152.x

Lau, J. Y. F., Belli, S. R., Gregory, A. M., & Eley, T. C. (2014). Interpersonal cognitive biases as genetic markers for pediatric depressive symptoms: Twin data from the emotions, cognitions, heredity and outcome (ECHO) study. *Development and Psychopathology, 26*(4), 1267–1276. https://doi.org/10.1017/S0954579414001011

Lau, J. Y. F., & Eley, T. C. (2008a). Attributional style as a risk marker of genetic effects for adolescent depressive symptoms. *Journal of Abnormal Psychology, 117*(4), 849–859. https://doi.org/10.1037/a0013943

Lau, J. Y. F., & Eley, T. C. (2008b). Disentangling gene-environment correlations and interactions on adolescent depressive symptoms. *Journal of Child Psychology and Psychiatry, 49*(2), 142–150. https://doi.org/10.1111/j.1469-7610.2007.01803.x

Lau, J. Y. F., Rijsdijk, F., & Eley, T. C. (2006). I think, therefore I am: A twin study of attributional style in adolescents. *Journal of Child Psychology and Psychiatry, 47*(7), 696–703. https://doi.org/10.1111/j.1469-7610.2005.01532.x

Lau, J. Y. F., & Waters, A. M. (2017). Annual research review: An expanded account of information-processing mechanisms in risk for child and adolescent anxiety and depression. *Journal of Child Psychology and Psychiatry, 58*(4), 387–407. https://doi.org/10.1111/jcpp.12653

Liu, Z., Liu, W., Yao, L., Yang, C., Xiao, L., Wan, Q., … Wang, G. (2013). Negative life events and corticotropin-releasing-hormone receptor1 gene in recurrent major depressive disorder. *Scientific Reports, 3*, 1548.

Lopez-Leon, S., Janssens, A., Ladd, A., Del-Favero, J., Claes, S. J., Oostra, B. A., & van Duijn, C. M. (2008). Meta-analyses of genetic studies on major depressive disorder. *Molecular Psychiatry, 13*(8), 772–785. https://doi.org/10.1038/sj.mp.4002088

MacLeod, C., & Cohen, I. (1993). Anxiety and the interpretation of ambiguity - A text comprehension study. *Journal of Abnormal Psychology, 102*(2), 238–247. https://doi.org/10.1037/0021-843X.102.2.238

MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology, 95*(1), 15–20. https://doi.org/10.1037//0021-843X.95.1.15

Mandelli, L., Serretti, A., Marino, E., Pirovano, A., Calati, R., & Colombo, C. (2007). Interaction between serotonin transporter gene, catechol-O-methyltransferase gene and stressful life events in mood disorders. *International Journal of Neuropsychopharmacology, 10*(4), 437–447.

Mathews, A., & MacLeod, C. (2005). Cognitive vulnerability to emotional disorders. *Annual Review of Clinical Psychology, 1*, 167–195. https://doi.org/10.1146/annurev.clinpsy.1.102803.143916

Miers, A. C., Blote, A. W., Bogels, S. M., & Westenberg, P. M. (2008). Interpretation bias and social anxiety in adolescents. *Journal of Anxiety Disorders, 22*(8), 1462–1471. https://doi.org/10.1016/j.janxdis.2008.02.010

Mogg, K., Bradbury, K. E., & Bradley, B. P. (2006). Interpretation of ambiguous information in clinical depression. *Behaviour Research and Therapy, 44*(10), 1411–1419. https://doi.org/10.1016/j.brat.2005.10.008

Mogg, K., & Bradley, B. P. (2006). Time course of attentional bias for fear-relevant pictures in spider-fearful individuals. *Behaviour Research and Therapy, 44*(9), 1241–1250. https://doi.org/10.1016/j.brat.2006.05.003

Mogg, K., & Bradley, B. P. (2016). Anxiety and attention to threat: Cognitive mechanisms and treatment with attention bias modification. *Behaviour Research and Therapy, 87*, 76–108. https://doi.org/10.1016/j.brat.2016.08.001

Mogg, K., & Bradley, B. P. (2018). Anxiety and threat-related attention: Cognitive-motivational framework and treatment. *Trends in Cognitive Sciences, 22*(3), 225–240. https://doi.org/10.1016/j.tics.2018.01.001

Mogg, K., Holmes, A., Garner, M., & Bradley, B. P. (2008). Effects of threat cues on attentional shifting, disengagement and response slowing in anxious individuals. *Behaviour Research and Therapy, 46*(5), 656–667. https://doi.org/10.1016/j.brat.2008.02.011

Moran, T. P., Jendrusina, A. A., & Moser, J. S. (2013). The psychometric properties of the late positive potential during emotion processing and regulation. *Brain Research, 1516*, 66–75. https://doi.org/10.1016/j.brainres.2013.04.018

Mullins, N., Power, R. A., Fisher, H. L., Hanscombe, K. B., Euesden, J., Iniesta, R., … Shi, J. (2016). Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. *Psychological Medicine, 46*(04), 759–770.

Munafo, M. R., Durrant, C., Lewis, G., & Flint, J. (2009). Gene X environment interactions at the serotonin transporter locus. *Biological Psychiatry, 65*(3), 211–219. https://doi.org/10.1016/j.biopsych.2008.06.009

Nanni, V., Uher, R., & Danese, A. (2012). Childhood maltreatment predicts unfavorable course of illness and treatment outcome in depression: A meta-analysis. *The American Journal of Psychiatry, 169*(2), 141–151. https://doi.org/10.1176/appi.ajp.2011.11020335

Neisser, U. (1967). *Cognitive psychology*. Englewood Cliffs, NJ: Prentice Hall.

Okbay, A., Baselmans, B., De Neve, J. E., Turley, P., Nivard, M., Fontana, M., … Cesarini, D. (2016). Novel genetic loci for neuroticism and depression identified using subjective well-being as a proxy-phenotype. *Behavior Genetics, 46*(6), 791–791.

Parsons, S., Kruijt, A. W., & Fox, E. (2018). Psychological science needs a standard practice of reporting the reliability of cognitive behavioural measurements. *PsyArxiv Preprints*. Retrieved from https://psyarxiv.com/6ka9z

Pergamin-Hight, L., Bakermans-Kranenburg, M. J., van Ijzendoorn, M. H., & Bar-Haim, Y. (2012). Variations in the promoter region of the serotonin transporter gene and biased attention for emotional information: A meta-analysis. *Biological Psychiatry, 71*(4), 373–379.

Peyrot, W. J., Milaneschi, Y., Abdellaoui, A., Sullivan, P. F., Hottenga, J. J., Boomsma, D. I., & Penninx, B. W. (2014). Effect of polygenic risk scores on depression in childhood trauma. *The British Journal of Psychiatry, 205*, 113–119. https://doi.org/10.1192/bjp.bp.113.143081

Peyrot, W. J., Van der Auwera, S., Milaneschi, Y., Dolan, C. V., Madden, P. A. F., Sullivan, P. F., … Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. (2018). Does childhood trauma moderate polygenic risk for depression? A meta-analysis of 5765 subjects from the Psychiatric Genomics Consortium. *Biological Psychiatry, 84*(2), 138–147. https://doi.org/10.1016/j.biopsych.2017.09.009

Phillips, W. J., Hine, D. W., & Thorsteinsson, E. B. (2010). Implicit cognition and depression: A meta-analysis. *Clinical Psychology Review, 30*(6), 691–709. https://doi.org/10.1016/j.cpr.2010.05.002

Pluess, M., & Belsky, J. (2013). Vantage sensitivity: Individual differences in response to positive experiences. *Psychological Bulletin, 139*(4), 901–916. https://doi.org/10.1037/a0030196

Polanczyk, G., Caspi, A., Williams, B., Price, T. S., Danese, A., Sugden, K., & Moffitt, T. E. (2009). Protective effect of CRHR1 gene variants on the development of adult depression following childhood maltreatment: Replication and extension. *Archives of General Psychiatry, 66*(9), 978–985.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology, 109*(2), 160–174.

Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology, 61*(3), 380–391.

Price, R., Kuckertz, J., Siegle, G., Ladouceur, C., Silk, J., Ryan, N., & Amir, N. (2015). Empirical recommendations for improving the stability of the dot-probe task in clinical research. *Psychological Assessment, 27*(2), 365–376.

Reilly, L. C., Ciesla, J. A., Felton, J. W., Weitlauf, A. S., & Anderson, N. L. (2012). Cognitive vulnerability to depression: A comparison of the weakest link, keystone and additive models. *Cognition and Emotion, 26*(3), 521–533.

Reutter, M., Hewig, J., Wieser, M. J., & Osinsky, R. (2017). The N2pc component reliably captures attentional bias in social anxiety. *Psychophysiology, 54*(4), 519–527.

Richards, A., & French, C. C. (1992). An anxiety-related bias in semantic activation when processing threat/neutral homographs. *Quarterly Journal of Experimental Psychology, 45*(3), 503–525.

Risch, N., Herrell, R., Lehner, T., Liang, K.-Y., Eaves, L., Hoh, J., … Merikangas, K. R. (2009). Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: A meta-analysis. *JAMA, 301*(23), 2462–2471.

Rubin, D. C., & Friendly, M. (1986). Predicting which words get recalled: Measures of free recall availability, goodness, emotionality and pronunciability for 925 nouns. *Memory and Cognition, 14*, 79–94.

Russo, R., Fox, E., Bellinger, L., & Nguyen-Van-Tam, D. P. (2001). Mood-congruent free recall bias in anxiety. *Cognition and Emotion, 15*, 419–433.

Russo, R., Fox, E., & Bowles, R. J. (1999). On the status of implicit memory bias in anxiety. *Cognition and Emotion, 13*, 435–456.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1–66.

Schoth, D. E., & Liossi, C. (2017). A systematic review of experimental paradigms for exploring biased interpretation of ambiguous information with emotional and neutral associations. *Frontiers in Psychology, 8*, 171. https://doi.org/10.3389/fpsyg.2017.00171

Schulman, P., Keith, D., & Seligman, M. E. P. (1993). Is optimism heritable - A study of twins. *Behaviour Research and Therapy, 31*(6), 569–574. https://doi.org/10.1016/0005-7967(93)90108-7

Scourfield, J., Rice, F., Thapar, A., Harold, G. T., Martin, N., & McGuffin, P. (2003). Depressive symptoms in children and adolescents: Changing aetiological influences with development. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 44*(7), 968–976. https://doi.org/10.1111/1469-7610.00181

Silberg, J., Rutter, M., Neale, M., & Eaves, L. (2001). Genetic moderation of environmental risk for depression and anxiety in adolescent girls. *The British Journal of Psychiatry, 179*, 116–121.

Staugaard, S. R. (2009). Reliability of two versions of the dot-probe task using photographic faces. *Psychology Science Quarterly, 51*(3), 339–350.

Stroop, J. R. (1935). Studies of interference in verbal reactions. *Journal of Experimental Psychology, 18*, 643–663.

Stuewig, J., & McCloskey, L. A. (2005). The relation of child maltreatment to shame and guilt among adolescents: Psychological routes to depression and delinquency. *Child Maltreatment, 10*(4), 324–336. https://doi.org/10.1177/1077559505279308

Sulik, M., Eisenberg, N., Lemery-Chalfant, K., Spinrad, T., Silva, K., Eggum, N., & Verrelli, B. (2012). Interactions between serotonin transporter gene haplotypes and quality of mothers' parenting predict the development of children's noncompliance. *Developmental Psychology, 48*(3), 740–754. https://doi.org/10.1037/a0025938

Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: Review and meta-analysis. *The American Journal of Psychiatry, 157*, 1552–1562.

Sumner, J. A., Griffith, J. W., & Mineka, S. (2010). Overgeneral autobiographical memory as a predictor of the course of depression: A meta-analysis. *Behavior Research and Therapy, 48*(7), 614–625.

Trouton, A., Spinath, F. M., & Plomin, R. (2002). Twins early development study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behavior problems in childhood. *Twin Research, 5*(5), 444–448. https://doi.org/10.1375/136905202320906255

Uher, R., & McGuffin, P. (2010). The moderation by the serotonin transporter gene of environmental adversity in the etiology of depression: 2009 update. *Molecular Psychiatry, 15*(1), 18–22. https://doi.org/10.1038/mp.2009.123

Van Bockstaele, B., Verschuere, B., Tibboel, H., De Houwer, J., Crombez, G., & Koster, E. H. W. (2014). A review of current evidence for the causal impact of attentional Bias on fear and anxiety. *Psychological Bulletin, 140*(3), 682–721. https://doi.org/10.1037/a0034834

van Ijzendoorn, M. H., & Bakermans-Kranenburg, M. J. (2015). Genetic differential susceptibility on trial: Meta-analytic support from randomized controlled experiments. *Development and Psychopathology, 27*(01), 151–162.

Weinberg, A., Venables, N. C., Proudfit, G. H., & Patrick, C. J. (2015). Heritability of the neural response to emotional pictures: Evidence from ERPs in an adult twin sample. *Social Cognitive and Affective Neuroscience, 10*(3), 424–434. https://doi.org/10.1093/scan/nsu059

Weissman, A. N., & Beck, A. T. (1978, November). Development and validation of the dysfunctional attitude scale. In *Paper Presented at the Annual Meeting of the Association for Advanced Behavior Therapy Chicago*.

Wenzlaff, R. M. (1993). The mental control of depression: Psychological obstacles to emotional Well-being. In D. M. Wegner & J. W. Pennebaker (Eds.), *Handbook of mental control* (pp. 239–257). Englewood Cliffs, NJ: Prentice-Hall.

Wichers, M., Myin-Germeys, I., Jacobs, N., Peeters, F., Kenis, G., Derom, C., … Van Os, J. (2007). Genetic risk of depression and stress-induced negative affect in daily life. *British Journal of Psychiatry, 191*, 218–223. https://doi.org/10.1192/bjp.bp.106.032201

Williams, J. M. G., & Broadbent, K. (1986). Autobiographical memory in suicide attempters. *Journal of Abnormal Psychology, 95*(2), 144–149.

Williams, J. M. G., Watts, F. N., MacLeod, C., & Mathews, A. M. (1988). *Cognitive psychology and emotional disorders*. Chichester, England: John Wiley.

Williams, J. M. G., Watts, F. N., MacLeod, C., & Mathews, A. M. (1997). *Cognitive psychology and emotional disorders* (2nd ed.). Chichester, England: John Wiley.

Wray, N. R., Lee, S. H., Mehta, D., Vinkhuyzen, A. A. E., Dudbridge, F., & Middeldorp, C. M. (2014). Research review: Polygenic methods and their application to psychiatric traits. *Journal of Child Psychology and Psychiatry, 55*(10), 1068–1087.

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., … Sullivan, P. F. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics, 50*(5), 668–681.

Zimmermann, P., Brückl, T., Nocon, A., Pfister, H., Binder, E. B., Uhr, M., … Holsboer, F. (2011). Interaction of FKBP5 gene variants and adverse life events in predicting depression onset: Results from a 10-year prospective community study. *American Journal of Psychiatry, 168*(10), 1107–1116.

Zvielli, A., Bernstein, A., & Koster, E. H. W. (2014). Temporal dynamics of attentional bias. *Clinical Psychological Science, 3*, 772–788.

Chapter 5
Pathways to Motivational Impairments in Psychopathology: Common Versus Unique Elements Across Domains

Deanna M. Barch, David Pagliaccio, Katherine Luking, Erin K. Moran, and Adam J. Culbreth

Introduction

Our ability as humans to engage in goal-directed actions that allow us to obtain outcomes that we desire is a core component of how we acquire/attain life satisfaction and achievement. Components of motivation are intricately intertwined with such goal-directed behavior, as they are part of the activation driving us to select outcomes that we find enjoyable or satisfying, to put in place effective action plans that allow us to achieve or obtain those outcomes and allow us to keep those goals/action plans represented over extended periods of time when necessary. Sadly, many forms of mental illness involve impairments in varying facets of motivation that are

D. M. Barch (✉)
Department of Psychological & Brain Science, Washington University in St. Louis, St. Louis, MO, USA

Department of Psychiatry, Washington University in St. Louis, St. Louis, MO, USA

Department of Radiology, Washington University in St. Louis, St. Louis, MO, USA
e-mail: dbarch@wustl.edu

D. Pagliaccio
Department of Psychiatry, Columbia University, New York, NY, USA

K. Luking
Department of Psychologial & Brain Sciences, Washington University in St. Louis, St. Louis, MO, USA

E. K. Moran
Department of Psychological & Brain Science, Washington University in St. Louis, St. Louis, MO, USA

Department of Psychiatry, Washington University in St. Louis, St. Louis, MO, USA

A. J. Culbreth
Department of Psychological & Brain Science, Washington University in St. Louis, St. Louis, MO, USA

M. Neta, I. J. Haas (eds.), *Emotion in the Mind and Body*, Nebraska Symposium on Motivation 66, https://doi.org/10.1007/978-3-030-27473-3_5

important contributors to the all too frequently impaired life function and reduced quality of life experienced by individuals with mental health challenges. As such, both the field of psychopathology research broadly and the Research Domain Criteria (RDoC) initiative have recognized the centrality of examining motivation and incentive processing in psychopathology. More specifically, the RDoC includes a "positive valence" systems (PVS) domain (Insel et al., 2010) that outlines a number of constructs that may be critical to understanding the nature and mechanisms of motivational impairments in psychopathology, including responses to receiving rewards or positive outcomes, the processes involved in learning which actions or stimuli predict reward, being able to anticipate future rewards, being able to combine choices and options to estimate the value and cost of different incentives and action plans, and being able to develop and implement, and when needed maintain, action plans that will achieve one's desired outcomes.

A key question in the field of psychopathology research is whether the varying manifestations of impaired motivation seen in different forms of psychopathology arise from a common set of mechanisms that operate transdiagnostically or whether there are one or more mechanisms that may uniquely contribute to motivational impairments in some forms of psychopathology and not others. This review will focus on the types of motivational impairments seen in disorders such as depression and schizophrenia. In schizophrenia, motivational impairments can take the form of reduced efforts to engage in occupational, educational, or social experiences. These aspects of the illness are often captured by what is referred to as "negative" symptoms (symptoms that involve the absence of behaviors or experiences that humans typically have). In the extreme form, individuals with schizophrenia may spend most of their time at home, relatively isolated, and often sitting for long periods of times engaged in relatively low-effort activities (watching TV, etc.). Individuals suffering from depression can also experience what on the surface may seem like similar types of impairments in motivated behavior. Some individuals with depression will also not engage in occupational, educational, or social behaviors that they might participate in when not depressed and may also spend much of their time alone and engage in very passive activities (sleeping, watching TV, etc.). A key question then is whether these seemingly similar types of motivation impairments arise from the same or from different mechanisms. This is a critical question, as if they arise from the same type of impairments; we might be able to develop treatments that are effective transdiagnostically. If not, we may need more disorder-specific interventions.

In the review below, we will argue that elements of the final common pathway linking reward to action in depression and schizophrenia may be shared and are likely to involve deficits in what we will refer to as effort-cost decision-making (ECDM), such that a proximal cause of reduced engagement in occupational, educational, and social pursuits in both depression and schizophrenia reflect a reduced willingness to exert effort to obtain potentially rewarding or positive outcomes (e.g., Barch, Treadway, & Schoen, 2014; Fervaha, Graff-Guerrero, et al., 2013; Gold et al., 2013; Wolf et al., 2014). ECDM requires a number of computations that are part of the RDoC PVS, as will be described in more detail below, including *reward*

responsiveness, the ability to appropriately update *reward values*, and the ability to generate accurate *reward predictions*. ECDM also requires cognitive computations, including *cognitive control* processes such as *goal representation and maintenance* and *performance monitoring*. A wealth of animal and human research suggests that ECDM computations are supported by a cortico-limbic-striatal circuit, including the dorsal and ventral striatum, ventromedial prefrontal, dorsal anterior cingulate, anterior insula, and dorsolateral prefrontal cortex (Croxson, Walton, O'Reilly, Behrens, & Rushworth, 2009; Haber & Behrens, 2014; Prevost, Pessiglione, Metereau, Clery-Melin, & Dreher, 2010; Salamone, Correa, Farrar, & Mingote, 2007; Treadway, Buckholtz, et al., 2012).

However, in this review, we also argue that that this shared proximal ECDM deficit in schizophrenia and depression may reflect differing distal mechanisms. Specifically, we argue that ECDM deficits in psychotic disorders such as schizophrenia reflect difficulties with cognitive control, internal representation of future and/or past events, and use of incentive information that is not currently available in the environment, which may result from impairments in the function and connectivity of the dorsolateral prefrontal cortex and the dorsal anterior cingulate cortex and their association with reward-processing systems. In contrast, we suggest that ECDM deficits in mood pathology such as depression may be more strongly related to reductions in hedonics and reward responsiveness/learning (RDoC PVS) and reward valuation, which may result from impairments in ventral/dorsal striatum, anterior insula, and ventral medial prefrontal cortex function and connectivity.

A Heuristic Model of the Motivation-Action-Outcome Pathway

There are a number of different ways to conceptualize the processes and mechanisms that help individuals translate between experiencing or anticipating an outcome as positive or reinforcing in some way and developing and implementing an effective action plan to achieve that outcome (Berridge, 2012; Berridge & Kringelbach, 2008; Braver et al., 2014; Medic et al., 2014; Schultz, 2016b), with many models and frameworks sharing a number of components. The RDoC PVS (http:\\www.nimh.nih.gov\research-priorities\rdoc\positive-valence-systems-workshop-proceedings.shtml) (Barch, Oquendo, Pacheco, & Morris, 2016) has tried to integrate work from varying models to provide a heuristic framework to guide research on potential mechanisms of impairment in psychopathology.

This organization groups PVS constructs into three superordinate constructs: *reward responsiveness*, *reward learning*, and *reward valuation*. *Reward responsiveness* includes sub-constructs of *initial responsiveness to reward*, *reward anticipation*, and *reward satiation*. *Reward learning* includes subconstructs of *habit*, *reinforcement learning*, and *reward prediction error*. *Reward valuation* includes sub-constructs of *probability*, *delay*, and *effort*. We have also used a complementary

model of the psychological processes and neural systems thought to link experienced or anticipated rewards/incentives with the action plans that need to be generated and maintained in order to obtain these rewards (Barch, Pagliaccio, & Luking, 2018, 2019; Kring & Barch, 2014). We have targeted seven components that we and others have argued are key to the translation of incentive or reward information into behavioral responses (Berridge, 2004, 2012; Berridge & Kringelbach, 2008; Braver et al., 2014; Medic et al., 2014; Schultz, 2007, 2016b; Wallis, 2007) (see Fig. 5.1). Of note, the model illustrated in Fig. 5.1 is not a process model in the sense of suggesting that one component follows another in only the order illustrated in the model. Instead, different components feedback on others and may interact throughout the course of engaging in "motivated" behaviors.

The first component is part of *reward responsiveness* and is termed *initial responsiveness to reward* in the RDoC PVS but has also been referred to as *hedonics or liking* (Fig. 5.1). This component captures the ability to "enjoy" a stimulus or event that may provide pleasure or reward. A number of lines of work suggest that hedonic responses (at least to primary sensory stimuli) may be mediated by activation of the opioid and GABA-ergic systems in the nucleus accumbens shell and its projections to the ventral pallidum, as well as in the orbital frontal cortex (OFC) (e.g., Berridge & Kringelbach, 2015; Berridge, Robinson, & Aldridge, 2009; Kringelbach & Berridge, 2017; Smith & Berridge, 2007).

The second component is also part of *reward responsiveness* and corresponds to *reward anticipation* and has also been described as *wanting* (Fig. 5.1). This relates in important ways to the third component, part of *reward learning*, referred to as

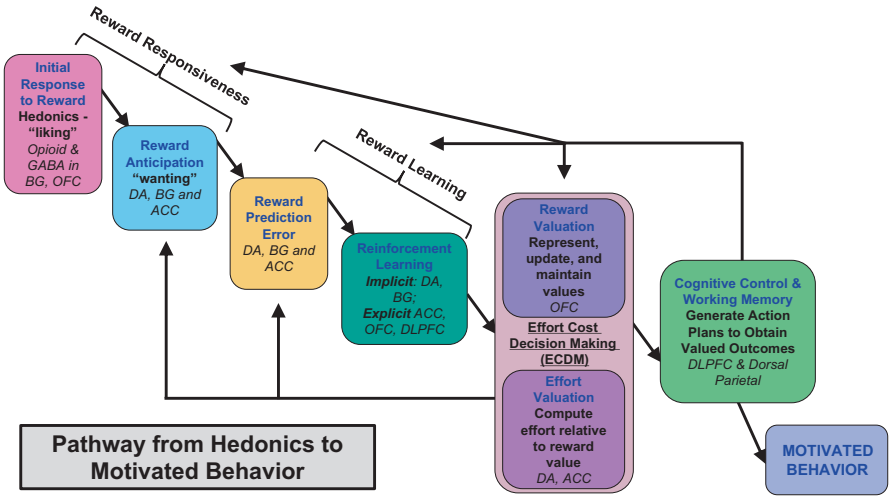

Fig. 5.1 Translating from the experience of reward/pleasure to motivated behavior. *ACC* anterior cingulate cortex, *BG* basal ganglia, *DA* dopamine, *DLPFC* dorsolateral prefrontal cortex, *OFC* orbital frontal cortex. Modified with permission. Note: The organization of this figure is not meant to apply to that processes occur only in the order illustrated, as many can feedback on each other to influenced motivated behavior

reward prediction error (*see below for description*). These components are mediated at least in part by the midbrain dopamine (DA) system, particularly projections to ventral and dorsal striatum (Berridge, 2004; Berridge & Kringelbach, 2015; Kringelbach & Berridge, 2017; Schultz, 2007). At least some DA neurons in the substantia nigra and ventral tegmental area (VTA) respond to stimuli that *predict* reward, as well as to rewards themselves. The degree of response depends importantly on predictability—if the reward was not expected, then the DA neurons fire more strongly (positive reward prediction error, discussed below) than if the reward were fully expected. Interestingly, there can be a transient attenuation in DA neuron firing (negative prediction error, discussed below) if a predicted reward does not occur (e.g., Schultz, 2007, 2016a, 2016b). Further, over time, DA neurons begin to fire to the predictive cues rather than to rewards themselves (Schultz, 2007, 2016a, 2016b). Similar effects have been found in humans using functional magnetic resonance imaging (fMRI) of the ventral and dorsal striatum (e.g., Knutson, Fong, Adams, Varner, & Hommer, 2001; Maia, 2009; Wang, Smith, & Delgado, 2016). These types of DA/striatal responses have been captured reasonably well by temporal difference models that simulate learning about stimuli that predict rewards (e.g., Montague, Dayan, & Sejnowski, 1996), though such a framework does not always map perfectly to human findings (Maia, 2009).

A fourth component is also part of *reward learning* and is called *probabilistic and reinforcement learning* in the RDoC PVS (Fig. 5.1). Such learning can be either implicit (i.e., outside of conscious awareness) or explicit (i.e., including the use of explicit representations about potential reward associations). The types of DA/striatal responses described above for reward prediction are thought to support aspects of reinforcement learning that may occur without conscious awareness (e.g., Frank, Seeberger, & O'Reilly, 2004). At the same time, there is evidence that the development of explicit representations that are accessible to conscious awareness can also drive reinforcement learning, albeit with a potentially different timecourse (e.g., Hazy, Frank, & O'Reilly, 2007). These more explicit forms of reinforcement learning also engage neural systems involved in working memory, cognitive control, and value representations, such as dorsal frontal and parietal regions and the OFC (e.g., Collins, Ciullo, Frank, & Badre, 2017; Collins & Frank, 2012, 2018; Gold, Waltz, et al., 2012; Hazy et al., 2007). By cognitive control, we mean the ability to maintain goal or task representations in order to focus attentional resources on task-relevant information while filtering out task-irrelevant information (e.g., Braver, 2012; Miller & Cohen, 2001).

A fifth component is *reward valuation* in the RDoC PVS (Fig. 5.1). There are a number of components to reward valuation, including integrating information about the intrinsic hedonic properties of a stimulus (sweet foods versus sour foods), the current state of the organism (e.g., value of brownies when hungry versus not) (Rolls, Sienkiewicz, & Yaxley, 1989), delay until a reward can occur (e.g., Rudebeck, Walton, Smyth, Bannerman, & Rushworth, 2006), the probability that a reward will occur (e.g., Cools, Clark, Owen, & Robbins, 2002), other potential rewards that are available in the environment (e.g., brownies versus ice cream), and, as discussed more below, the amount of effort that one might need to allocate to obtain that

reward. Integrating these various sources of information in order to compute the "value" of a particular stimulus or outcome in the current time and context is thought to be mediated as least in part by the OFC (e.g., Conen & Padoa-Schioppa, 2016; Padoa-Schioppa & Cai, 2011; Padoa-Schioppa & Conen, 2017). Functional neuroimaging studies in humans have also demonstrated activation of OFC under modulation of value representations (e.g., O'Doherty, 2007), particularly those in which response contingencies need to be updated, such as reversal learning (e.g., Cools, Lewis, Clark, Barker, & Robbins, 2007; Suzuki, Cross, & O'Doherty, 2017; Tobia et al., 2014; Yan et al., 2016). In addition, humans with OFC lesions can show reversal learning impairments (e.g., Fellows & Farah, 2005).

A sixth component in our model, a subset of *reward valuation* in the RDoC PVS, is the ability to *compute effort relative to reward value* (Fig. 5.1) or what we discussed above as ECDM. This construct refers to determining the cost of engaging in actions necessary to obtain a desired outcome and determining how much of that cost you are willing to undertake or how much effort you are willing to allocate. There are a number of lines of research that suggest that the dorsal anterior cingulate cortex (dACC) may be important for evaluating the cognitive and physical effort associated with different action plans (Holroyd & McClure, 2015; Shenhav, Botvinick, & Cohen, 2013) with contributions of DA input from the nucleus accumbens and related forebrain circuitry (e.g., Botvinick, Huffstetler, & McGuire, 2009; Salamone et al., 2007). As one example, dACC lesions, as well as depletions of accumbens DA, led animals to choose low-effort but low-reward options over higher-reward but higher-effort options (e.g., Hosking, Cocker, & Winstanley, 2015; Rudebeck et al., 2006; Rushworth, Behrens, Rudebeck, & Walton, 2007). Further, the DA system is also critical for ECDM, given that depletion of DA in animals will impair willingness to work for rewards, without changing the hedonic response to rewards (Salamone et al., 2007, 2016). In addition to the dACC, the anterior insula is thought to be important in evaluating the cost of effort, with higher activity when effort is perceived as more costly (Prevost et al., 2010) or when experiencing worse outcomes than expected (Kurniawan, Guitart-Masip, Dayan, & Dolan, 2013). The ventral medial prefrontal cortex has also been shown to be critical to reward valuation, with the suggestion that it may help maintain and integrate such reward values (Treadway, Buckholtz, et al., 2012).

A seventh component is related to the *cognitive systems* domain of RDoC and reflects the ability to *generate and execute goal-directed action plans* necessary to achieve the valued outcome (Fig. 5.1). Many researchers have argued for the role of the dorsolateral PFC in relation to reward and motivation (e.g., Braver & Cohen, 1999; Miller & Cohen, 2001; Wallis, 2007), though a broader network of dorsal frontal and parietal regions, often referred to as the frontal-parietal networks, is also critical for representing and maintaining action plans (Braver, 2012; Dosenbach, Fair, Cohen, Schlaggar, & Petersen, 2008). Importantly, the dorsolateral prefrontal cortex may help represent the value of cognitive effortful rewards in particular, highlighting a potentially important link to ECDM (Massar, Libedinsky, Weiyan, Huettel, & Chee, 2015).

To illustrate how these components may play a role in motivation-related behaviors, let us consider a simplistic example. Imagine that you were debating whether or not to eat some chocolate cake. One consideration is the degree to which you enjoy chocolate cake (reward responsiveness and/or anticipation) as well as how recently you have eaten chocolate cake and what other options you have available (reward satiation and valuation). Another consideration is whether you have the cake in your house now or whether you will have to go out to buy it to get the ingredients to make it (effort allocation). A further consideration is whether you know where to find good chocolate cake (perhaps reward learning) and the degree to which you might have to plan ahead to get the relevant ingredients or to pick up the cake later in the day, requiring you maintain your plan over time (generating and executing action plans). While this is a simple example, it nonetheless captures the ways in which these components are needed to support motivated goal-directed behaviors.

Here we review evidence for psychosis- and depression-related impairments in these components of the model that link hedonic experiences of rewards with goal-directed actions that allow individuals to engage in "motivated" behaviors that allow them to obtain rewards. Where available, we also review neuroimaging evidence to highlight the neurobiological correlates of each type of impairment. As noted above, our guiding framework is the idea that ECDM deficits are a common proximal contributor to motivational impairments across psychosis and depression but that the more distal contributors to ECDM deficits differ in psychosis versus depression.

Mechanisms

Hedonics, Liking, and Responding to Rewards

Depression

There is a robust literature demonstrating that adults and adolescents with or at risk for depression have impaired hedonic responses to both pleasurable stimuli (primary reward) and monetary (secondary) rewards (Keren et al., 2018). Such group differences have been reported using behavioral measures as well as event-related potential (ERP) and fMRI measures of brain function (e.g., Foti, Kotov, Klein, & Hajcak, 2011; Zhang, Chang, Guo, Zhang, & Wang, 2013) and have been related to elevated levels of anhedonia.

Monetary Rewards Depressed individuals show reduced less change in their behavior as a function of reward (e.g., Fletcher et al., 2015; Pizzagalli, Iosifescu, Hallett, Ratner, & Fava, 2008) and reduced ability to learn from reward (e.g., Maddox, Gorlick, Worthy, & Beevers, 2012). Reduced reward sensitivity has been related to self-reported anhedonia (e.g., Pizzagalli et al., 2008; Vrieze et al., 2013) and can be predictive of treatment response (Burkhouse et al., 2016; Vrieze et al.,

2013). Using ERPs, a number of studies have examined the feedback-related negativity (FN) in depression. FN is an ERP component elicited by reward or loss feedback and is thought to reflect activity in the ventral striatum, caudate, and the dorsal ACC (e.g., Carlson, Foti, Mujica-Parodi, Harmon-Jones, & Hajcak, 2011; Liu et al., 2014). Depressed children and adults show decreased FN to rewards (e.g., Belden et al., 2016; Burkhouse, Gorka, Afshar, & Phan, 2017; Foti, Carlson, Sauder, & Proudfit, 2014). Risk for depression (Kujawa, Proudfit, & Klein, 2014; Liu et al., 2016; Whitton et al., 2016) and increased depressive symptoms are also related to reduced FN in adults and children (e.g., Ait Oumeziane & Foti, 2016; Bress, Smith, Foti, Klein, & Hajcak, 2012) and prospectively predict future onset of depression in adolescents (Bress, Foti, Kotov, Klein, & Hajcak, 2013; Nelson, Perlman, Klein, Kotov, & Hajcak, 2016). In addition, the severity of clinically rated anhedonia predicts the magnitude of reduced FN in depressed adults (Liu et al., 2014).

Human neuroimaging studies of reward processing have also found that depression is associated with decreased activation following positive feedback (i.e., reward) in reward-related brain areas such as the caudate, putamen, ACC, and insula (e.g., Luking, Pagliaccio, Luby, & Barch, 2016a; Redlich et al., 2015; Satterthwaite et al., 2015; Zhang et al., 2013). However, not all studies have found such reductions in striatal activity to reward receipt (Ubl et al., 2015). Such reductions have been associated with anhedonia symptoms (e.g., Stoy et al., 2012), have been detected in individuals at risk for depression (e.g., Luking et al., 2016a; Stringaris et al., 2015), and can predict the development of depression in adolescents (Morgan, Olino, McMakin, Ryan, & Forbes, 2013; Stringaris et al., 2015). Research has also shown increased ventral striatal responses to reward following successful treatment (Stoy et al., 2012). These findings are consistent with the hypothesis that current depression and risk for depression are often robustly associated with reduced behavioral and neural responsivity to monetary (secondary) rewards.

Primary Rewards Studies examining self-reported hedonic responses to tastes/odors have been mixed and generally do not find strong evidence for behavioral differences associated with depression. Particularly, hedonic response ratings of sucrose and odor stimuli are generally not different when comparing depressed patients and healthy controls (Berlin, Givry-Steiner, Lecrubier, & Puech, 1998; Clepce, Gossler, Reich, Kornhuber, & Thuerauf, 2010; Dichter, Smoski, Kampov-Polevoy, Gallop, & Garbutt, 2010), and depressive symptom severity in a nonclinical sample did not correlate with pleasantness ratings of sweet, sour, salty, or bitter tastes (Scinska et al., 2004). There is some evidence that elevated levels of anhedonia negatively predict hedonic responses to sucrose across individuals with depression and schizophrenia as well as healthy individuals (Berlin et al., 1998) and that measures of anticipatory anhedonia negatively predict anticipated hedonic responses to chocolate but not actual or recalled responses (Chentsova-Dutton & Hanley, 2010).

fMRI studies provide more consistent evidence for reduced reactivity to primary rewards/pleasant stimuli in depression than the self-report studies discussed above. For example, in one study, individuals with remitted depression showed no difference

in the rating of pleasant food images/tastes as compared to controls but did show decreased ventral striatal response relative to never depressed controls (McCabe, Cowen, & Harmer, 2009). Further, adolescents/young adults at elevated risk for depression based either on a parental history of depression or high self-reports of depression showed lower OFC and/or ACC responses to pleasant food images/tastes (McCabe, Woffindale, Harmer, & Cowen, 2012; Rzepa, Fisk, & McCabe, 2017). fMRI studies utilizing other types of pleasant stimuli, such as social stimuli, happy faces, or pleasant scenes, have also found reduced striatal responses in depressed patients or high-risk groups as compared to controls (Gotlib et al., 2005; Smoski et al., 2011) (Kerestes et al., 2016; Olino, Silk, Osterritter, & Forbes, 2015). Importantly, reduced striatal responses to pleasant stimuli specifically related to elevated levels of anhedonia, rather than to general depressive symptom severity (Keedwell, Andrew, Williams, Brammer, & Phillips, 2005).

Schizophrenia

Monetary Rewards In contrast to the work on depression, neuroimaging studies examining striatal responses to the receipt of monetary rewards in schizophrenia have shown a consistent pattern of intact responses, with robust ventral striatal responses to the receipt of money in unmedicated patients (Nielsen, Rostrup, Wulff, Bak, Lublin, et al., 2012) and patients treated with either typical (primarily targeting dopamine receptors) or atypical (targeting other neurotransmitter systems in addition to dopamine) antipsychotics (e.g., Dowd & Barch, 2012; Gilleen, Shergill, & Kapur, 2014). Further, studies have also shown intact feedback negativity (FN) responses, the ERP component in response to explicit feedback, to the receipts of rewards and losses in schizophrenia (e.g., Llerena, Wynn, Hajcak, Green, & Horan, 2016; Morris, Holroyd, Mann-Wrobel, & Gold, 2011). However, while striatal responses to reward receipt seem to be largely intact in schizophrenia, some of these studies did report abnormal cortical responses to reward receipt. Particularly, prior work has noted reduced reward-related responses in medial PFC (Schlagenhauf et al., 2009), abnormal responses in both medial and lateral PFC (Waltz et al., 2010), and reduced salience coding in ventrolateral PFC in schizophrenia patients, which was correlated with negative symptom severity (Walter et al., 2010).

Primary Rewards A less consistent picture in regard to deficits in schizophrenia emerges from functional neuroimaging studies examining brain responses to other types of pleasurable or rewarding stimuli in schizophrenia (Crespo-Facorro et al., 2001; Paradiso et al., 2003). Plailly, d'Amato, Saoud, and Royet (2006) found reduced activation in schizophrenia within the insula and OFC during hedonicity judgments of positive and negative odors. Schneider et al. (2007) also found reduced activation of the insula during the experience of positive olfactory stimuli in schizophrenia. Taylor, Phan, Britton, and Liberzon (2005) showed reduced phasic ventral striatal responses comparing positive versus neutral picture viewing in both medicated and unmedicated individuals with schizophrenia. Other research has found

evidence for reduced striatal responses to food cues (Grimm, Vollstadt-Klein, Krebs, Zink, & Smolka, 2012) and the receipt of juice, with the magnitude of this reduction associated with the severity of anhedonia (Waltz et al., 2009), though medications may have been a confound in both of these studies. In terms of behavior, individuals with schizophrenia often show impaired judgment of odors, although this is not limited to positive odors (Auster, Cohen, Callaway, & Brown, 2014; Kamath, Lasutschinkow, Ishizuka, & Sawa, 2018; Urban-Kowalczyk, Smigielski, & Kotlicka-Antczak, 2018; Zou et al., 2018), with some evidence that impaired odor judgment is associated with increased anhedonia (Kamath et al., 2018; Zou et al., 2018). Schizophrenia patients also some evidence of altered sensory-specific satiety responses to foods (Waltz et al., 2015).

Summary of Initial Responsiveness to Reward in Depression Compared to Schizophrenia The literature on reward responsiveness in depression and schizophrenia suggests quite different patterns, with much stronger evidence for impairments in initial responsiveness to reward, at least monetary reward, among individuals with depression than among individuals with schizophrenia. Such findings are consistent with the hypothesis that in depression, impairments in motivated behavior may be linked to impairments in initial reward responsiveness. In contrast, the literature does not provide evidence suggesting impaired motivated behavior in schizophrenia arises from initial hedonic responses and instead may reflect alterations in the way information about hedonic experience is stored, represented, maintained, or used.

Reward Anticipation, Reward Prediction Error, and Reinforcement Learning

Depression

Reward Anticipation Not surprisingly given the work on responses to reward in depression, there is evidence that individuals with depression show reduced responses to the anticipation of reward. For example, individuals with depression or with a family history of depression show reduced frontal EEG asymmetries during reward anticipation (e.g., Nelson et al., 2013; Nelson, Kessel, Klein, & Shankman, 2018; Nelson, Shankman, & Proudfit, 2014; Shankman et al., 2013; Shankman, Klein, Tenke, & Bruder, 2007). Further, several studies have found reduced activation in various regions of the striatum during reward anticipation among individuals with current depression or individuals at risk for depression (e.g., Rzepa et al., 2017; Stringaris et al., 2015; Takamura et al., 2017; Ubl et al., 2015), as well as increased activity in the ACC (e.g., Dichter, Kozink, McClernon, & Smoski, 2012; Gorka et al., 2014), and have found that reduced striatal responses to reward predicts depression onset (Stringaris et al., 2015). However, the literature is not fully consistent, as other work has found no differences in striatal activation during reward

anticipation between healthy individuals and those with current (Gorka et al., 2014; Knutson, Bhanji, Cooney, Atlas, & Gotlib, 2008) or remitted depression (Dichter et al., 2012), although in at least one case, it was not clear that any participant, including controls, showed activity in the striatum during reward anticipation (Chase et al., 2013). Is not yet clear why this literature is someone mixed, though it may related to the severity of the depression, the stage of illness, or the age of the participants.

Reward Prediction Error The majority of prediction error studies in depression have found evidence for reduced or disrupted positive prediction errors in depression in the striatum (e.g., Greenberg et al., 2015; Kumar et al., 2018; Robinson, Cools, Carlisi, Sahakian, & Drevets, 2012) and/or the OFC (Rothkirch, Tonn, Kohler, & Sterzer, 2017), with the magnitude of these reductions associated with anhedonia (Greenberg et al., 2015; Rothkirch et al., 2017). However, several studies did not find reduced positive prediction errors in the striatum in depression, at least at the group level (e.g., Rothkirch et al., 2017; Rutledge et al., 2017; Ubl et al., 2015), and one study that found reduced prediction errors in the striatum also found increased prediction error responses in the VTA (Kumar et al., 2008).

Reinforcement Learning Several studies have shown that individuals with depression show impaired reinforcement learning on both implicit and explicit tasks. Implicit tasks are ones where an individual changes their behavior in response to reward but is not aware that they are doing so. This influence of reward can be reflected in a bias to choose a response more likely to receive reward feedback, such as in the probabilistic learning task developed by Pizzagalli (e.g., Fletcher et al., 2015; Pizzagalli et al., 2008). However, individuals with depression also show impairments on more explicit learning tasks, where individuals are aware that one stimulus is more likely to be associated with reward than another and are asked to figure out which one is more likely to be rewarded (Admon et al., 2017; Kumar et al., 2018; Morkl, Blesl, Jahanshahi, Painold, & Holl, 2016). This is seen in remitted depression (Pechtel, Dutra, Goetz, & Pizzagalli, 2013) as well as those at risk for depression (Liu et al., 2016; Luking, Neiman, Luby, & Barch, 2017) and is worse in individuals with depression who have higher anhedonia (e.g., Liverant et al., 2014). Impairments in depression have been found on a reinforcement learning task similar to the weather prediction task (e.g., Herzallah et al., 2013), where it is typically difficult to develop explicit representations of the reward contingencies because of the number of possible combinations of stimuli that could predict rewards. In contrast, the literature on explicit reinforcement learning in depression suggests potentially intact performance. For example, there are a number of studies showing that individuals with depression perform similarly to healthy controls on the same probabilistic selection task that shows consistent impairments in schizophrenia (e.g., Whitmer, Frank, & Gotlib, 2012).

Although not all studies are fully consistent, when aggregated together, the reward anticipation, prediction error, and reinforcement learning literatures provide

support for the hypotheses that a dysfunction in neural responses to reward in the striatum and potentially OFC are an important component of altered reward processing and motivation in depression (Keren et al., 2018). One possibility is that these alterations in reward processing could potentially be reflecting altered DA function (Admon et al., 2017; Minami et al., 2017; Walsh, Browning, Drevets, Furey, & Harmer, 2018). The evidence, albeit modest, for altered dACC responses is also intriguing. Shankman and others have hypothesized that increased dACC activation may reflect "conflict" that individuals with depression experience when asked to anticipate processing positive stimuli that conflict with their current negative emotional state (Gorka et al., 2014). If so, this would suggest that altered dACC activation is an outcome of the phenomenology of depression rather than potentially playing a causal role in anticipatory pleasure impairments. Importantly however, there is also a large literature on altered error-related negativities in depression (e.g., Gorka, Burkhouse, Afshar, & Phan, 2017; Meyer, Bress, Hajcak, & Gibb, 2018; Vaidyanathan, Nelson, & Patrick, 2012; Weinberg, Liu, & Shankman, 2016; Whitton et al., 2017). The error-related negativity (also referred to as reward positivity) is an ERP component that typically occurs in the range of 300–400 ms post-feedback about reward or loss, with a more negative going component for loss feedback compared to reward feedback. Such altered error-related negativity is thought to reflect, at least in part, altered activity in the dACC. As such, more work is needed to establish what role dACC alterations may play in experienced or anticipated hedonic processing deficits associated with depressive pathology.

Schizophrenia

Reward Anticipation There is a mixed self-report literature on anticipated pleasure in schizophrenia, with some studies suggesting impairments (e.g., Moran, Culbreth, Kandala, & Barch, in submission; Mote, Minzenberg, Carter, & Kring, 2014) and others not (e.g., Tremeau, Antonius, Nolan, Butler, & Javitt, 2014). There are few behavioral studies in schizophrenia that directly measure reward anticipation/prediction, though one such study did find evidence for reduced anticipation (Heerey & Gold, 2007). Much of the focus instead has been on neuroimaging studies. The majority of studies have reported reduced ventral striatum activity to cues predicting reward in schizophrenia (Radua et al., 2015). These results have been found in unmedicated individuals with schizophrenia (e.g., Nielsen, Rostrup, Wulff, Bak, Lublin, et al., 2012; Nielsen, Rostrup, Broberg, Wulff, & Glenthoj, 2018) and medicated individuals (e.g., Moran et al., in submission; Subramaniam et al., 2015). These deficits may not be present in individuals taking atypical medication (Juckel et al., 2006) nor in prodromal individuals (Juckel et al., 2012), though some of these results are in small samples and need replication. Other work has noted reduced ventral striatal responses to anticipation cues in antipsychotic-naïve schizophrenia patients, which improved following atypical antipsychotic treatment (Nielsen, Rostrup, Wulff, Bak, Broberg, et al., 2012; Nielsen, Rostrup, Wulff, Bak, Lublin, et al., 2012). Several studies also showed a relationship between negative symptom

severity and deficits in anticipatory ventral striatal activity (e.g., Dowd & Barch, 2012; Kluge et al., 2018; Stepien et al., 2018; Waltz et al., 2010).

Reward Prediction Error A number of studies have now shown altered prediction error responses in schizophrenia (Chase, Loriemi, Wensing, Eickhoff, & Nickl-Jockschat, 2018; Culbreth, Westbrook, Xu, Barch, & Waltz, 2016), both in terms of reductions in responses to unpredicted rewards and larger than expected responses to predicted rewards (e.g., Reinen et al., 2016; Schlagenhauf et al., 2014). Importantly, Insel and colleagues found that individuals with chronic schizophrenia taking higher doses of medication showed smaller prediction error responses (Insel et al., 2014). In contrast, other studies have found intact prediction error responses in the striatum among medicated individuals (Culbreth, Westbrook, Xu, et al., 2016; Dowd, Frank, Collins, Gold, & Barch, 2016; Waltz et al., 2009), including treatment-resistant individuals (Culbreth, Westbrook, Xu, et al., 2016) and even evidence for increased prediction error responses (White, Kraguljac, Reid, & Lahti, 2015). However, the fact that reduced prediction error responses have also been seen in unmedicated individuals (Schlagenhauf et al., 2014) argues against such abnormalities resulting only from medication effects in schizophrenia and suggests that further work is needed to understand under what conditions prediction error responses are intact versus not in schizophrenia.

Reinforcement Learning Intriguingly, several behavioral studies have suggested that reinforcement learning is intact in schizophrenia when learning is fairly implicit (e.g., Bansal et al., 2018; Barch et al., 2017; Heerey, Bell-Warren, & Gold, 2008; Somlai, Moustafa, Keri, Myers, & Gluck, 2011), (though see Siegert, Weatherall, & Bell, 2008; Taylor et al., 2018 for differing results). Similarly, several studies using the weather prediction task have shown a relatively intact learning rate but impaired asymptotic performance, which provides mixed evidence for striatal learning impairments (e.g., Keri, Nagy, Kelemen, Myers, & Gluck, 2005). When the reinforcement learning paradigms become more difficult and require the explicit use of representations about stimulus-reward contingencies, individuals with schizophrenia show more consistent evidence of impaired reinforcement learning (e.g., Culbreth, Gold, Cools, & Barch, 2016; Gold, Waltz, et al., 2012) (Barch et al., 2017; Hartmann-Riemer et al., 2017; Hernaus, Gold, Waltz, & Frank, 2018; Morris, Cyrzon, Green, Le Pelley, & Balleine, 2018; Vanes, Mouchlianitis, Collier, Averbeck, & Shergill, 2018). Interestingly, these impairments may be greater when individuals with schizophrenia must learn from reward versus from punishment (e.g., Gold, Waltz, et al., 2012; Reinen et al., 2014), though some studies also find impaired learning from punishment (e.g., Fervaha, Agid, Foussias, & Remington, 2013). Further, there is work suggesting that working memory impairments may make a significant contribution to reinforcement learning deficits in schizophrenia (Collins, Albrecht, Waltz, Gold, & Frank, 2017; Collins, Brown, Gold, Waltz, & Frank, 2014). In addition, there is a literature reporting altered activity in cortical regions involved in cognitive control during anticipation/prediction error (e.g., Gilleen et al., 2014) and during reinforcement learning (e.g., Culbreth, Gold, et al., 2016; Dowd et al., 2016; Waltz et al., 2013). These results are consistent with the literature documenting altered cognitive control

function in schizophrenia and with the growing literature suggesting important interactions between what have been referred to as "model-free" learning systems (e.g., DA in the striatum) and "model-based" learning systems that engage prefrontal and parietal systems that support representations of action-outcome models (e.g., Otto, Skatova, Madlon-Kay, & Daw, 2015), with evidence for impaired model-based learning in schizophrenia (Culbreth, Westbrook, Daw, Botvinick, & Barch, 2016). These data point to the need to examine interactions between these control systems and DA-mediated reinforcement learning systems.

Summary of Reward Anticipation, Prediction, and Reinforcement Learning in Depression Compared to Schizophrenia

As described above, there are at least two pathways to impaired reinforcement learning—altered striatal-mediated stimulus-response learning and the use of cognitive control processes to develop and maintain explicit representations of action-outcome contingencies that can guide behavior. Our hypothesis is that the former is more impaired in depression and the latter more impaired in schizophrenia. The depression literature provides evidence for reductions in reward anticipation and striatal prediction error responses, as well as evidence for impairments in "implicit" reinforcement learning on tasks that are thought to reflect slow striatally mediated reinforcement learning. This is consistent with the hypothesis that in depression, motivational impairments may originate in dysfunctional responses to rewards, which may in turn propagate forward to impair other components of reward processing. In contrast, the work in schizophrenia suggests relatively intact learning on simple reinforcement learning paradigms that may be relatively implicit in nature. These findings with more implicit learning tasks in schizophrenia are in strong contrast to the evidence for impaired performance on more difficult tasks that also engage explicit learning. This raises the question of whether these reinforcement learning impairments seen in schizophrenia result from alterations in striatally mediated implicit learning mechanisms versus more cortically mediated explicit learning mechanisms. The neuroimaging literature indicates reduced ventral striatal reward anticipation responses in unmedicated and typically medicated individuals with schizophrenia (with mixed evidence in those taking atypical antipsychotics) and evidence for reduced positive prediction errors, at least among unmedicated individuals. A number of studies have also found altered activation in frontal regions during reward anticipation or reinforcement learning.

Reward Valuation

Many different paradigms can be interpreted in the context of value representations (Gold, Waltz, et al., 2012), and there is increasingly sophisticated work using computational approaches to evaluate valuation of rewards and incentives.

In addition, there are two paradigms that have been in the field for many years and are frequently used as probes of lateral and medial OFC function: probabilistic reversal learning and the Iowa Gambling Task (in which participants have to learn to choose an optimal deck of cards, with decks that vary in the magnitude of win and loss feedback) (Bechara, Damasio, Damasio, & Anderson, 1994). Both paradigms require individuals to integrate information about rewards and punishments across trials and to update value representations appropriately.

Depression

In depression, some studies have found impaired performance on the Iowa Gambling Task (e.g., Hegedus et al., 2018; Must, Horvath, Nemeth, & Janka, 2013), though others have not (e.g., Deisenhammer, Schmid, Kemmler, Moser, & Delazer, 2018; Wang et al., 2008). There is also evidence of impairments in reversal learning in depression (e.g., Hall, Milne, & Macqueen, 2014). There is no evidence directly linking such impairments to OFC, and the sparse imaging literature on reversal learning in depression points to altered striatal responses associated with impaired reversal learning (e.g., Hall et al., 2014).

Schizophrenia

The findings on reward valuation in schizophrenia are similar to depression overall. Individuals with schizophrenia often do poorly on Iowa Gambling Task (e.g., Brown et al., 2015; Kim, Kang, & Lim, 2016; Kim, Lee, & Lee, 2009; Nestor et al., 2014; Zhang et al., 2015), though not always (e.g., Turnbull, Evans, Kemish, Park, & Bowman, 2006). Several studies also suggest impaired reversal learning in schizophrenia (e.g., Reddy, Waltz, Green, Wynn, & Horan, 2016; Waltz & Gold, 2007), though some studies have not (e.g., Jazbec et al., 2007), particularly when the difficulty of initial learning and reversal is made more similar (MacDonald et al., in submission). The imaging studies on reversal learning in schizophrenia also do not point to altered activation of the OFC in relation to these deficits, instead indicating either alterations in striatal prediction error responses (Schlagenhauf et al., 2014), deactivation of default-mode regions (Waltz et al., 2013), or impaired activation of cognitive control networks (Culbreth, Gold, et al., 2016). Thus, while there may be impairments in value computations in schizophrenia, there is little direct evidence that they reflect OFC dysfunction. However, there is evidence for altered value-based responding on other paradigms (Albrecht, Waltz, Frank, & Gold, 2016; Hernaus et al., 2018; Martinelli, Rigoli, Dolan, & Shergill, 2018; Strauss, Visser, Keller, Gold, & Buchanan, 2018; Waltz & Gold, 2016), though there is also evidence for intact learning of values (Collins, Albrecht, et al., 2017; Collins, Ciullo, et al., 2017).

Effort Cost Decision-Making (Effort Valuation)

The last 10 years have seen a burgeoning of research on effort allocation in both human and animal work. There is good evidence that DA plays a key role in regulating physical effort allocation, in that blockade of DA, especially in the accumbens, reduces physical effort allocation (e.g., Salamone, Correa, Nunes, Randall, & Pardo, 2012) and increased D2 receptor expression in the nucleus accumbens of adult mice increases physical effort expenditure (Trifilieff et al., 2013). In humans, increased DA release in response to d-amphetamine in the left striatum and the left ventromedial PFC was associated with increased willingness to expend physical effort (Treadway, Buckholtz, et al., 2012) and administration of DA-enhancing medications increases willingness to exert effort in Parkinson's disease (Chong et al., 2015).

There is also evidence for a role for the medial PFC in modulating effort allocation. Computational work has argued for a role for dACC in computing the expected value of control (Shenhav et al., 2013; Shenhav, Cohen, & Botvinick, 2016; Vassena, Holroyd, & Alexander, 2017), suggesting that the dACC integrates information about the expected value of the outcome, the expected cognitive control needed to obtain that outcome, and the expected cost of that cognitive control, in order to make decisions about the utility of expending effort. This hypothesized function of the dACC is consistent with the rodent and primate literature showing that lesions/inactivation of the dorsal ACC reduced both physical and mental effort allocation (e.g., Croxson, Walton, Boorman, Rushworth, & Bannerman, 2014; Hosking et al., 2015; Walton, Bannerman, Alterescu, & Rushworth, 2003), with evidence that rodent ACC neurons encode cost-benefit computations (e.g., Hillman & Bilkey, 2012) and with the human literature showing activation of the dACC during effort-based decision-making (Croxson et al., 2009; Prevost et al., 2010).

Depression

All of the studies to date on effort allocation in depression have focused on physical effort. Individuals with current syndromal and subsyndromal depression show reduced effort allocation as a function of increasing monetary incentives (e.g., Hershenberg et al., 2016; Treadway, Bossaller, Shelton, & Zald, 2012; Yang et al., 2014). There is also some evidence that individual differences in self-reported anticipatory and consummatory pleasure relate to individual differences in the severity of effort-allocation impairments (Yang et al., 2014). Individuals with remitted depression do not show effects as a group, though they do still show these individual difference relationships (Yang et al., 2014). In a novel study using viewing of humorous cartoons as the incentive, Sherdell, Waugh, and Gotlib (2012) did not find group differences in effort allocation, though they did find that those individuals with major depression who self-reported increased anticipatory anhedonia did show reduced effort allocation (Sherdell et al., 2012). There is a need for neuroimaging studies of ECDM in depression.

Schizophrenia

The paradigms focused on physical effort as measured by finger tapping have found relatively consistent evidence for impairment in schizophrenia (e.g., Barch et al., 2014; Gold et al., 2013; Huang et al., 2016; McCarthy, Treadway, Bennett, & Blanchard, 2016; Moran, Culbreth, & Barch, 2017; Reddy et al., 2015; Serper, Payne, Dill, Portillo, & Taliercio, 2017; Treadway, Peterman, Zald, & Park, 2015). In addition, the majority of the studies found that the degree of reduction in effort allocation was associated with either negative symptoms (e.g., Gold et al., 2013; Moran et al., 2017; Strauss et al., 2016; Treadway et al., 2015; Wang et al., 2015) or functional status (Barch et al., 2014), though one recent study found the opposite (McCarthy et al., 2016). Two studies using grip strength showed differing results—one found a significant reduction in effort allocation among individuals with schizophrenia rated clinically as having higher apathy (Hartmann et al., 2014), while the other study found no effects of either diagnosis or symptom severity (Docx et al., 2015). Several recent studies have also examined cognitive effort allocation. One study using a progressive ratio task found evidence for reduced effort allocation in schizophrenia, although the design of the task was such that cognitive effort was confounded with physical effort (Wolf et al., 2014) and another found a correlation between negative symptoms and progressive ratio (Strauss et al., 2016). In contrast, Gold et al. found little evidence of reduced cognitive effort in schizophrenia across three studies, though these studies did suggest that individuals with schizophrenia had difficulty detecting variations in cognitive effort among conditions (Gold et al., 2014). However, more recent work using both the same paradigm (Reddy et al., 2015) and a different paradigm that assesses discounting as a function of effort (Culbreth, Westbrook, & Barch, 2016) did provide evidence for impaired cognitive effort allocation in schizophrenia.

Only a few studies have examined the neural correlates of aberrant effort-cost decision-making in schizophrenia. Huang and colleagues (Huang et al., 2016) instructed individuals with schizophrenia and healthy controls to complete a button-pressing task during neuroimaging, finding greater BOLD activation in the ventral striatum during effort-based choice was associated with greater willingness to exert effort across both individuals with schizophrenia and controls. Further, people with schizophrenia showed reduced BOLD activation in the ventral striatum, the posterior cingulate gyrus, and the left medial frontal gyrus as a function of reward value and reward probability compared to healthy controls (Huang et al., 2016). Wolf et al. (2014) found that increased BOLD activation of the ventral striatum and the dorsolateral prefrontal cortex during reward processing was significantly related to increased willingness to exert effort on a behavioral task among individuals with schizophrenia. More recently, Park and colleagues showed somewhat surprisingly *greater* activation of the caudate for individuals with schizophrenia compared to healthy controls as a function of effort. However, this task did not include a choice but rather required individuals to perform either a hard or an easy option. As such, it is not clear Park and colleagues' (2017) findings relate to the larger effort-based decision-making literature that is based on choice of whether or not to allocate effort. Based on this small number of studies, the literature suggests potential

contributions to ECDM deficits in schizophrenia from the ventral striatum, cingulate gyrus, and the dorsolateral PFC, though clearly more work is needed in this domain.

Summary of Effort Cost Decision-Making in Depression Compared to Schizophrenia The literature on effort allocation is larger in schizophrenia than in depression, but to date both literatures provide robust evidence of reduced physical effort allocation in both schizophrenia and depression. There is some evidence of a similar deficit in cognitive effort allocation in schizophrenia, but to date there is no work on cognitive effort allocation in depression, an area in need of research. There is no neuroimaging literature yet on ECDM in depression, but the small literature in schizophrenia suggests a combination of contributions from the striatum, dACC, and dorsolateral PFC.

Cognitive Control and Goal-Directed Action

Depression

Individuals with depression can show cognitive control deficits (Ahern & Semkovska, 2017; McDermott & Ebmeier, 2009; Rock, Roiser, Riedel, & Blackwell, 2014), though less severe than typically seen in psychosis and it is less clear to what degree these deficits may be state-dependent and/or related to factors such as slowed processing speed. There is some meta-analytic evidence for structural alterations in brain regions often associated with cognitive control in depression, such as the dorsolateral prefrontal cortex (DLPFC) (Wise et al., 2017). There is also evidence for altered prefrontal activity in depression during emotion regulation paradigms, though with variation in the pattern across studies (e.g., Johnstone, van Reekum, Urry, Kalin, & Davidson, 2007; Sheline et al., 2009). Two studies looked at incentive-modulated cognitive control in adolescent depression, both of which found intact effects of incentives on reducing anti-saccade errors among depressed adolescents but reduced effects of incentives on latencies (e.g., Hardin, Schroth, Pine, & Ernst, 2007; Jazbec, McClure, Hardin, Pine, & Ernst, 2005). Such reduced effects would be expected if reward were experienced as less hedonically pleasurable for depressed individuals (i.e., bottom-up hedonic/"liking" deficits feeding forward to produce other deficits), but further work is needed to understand the degree to which cognitive control impairments might also contribute to altered goal-direction action in depression.

Schizophrenia

Numerous reviews have outlined the evidence for impairments in goal representation and cognitive control in schizophrenia (Barch & Ceaser, 2012; Lesh et al., 2013), as well as the evidence for altered activation, connectivity, and structure of brain regions such as the DLPFC (e.g., Bora, Fornito, Yucel, & Pantelis, 2012;

Minzenberg, Laird, Thelen, Carter, & Glahn, 2009; Ragland et al., 2009; Zhang, Picchioni, Allen, & Toulopoulou, 2016). Notably, several studies suggest that individuals with schizophrenia are not able to improve their performance on cognitive tasks when offered monetary incentives (e.g., Green, Satz, Ganzell, & Vaclav, 1992; Rassovsky, Green, Nuechterlein, Breitmeyer, & Mintz, 2005).

A recent study examined whether or not individuals with schizophrenia could improve cognitive control on a response inhibition task. Patients were able to speed their responses when presented with specific cues about winning reward and to a certain extent could speed their responses on trials in the reward "context" even when they could not earn money, an effect thought to reflect the maintenance of reward information through proactive control mechanisms. However, the individuals with schizophrenia showed a significantly smaller incentive context effect than controls (Mann, Footer, Chung, Driscoll, & Barch, 2013). There is now a published fMRI study examining whether incentives modulate DLPFC activity during a cognitive control task in schizophrenia. This study did not find any behavioral differences and found a pattern of increased sustained DLPFC activity during reward blocks in individuals with schizophrenia as a group, combined with blunted sustained activation during reward blocks in the putamen. However, individual differences in anhedonia symptom severity were significantly associated with reduced sustained DLPFC activation in the same region that showed overall increased activity as a function of reward (Chung & Barch, 2016).

Summary of Cognitive Control in Depression Compared to Schizophrenia

As described there is evidence for impairments in cognitive control in both depression and schizophrenia, though the literature in schizophrenia is far more extensive than that in depression. However, there is little evidence in depression about the relationships between cognitive control impairments and difficulties with any component of motivated behavior. In contrast, in schizophrenia there is a modest literature suggesting a link between impaired cognitive control and altered responsivity to rewards, though more work in this domain is clearly needed for both depression and schizophrenia.

Summary and Future Directions

The literature on the various components that form the pathway from reward to action provides evidence for both shared and differential impairments across depression and schizophrenia. In terms of common impairments, the literature provides evidence for both behavioral and neuroimaging indicators of reward prediction and prediction errors, as well as strong evidence for impairments in effort allocation. Importantly, however, there were also clear differences in the patterns of deficits across other components that suggest differential etiological pathways leading to

impairments in motivated behavior associated with depression versus schizophrenia. Specifically, as reviewed here, the prior literature provides evidence that (1) self-report, physiological, and neural (i.e., striatum and anterior insula) indicators of responsiveness to pleasurable stimuli and rewards are reduced among individuals with mood pathology or at risk for mood pathology, especially those who self-report higher levels of anhedonia (Ait Oumeziane & Foti, 2016; Bress et al., 2012, 2013; Bylsma, Morris, & Rottenberg, 2008; Foti et al., 2011, 2014; Foti & Hajcak, 2009; Kujawa et al., 2014; Luking et al., 2016a; Luking, Pagliaccio, Luby, & Barch, 2016b; Satterthwaite et al., 2015; Zhang et al., 2013); (2) individuals with mood pathology show impairments in reward receipt that are as robust as those found for reward anticipation (Zhang et al., 2013); (3) individuals with or at risk for mood pathology show impaired implicit reinforcement learning thought to be dependent on striatal function (Henriques, Glowacki, & Davidson, 1994; Herzallah et al., 2010, 2013; Liverant et al., 2014; Luking et al., 2017; Morkl et al., 2016; Pechtel et al., 2013; Pizzagalli et al., 2008; Vrieze et al., 2013); (4) evidence of intact cognitive control functions related to performance monitoring in the same depressed individuals who show impaired response to reward feedback (Bakic et al., 2016); and (5) some evidence of a relationship between caudate activity and ECDM deficits in depression (Yang et al., 2016).

In contrast, the prior literature provides evidence consistent with our hypotheses that different mechanisms may be contributing to ECDM deficits in schizophrenia compared to depression, including evidence that (1) self-report, physiological, and neural indicators of responsiveness to pleasurable stimuli and rewards are relatively intact in schizophrenia (Barch, Pagliaccio, & Luking, 2015; Kring & Barch, 2014; Mucci et al., 2015; Radua et al., 2015; Subramaniam et al., 2015; Wolf et al., 2014); (2) individuals with schizophrenia show stronger impairments in striatal activity related to reward anticipation than to reward receipt (i.e., intact reward responsiveness) (Barch et al., 2017, #33308; Dowd et al., 2016; Radua et al., 2015; Subramaniam et al., 2015); (3) individuals with schizophrenia show intact implicit reinforcement learning thought to be dependent on the striatum (Barch, Carter, et al., 2017; Heerey et al., 2008) but impaired explicit reinforcement learning thought to engage DLPFC and dACC (Barch et al., 2017; Culbreth, Gold, et al., 2016; Dowd et al., 2016; Gold, Waltz, et al., 2012; Waltz, Frank, Robinson, & Gold, 2007); (4) explicit reinforcement learning deficits in schizophrenia are more strongly related to dysfunction of DLPFC and dACC than VS activity (Culbreth, Gold, et al., 2016; Dowd et al., 2016; Walter et al., 2010); (5) there are consistent impairments in cognitive control and internal representation of reward (Barch & Ceaser, 2012; Henderson et al., 2012; Mann et al., 2013) in schizophrenia, related to dysfunction in DLPFC and dACC (Lesh et al., 2013; Lesh, Niendam, Minzenberg, & Carter, 2011; Minzenberg et al., 2009); and (6) robust relationships of cognitive control deficits (Gold, Barch, et al., 2012), explicit reinforcement learning impairments (Barch et al., 2017; Gold, Waltz, et al., 2012), and DLPFC dysfunction (Chung & Barch, 2016; Dowd et al., 2016) to impairments in motivation and function in schizophrenia (Gold, Barch, et al., 2012), but no such relationship for implicit reinforcement learning (Heerey et al., 2008).

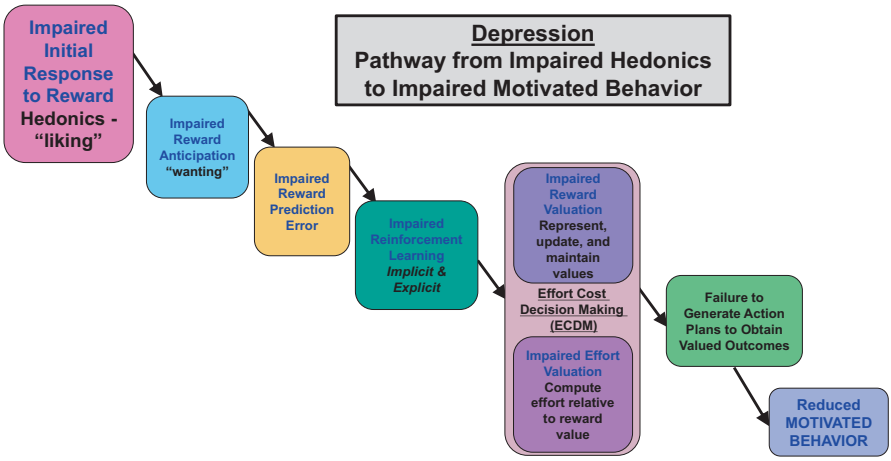

Fig. 5.2 Potential pathway to motivational impairments in depression. *ACC* anterior cingulate cortex, *BG* basal ganglia, *DA* dopamine, *DLPFC* dorsolateral prefrontal cortex, *OFC* orbital frontal cortex. Modified with permission

When integrated, as shown in Fig. 5.2, these patterns are consistent with the hypothesis that impaired motivated behavior in depression may be more related to deficits in hedonic experience and initial responsiveness to reward that propagate forward to result in impaired goal-directed and motivated behavior. As reviewed above, the animal literature suggests that hedonic responsivity is associated with opioid and GABA-ergic function in the striatum. Intriguingly, there is some literature on opioid mechanisms in depression (Lalanne, Ayranci, Kieffer, & Lutz, 2014; Murphy, 2015) and an emerging interest in modulation of the kappa opioid system as a treatment for depression (Callaghan et al., 2018; Connolly & Thase, 2012; Miller et al., 2018), with a specific focus on anhedonia. Thus, this pattern of impairments is consistent with the hypothesis that altered opioid function may contribute to hedonic impairment in depression. Nonetheless, the pattern of findings in depression could also indicate altered DA function in the striatum (Cannon et al., 2009), though the literature on DA alterations is mixed and relatively small (Camardese et al., 2014; Savitz & Drevets, 2013).

In contrast, as shown in Fig. 5.3, schizophrenia may be more related to impaired goal representation and utilization mechanisms rather than to deficits in hedonic experience or initial responsiveness to reward (Barch et al., 2018; Barch, Carter, et al., 2017; Kring & Barch, 2014). Such impairments may reflect either or both altered DA function and altered activation of dorsal frontal-parietal cognitive control systems. Recent meta-analyses point to robust evidence for increased DA synthesis availability and some evidence for D2 receptor overexpression (Fusar-Poli & Meyer-Lindenberg, 2013; Howes et al., 2012), as well as replicable evidence for altered activity of cognitive control systems (Minzenberg et al., 2009). Such a framework would suggest that even though individuals with schizophrenia can

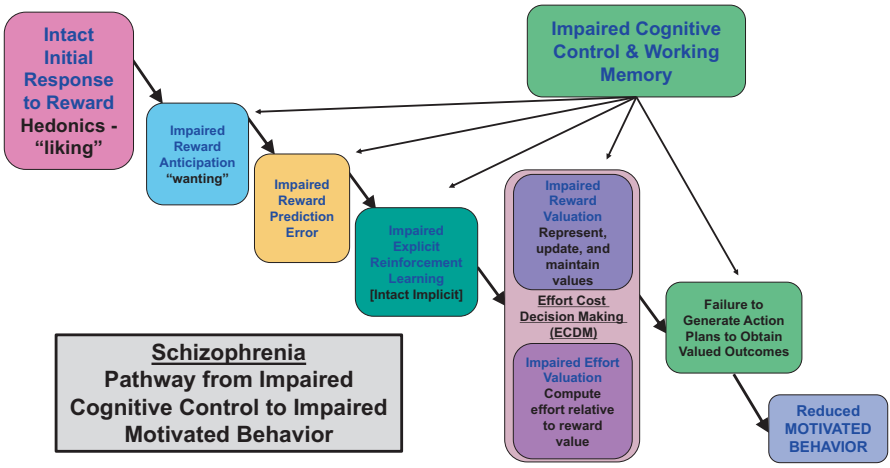

Fig. 5.3 Potential pathway to motivational impairments in schizophrenia. *ACC* anterior cingulate cortex, *BG* basal ganglia, *DA* dopamine, *DLPFC* dorsolateral prefrontal cortex, *OFC* orbital frontal cortex. Modified with permission

experience reward and pleasure from a variety of stimuli, they may have difficulty learning appropriate reward or salience representations (Howes & Kapur, 2009) and difficulty representing and maintaining such incentive information over time so that this information can drive further action selection and motivated behavior (Barch & Dowd, 2010; Kring & Barch, 2014).

The following scenarios may illustrate how such differing patterns of impairments may lead to altered motivated behavior. An individual with depression may report that they do not find social interactions enjoyable, whether or not someone else organizes such interactions for them. Such impairments in initial responsiveness to incentives might reflect altered opioid signaling in the striatum, though this is highly speculative. As a consequence, individuals with depression (at least those with impaired hedonic capacity) may not learn about cues associated with positive social interactions and/or anticipate pleasure associated with such interactions and, thus, may not allocate effort to making social plans since they do not anticipate them being particularly enjoyable. In contrast, an individual with schizophrenia may report that they get pleasure from interacting with friends and enjoy such interactions if someone else arranges them and provides transportation (i.e., intact hedonics). However, they may have difficulty engaging in the behaviors necessary to organize such social interactions on their own (Barch & Dowd, 2010; Kring & Moran, 2008). The need to determine who to call or invite, figuring out where to meet, and determining what activities one might do necessitates maintenance of information that links associations of the social interaction's rewarding properties to the allocation of effort to organize such experiences. These functions may require the ability to maintain incentive cues or context over time—a process that may be reliant on cognitive control and working memory mechanisms, which are

compromised in schizophrenia (Barch & Dowd, 2010). Thus, they may not exert effort to organize social interactions because they have difficulties using representations about future reward to drive behavior, but not because they would not find those social interactions enjoyable.

These are simplistic examples and do not fully capture the many facets of context and individual traits that drive the complexity of motivated behaviors that we need to engage in every day nor do they fully capture the many interactions among these components and systems. Further, our hypothesis does not clarify why there are individual differences among people with the same putative diagnosis in the degree of motivational impairments or why the severity of such deficits may vary over time. However, this hypothesis does provide one framework for organizing research on mechanisms of motivational impairment in psychopathology, allowing us to begin to determine which components reflect truly transdiagnostic impairments, versus those that may be more unique or selective to particular forms of psychopathology. The field now needs to take up the challenge of addressing these open questions, moving beyond single-diagnosis studies and even beyond studies that focus on only the extreme of clinically diagnosed individuals. Instead, we need studies that examine multiple forms of psychopathology with the same methods and approaches, working with populations whose impairments span the range from relatively normative to severely clinically impaired. These studies are starting to emerge in the literature, and as their results evolve, they will help us better understand the structure of psychopathology and eventually help to drive the development of hopefully more effective and targeted interventions that enhance quality of life, reduce public health burden, and avoid the development of psychopathology in the first place.

Acknowledgments Parts of this chapter have been reprinted with permission from Oxford University Press and Cambridge University Press and come from Barch, D. M., Pagliaccio, D., & Luking, K. (2019). Positive Valence System Dysregulation in Psychosis: A comparative Analysis. In Gruber, J. (Ed). Oxford Handbook of Positive Emotion and Psychopathology and Barch, D. M., Pagliaccio, D., & Luking, K. (2018). Motivational Impairments in Psychotic and Depressive Pathology: Psychological and Neural Mechanisms. In Sangha, S., & Foti, D., Eds. *Neurobiology of Abnormal Emotion and Motivated Behaviors: Integrating Animal and Human Research*. Pages 278–304.

References

Admon, R., Kaiser, R. H., Dillon, D. G., Beltzer, M., Goer, F., Olson, D. P., … Pizzagalli, D. A. (2017). Dopaminergic enhancement of striatal response to reward in major depression. *The American Journal of Psychiatry, 174*(4), 378–386. https://doi.org/10.1176/appi.ajp.2016.16010111

Ahern, E., & Semkovska, M. (2017). Cognitive functioning in the first-episode of major depressive disorder: A systematic review and meta-analysis. *Neuropsychology, 31*(1), 52–72. https://doi.org/10.1037/neu0000319

Ait Oumeziane, B., & Foti, D. (2016). Reward-related neural dysfunction across depression and impulsivity: A dimensional approach. *Psychophysiology, 53*(8), 1174–1184. https://doi.org/10.1111/psyp.12672

Albrecht, M. A., Waltz, J. A., Frank, M. J., & Gold, J. M. (2016). Probability and magnitude evaluation in schizophrenia. *Schizophrenia Research: Cognition, 5*, 41–46. https://doi.org/10.1016/j.scog.2016.06.003

Auster, T. L., Cohen, A. S., Callaway, D. A., & Brown, L. A. (2014). Objective and subjective olfaction across the schizophrenia spectrum. *Psychiatry, 77*(1), 57–66. https://doi.org/10.1521/psyc.2014.77.1.57

Bakic, J., Pourtois, G., Jepma, M., Duprat, R., De Raedt, R., & Baeken, C. (2016). Spared internal but impaired external reward prediction error signals in major depressive disorder during reinforcement learning. *Depression and Anxiety*. https://doi.org/10.1002/da.22576

Bansal, S., Robinson, B. M., Geng, J. J., Leonard, C. J., Hahn, B., Luck, S. J., & Gold, J. M. (2018). The impact of reward on attention in schizophrenia. *Schizophrenia Research: Cognition, 12*, 66–73. https://doi.org/10.1016/j.scog.2018.05.001

Barch, D. M., Carter, C. S., Gold, J. M., Johnson, S. L., Kring, A. M., MacDonald, A. W., … Strauss, M. E. (2017). Explicit and implicit reinforcement learning across the psychosis spectrum. *Journal of Abnormal Psychology, 126*(5), 694–711. https://doi.org/10.1037/abn0000259

Barch, D. M., & Ceaser, A. E. (2012). Cognition in schizophrenia: Core psychological and neural mechanisms. *Trends in Cognitive Science, 16*, 27–34.

Barch, D. M., & Dowd, E. C. (2010). Goal representations and motivational drive in schizophrenia: the role of prefrontal-striatal interactions. *Schizophrenia Bulletin, 36*(5), 919–934. https://doi.org/10.1093/schbul/sbq068

Barch, D. M., Oquendo, M. A., Pacheco, J., & Morris, S. (2016). *Behavioral assessment methods for RDoC constructs: A report by the National Advisory Mental Health Council Workgroup on tasks and measures for RDoC*. Washington, DC: National Institutes of Mental Health.

Barch, D. M., Pagliaccio, D., & Luking, K. (2015). Mechanisms underlying motivational deficits in psychopathology: Similarities and differences in depression and schizophrenia. *Current Topics in Behavioral Neurosciences*. https://doi.org/10.1007/7854_2015_376

Barch, D. M., Pagliaccio, D., & Luking, K. (2018). Motivational impairments in psychotic and depressive pathology: Psychological and neural mechanisms. In S. Sangha & D. Foti (Eds.), *Neurobiology of abnormal emotion and motivated behaviors: Integrating animal and human research* (pp. 278–304). London, England: Academic Press.

Barch, D. M., Pagliaccio, D., & Luking, K. (2019). Positive valence system dysregulation in psychosis: A comparative analysis. In J. Gruber (Ed.), *Handbook of positive emotion and psychopathology* (pp. 253–283). London, England: Oxford University Press.

Barch, D. M., Treadway, M. T., & Schoen, N. (2014). Effort, anhedonia, and function in schizophrenia: Reduced effort allocation predicts amotivation and functional impairment. *Journal of Abnormal Psychology, 123*(2), 387–397. https://doi.org/10.1037/a0036299

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, H. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50*, 7–15.

Belden, A. C., Irvin, K., Hajcak, G., Kappenman, E. S., Kelly, D., Karlow, S., … Barch, D. M. (2016). Neural correlates of reward processing in depressed and healthy preschool-age children. *Journal of the American Academy of Child and Adolescent Psychiatry, 55*(12), 1081–1089. https://doi.org/10.1016/j.jaac.2016.09.503

Berlin, I., Givry-Steiner, L., Lecrubier, Y., & Puech, A. J. (1998). Measures of anhedonia and hedonic responses to sucrose in depressive and schizophrenic patients in comparison with healthy subjects. *European Psychiatry, 13*(6), 303–309. https://doi.org/10.1016/S0924-9338(98)80048-5

Berridge, K. C. (2004). Motivation concepts in behavioral neuroscience. *Physiology & Behavior, 81*(2), 179–209.

Berridge, K. C. (2012). From prediction error to incentive salience: Mesolimbic computation of reward motivation. *The European Journal of Neuroscience, 35*(7), 1124–1143. https://doi.org/10.1111/j.1460-9568.2012.07990.x

Berridge, K. C., & Kringelbach, M. L. (2008). Affective neuroscience of pleasure: Reward in humans and animals. *Psychopharmacology, 199*(3), 457–480. https://doi.org/10.1007/s00213-008-1099-6

Berridge, K. C., & Kringelbach, M. L. (2015). Pleasure systems in the brain. *Neuron, 86*(3), 646–664. https://doi.org/10.1016/j.neuron.2015.02.018

Berridge, K. C., Robinson, T. E., & Aldridge, J. W. (2009). Dissecting components of reward: 'Liking', 'wanting', and learning. *Current Opinion in Pharmacology, 9*(1), 65–73. https://doi.org/10.1016/j.coph.2008.12.014

Bora, E., Fornito, A., Yucel, M., & Pantelis, C. (2012). The effects of gender on grey matter abnormalities in major psychoses: A comparative voxelwise meta-analysis of schizophrenia and bipolar disorder. *Psychological Medicine, 42*(2), 295–307. https://doi.org/10.1017/S0033291711001450

Botvinick, M. M., Huffstetler, S., & McGuire, J. T. (2009). Effort discounting in human nucleus accumbens. *Cognitive, Affective, & Behavioral Neuroscience, 9*(1), 16–27.

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences, 16*(2), 106–113. https://doi.org/10.1016/j.tics.2011.12.010

Braver, T. S., & Cohen, J. D. (1999). Dopamine, cognitive control, and schizophrenia: The gating model. *Progress in Brain Research, 121*, 327–349.

Braver, T. S., Krug, M. K., Chiew, K. S., Kool, W., Westbrook, J. A., Clement, N. J., … Momcai Group. (2014). Mechanisms of motivation-cognition interaction: Challenges and opportunities. *Cognitive, Affective, & Behavioral Neuroscience, 14*(2), 443–472. https://doi.org/10.3758/s13415-014-0300-0

Bress, J. N., Foti, D., Kotov, R., Klein, D. N., & Hajcak, G. (2013). Blunted neural response to rewards prospectively predicts depression in adolescent girls. *Psychophysiology, 50*(1), 74–81. https://doi.org/10.1111/j.1469-8986.2012.01485.x

Bress, J. N., Smith, E., Foti, D., Klein, D. N., & Hajcak, G. (2012). Neural response to reward and depressive symptoms in late childhood to early adolescence. *Biological Psychology, 89*(1), 156–162. https://doi.org/10.1016/j.biopsycho.2011.10.004

Brown, E. C., Hack, S. M., Gold, J. M., Carpenter, W. T., Jr., Fischer, B. A., Prentice, K. P., & Waltz, J. A. (2015). Integrating frequency and magnitude information in decision-making in schizophrenia: An account of patient performance on the Iowa Gambling Task. *Journal of Psychiatric Research, 66–67*, 16–23. https://doi.org/10.1016/j.jpsychires.2015.04.007

Burkhouse, K. L., Gorka, S. M., Afshar, K., & Phan, K. L. (2017). Neural reactivity to reward and internalizing symptom dimensions. *Journal of Affective Disorders, 217*, 73–79. https://doi.org/10.1016/j.jad.2017.03.061

Burkhouse, K. L., Kujawa, A., Kennedy, A. E., Shankman, S. A., Langenecker, S. A., Phan, K. L., & Klumpp, H. (2016). Neural reactivity to reward as a predictor of cognitive behavioral therapy response in anxiety and depression. *Depression and Anxiety, 33*(4), 281–288. https://doi.org/10.1002/da.22482

Bylsma, L. M., Morris, B. H., & Rottenberg, J. (2008). A meta-analysis of emotional reactivity in major depressive disorder. *Clinical Psychology Review, 28*(4), 676–691. https://doi.org/10.1016/j.cpr.2007.10.001

Callaghan, C. K., Rouine, J., Dean, R. L., Knapp, B. I., Bidlack, J. M., Deaver, D. R., & O'Mara, S. M. (2018). Antidepressant-like effects of 3-carboxamido seco-nalmefene (3CS-nalmefene), a novel opioid receptor modulator, in a rat IFN-alpha-induced depression model. *Brain, Behavior, and Immunity, 67*, 152–162. https://doi.org/10.1016/j.bbi.2017.08.016

Camardese, G., Di Giuda, D., Di Nicola, M., Cocciolillo, F., Giordano, A., Janiri, L., & Guglielmo, R. (2014). Imaging studies on dopamine transporter and depression: A review of literature and suggestions for future research. *Journal of Psychiatric Research, 51*, 7–18. https://doi.org/10.1016/j.jpsychires.2013.12.006

Cannon, D. M., Klaver, J. M., Peck, S. A., Rallis-Voak, D., Erickson, K., & Drevets, W. C. (2009). Dopamine type-1 receptor binding in major depressive disorder assessed using positron emission tomography and [11C]NNC-112. *Neuropsychopharmacology, 34*(5), 1277–1287. https://doi.org/10.1038/npp.2008.194

Carlson, J. M., Foti, D., Mujica-Parodi, L. R., Harmon-Jones, E., & Hajcak, G. (2011). Ventral striatal and medial prefrontal BOLD activation is correlated with reward-related electrocortical activity: A combined ERP and fMRI study. *NeuroImage, 57*(4), 1608–1616. https://doi.org/10.1016/j.neuroimage.2011.05.037

Chase, H. W., Loriemi, P., Wensing, T., Eickhoff, S. B., & Nickl-Jockschat, T. (2018). Meta-analytic evidence for altered mesolimbic responses to reward in schizophrenia. *Human Brain Mapping*. https://doi.org/10.1002/hbm.24049

Chase, H. W., Nusslock, R., Almeida, J. R., Forbes, E. E., LaBarbara, E. J., & Phillips, M. L. (2013). Dissociable patterns of abnormal frontal cortical activation during anticipation of an uncertain reward or loss in bipolar versus major depression. *Bipolar Disorders, 15*(8), 839–854. https://doi.org/10.1111/bdi.12132

Chentsova-Dutton, Y., & Hanley, K. (2010). The effects of anhedonia and depression on hedonic responses. *Psychiatry Research, 179*(2), 176–180. https://doi.org/10.1016/j.psychres.2009.06.013

Chong, T. T., Bonnelle, V., Manohar, S., Veromann, K. R., Muhammed, K., Tofaris, G. K., … Husain, M. (2015). Dopamine enhances willingness to exert effort for reward in Parkinson's disease. *Cortex, 69*, 40–46. https://doi.org/10.1016/j.cortex.2015.04.003

Chung, Y. S., & Barch, D. M. (2016). Frontal-striatum dysfunction during reward processing: Relationships to amotivation in schizophrenia. *Journal of Abnormal Psychology, 125*(3), 453–469. https://doi.org/10.1037/abn0000137

Clepce, M., Gossler, A., Reich, K., Kornhuber, J., & Thuerauf, N. (2010). The relation between depression, anhedonia and olfactory hedonic estimates—A pilot study in major depression. *Neuroscience Letters, 471*(3), 139–143. https://doi.org/10.1016/j.neulet.2010.01.027

Collins, A., & Frank, M. J. (2018). Within and across-trial dynamics of human EEG reveal cooperative interplay between reinforcement learning and working memory. *Proceeding of the National Academy of Sciences, 115*, 201720963.

Collins, A. G., Brown, J. K., Gold, J. M., Waltz, J. A., & Frank, M. J. (2014). Working memory contributions to reinforcement learning impairments in schizophrenia. *The Journal of Neuroscience, 34*(41), 13747–13756. https://doi.org/10.1523/JNEUROSCI.0989-14.2014

Collins, A. G., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *The European Journal of Neuroscience, 35*(7), 1024–1035. https://doi.org/10.1111/j.1460-9568.2011.07980.x

Collins, A. G. E., Albrecht, M. A., Waltz, J. A., Gold, J. M., & Frank, M. J. (2017). Interactions among working memory, reinforcement learning, and effort in value-based choice: A new paradigm and selective deficits in schizophrenia. *Biological Psychiatry, 82*(6), 431–439. https://doi.org/10.1016/j.biopsych.2017.05.017

Collins, A. G. E., Ciullo, B., Frank, M. J., & Badre, D. (2017). Working memory load strengthens reward prediction errors. *The Journal of Neuroscience, 37*(16), 4332–4342. https://doi.org/10.1523/JNEUROSCI.2700-16.2017

Conen, K. E., & Padoa-Schioppa, C. (2016). The dynamic nature of value-based decisions. *Nature Neuroscience, 19*(7), 866–867. https://doi.org/10.1038/nn.4329

Connolly, K. R., & Thase, M. E. (2012). Emerging drugs for major depressive disorder. *Expert Opinion on Emerging Drugs, 17*(1), 105–126. https://doi.org/10.1517/14728214.2012.660146

Cools, R., Clark, L., Owen, A. M., & Robbins, T. W. (2002). Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *Journal of Neuroscience, 22*(11), 4563–4567.

Cools, R., Lewis, S. J., Clark, L., Barker, R. A., & Robbins, T. W. (2007). L-DOPA disrupts activity in the nucleus accumbens during reversal learning in Parkinson's disease. *Neuropsychopharmacology, 32*(1), 180–189.

Crespo-Facorro, B., Paradiso, S., Andreasen, N. C., O'Leary, D. S., Watkins, G. L., Ponto, L. L. B., & Hichwa, R. D. (2001). Neural mechanisms of anhedonia in schizophrenia. *Journal of the American Medical Association, 286*(4), 427–435.

Croxson, P. L., Walton, M. E., Boorman, E. D., Rushworth, M. F., & Bannerman, D. M. (2014). Unilateral medial frontal cortex lesions cause a cognitive decision-making deficit in rats. *The European Journal of Neuroscience, 40*(12), 3757–3765. https://doi.org/10.1111/ejn.12751

Croxson, P. L., Walton, M. E., O'Reilly, J. X., Behrens, T. E., & Rushworth, M. F. (2009). Effort-based cost-benefit valuation and the human brain. *The Journal of Neuroscience, 29*(14), 4531–4541.

Culbreth, A., Westbrook, A., & Barch, D. (2016). Negative symptoms are associated with an increased subjective cost of cognitive effort. *Journal of Abnormal Psychology, 125*(4), 528–536. https://doi.org/10.1037/abn0000153

Culbreth, A. J., Gold, J. M., Cools, R., & Barch, D. M. (2016). Impaired activation in cognitive control regions predicts reversal learning in schizophrenia. *Schizophrenia Bulletin, 42*(2), 484–493. https://doi.org/10.1093/schbul/sbv075

Culbreth, A. J., Westbrook, A., Daw, N. D., Botvinick, M., & Barch, D. M. (2016). Reduced model-based decision-making in schizophrenia. *Journal of Abnormal Psychology, 125*(6), 777–787. https://doi.org/10.1037/abn0000164

Culbreth, A. J., Westbrook, A., Xu, Z., Barch, D. M., & Waltz, J. A. (2016). Intact ventral striatal prediction error signaling in medicated schizophrenia patients. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 1*(5), 474–483. https://doi.org/10.1016/j.bpsc.2016.07.007

Deisenhammer, E. A., Schmid, S. K., Kemmler, G., Moser, B., & Delazer, M. (2018). Decision making under risk and under ambiguity in depressed suicide attempters, depressed non-attempters and healthy controls. *Journal of Affective Disorders, 226*, 261–266. https://doi.org/10.1016/j.jad.2017.10.012

Dichter, G. S., Kozink, R. V., McClernon, F. J., & Smoski, M. J. (2012). Remitted major depression is characterized by reward network hyperactivation during reward anticipation and hypoactivation during reward outcomes. *Journal of Affective Disorders, 136*(3), 1126–1134. https://doi.org/10.1016/j.jad.2011.09.048

Dichter, G. S., Smoski, M. J., Kampov-Polevoy, A. B., Gallop, R., & Garbutt, J. C. (2010). Unipolar depression does not moderate responses to the Sweet Taste Test. *Depression and Anxiety, 27*(9), 859–863. https://doi.org/10.1002/da.20690

Docx, L., de la Asuncion, J., Sabbe, B., Hoste, L., Baeten, R., Warnaerts, N., & Morrens, M. (2015). Effort discounting and its association with negative symptoms in schizophrenia. *Cognitive Neuropsychiatry, 20*(2), 172–185. https://doi.org/10.1080/13546805.2014.993463

Dosenbach, N. U., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences, 12*, 99–105.

Dowd, E. C., & Barch, D. M. (2012). Pavlovian reward prediction and receipt in schizophrenia: Relationship to anhedonia. *PLoS One, 7*(5), e35622. https://doi.org/10.1371/journal.pone.0035622

Dowd, E. C., Frank, M. J., Collins, A., Gold, J. M., & Barch, D. M. (2016). Probabilistic reinforcement learning in patients with schizophrenia: Relationships to anhedonia and avolition. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 1*(5), 460–473. https://doi.org/10.1016/j.bpsc.2016.05.005

Fellows, L. K., & Farah, M. J. (2005). Different underlying impairments in decision-making following ventromedial and dorsolateral frontal lobe damage in humans. *Cerebral Cortex, 15*(1), 58–63.

Fervaha, G., Agid, O., Foussias, G., & Remington, G. (2013). Impairments in both reward and punishment guided reinforcement learning in schizophrenia. *Schizophrenia Research, 150*(2–3), 592–593. https://doi.org/10.1016/j.schres.2013.08.012

Fervaha, G., Graff-Guerrero, A., Zakzanis, K. K., Foussias, G., Agid, O., & Remington, G. (2013). Incentive motivation deficits in schizophrenia reflect effort computation impairments during cost-benefit decision-making. *Journal of Psychiatric Research, 47*(11), 1590–1596. https://doi.org/10.1016/j.jpsychires.2013.08.003

Fletcher, K., Parker, G., Paterson, A., Fava, M., Iosifescu, D., & Pizzagalli, D. A. (2015). Anhedonia in melancholic and non-melancholic depressive disorders. *Journal of Affective Disorders, 184*, 81–88. https://doi.org/10.1016/j.jad.2015.05.028

Foti, D., Carlson, J. M., Sauder, C. L., & Proudfit, G. H. (2014). Reward dysfunction in major depression: Multimodal neuroimaging evidence for refining the melancholic phenotype. *NeuroImage, 101*, 50–58. https://doi.org/10.1016/j.neuroimage.2014.06.058

Foti, D., & Hajcak, G. (2009). Depression and reduced sensitivity to non-rewards versus rewards: Evidence from event-related potentials. *Biological Psychology, 81*(1), 1–8. https://doi.org/10.1016/j.biopsycho.2008.12.004

Foti, D., Kotov, R., Klein, D. N., & Hajcak, G. (2011). Abnormal neural sensitivity to monetary gains versus losses among adolescents at risk for depression. *Journal of Abnormal Child Psychology, 39*(7), 913–924. https://doi.org/10.1007/s10802-011-9503-9

Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: Cognitive reinforcement learning in parkinsonism. *Science, 306*(5703), 1940–1943.

Fusar-Poli, P., & Meyer-Lindenberg, A. (2013). Striatal presynaptic dopamine in schizophrenia, part II: Meta-analysis of [(18)F/(11)C]-DOPA PET studies. *Schizophrenia Bulletin, 39*(1), 33–42. https://doi.org/10.1093/schbul/sbr180

Gilleen, J., Shergill, S. S., & Kapur, S. (2014). Impaired subjective well-being in schizophrenia is associated with reduced anterior cingulate activity during reward processing. *Psychological Medicine*, 1–12. https://doi.org/10.1017/S0033291714001718

Gold, J. M., Barch, D. M., Carter, C. S., Dakin, S., Luck, S. J., MacDonald, A. W., 3rd, … Strauss, M. (2012). Clinical, functional, and intertask correlations of measures developed by the Cognitive Neuroscience Test Reliability and Clinical Applications for Schizophrenia Consortium. *Schizophrenia Bulletin, 38*(1), 144–152. https://doi.org/10.1093/schbul/sbr142

Gold, J. M., Kool, W., Botvinick, M. M., Hubzin, L., August, S., & Waltz, J. A. (2014). Cognitive effort avoidance and detection in people with schizophrenia. *Cognitive, Affective, & Behavioral Neuroscience*. https://doi.org/10.3758/s13415-014-0308-5

Gold, J. M., Strauss, G. P., Waltz, J. A., Robinson, B. M., Brown, J. K., & Frank, M. J. (2013). Negative symptoms of schizophrenia are associated with abnormal effort-cost computations. *Biological Psychiatry*. https://doi.org/10.1016/j.biopsych.2012.12.022

Gold, J. M., Waltz, J. A., Matveeva, T. M., Kasanova, Z., Strauss, G. P., Herbener, E. S., … Frank, M. J. (2012). Negative symptoms and the failure to represent the expected reward value of actions: Behavioral and computational modeling evidence. *Archives of General Psychiatry, 69*(2), 129–138. https://doi.org/10.1001/archgenpsychiatry.2011.1269

Gorka, S. M., Burkhouse, K. L., Afshar, K., & Phan, K. L. (2017). Error-related brain activity and internalizing disorder symptom dimensions in depression and anxiety. *Depression and Anxiety, 34*(11), 985–995. https://doi.org/10.1002/da.22648

Gorka, S. M., Huggins, A. A., Fitzgerald, D. A., Nelson, B. D., Phan, K. L., & Shankman, S. A. (2014). Neural response to reward anticipation in those with depression with and without panic disorder. *Journal of Affective Disorders, 164*, 50–56. https://doi.org/10.1016/j.jad.2014.04.019

Gotlib, I. H., Sivers, H., Gabrieli, J. D., Whitfield-Gabrieli, S., Goldin, P., Minor, K. L., & Canli, T. (2005). Subgenual anterior cingulate activation to valenced emotional stimuli in major depression. *Neuroreport, 16*(16), 1731–1734.

Green, M. F., Satz, P., Ganzell, S., & Vaclav, J. F. (1992). Wisconsin Card Sorting Test performance in schizophrenia: Remediation of a stubborn deficit. *The American Journal of Psychiatry, 149*(1), 62–67.

Greenberg, T., Chase, H. W., Almeida, J. R., Stiffler, R., Zevallos, C. R., Aslam, H. A., … Phillips, M. L. (2015). Moderation of the relationship between reward expectancy and prediction error-related ventral striatal reactivity by anhedonia in unmedicated major depressive disorder: Findings from the EMBARC study. *The American Journal of Psychiatry, 172*(9), 881–891. https://doi.org/10.1176/appi.ajp.2015.14050594

Grimm, O., Vollstadt-Klein, S., Krebs, L., Zink, M., & Smolka, M. N. (2012). Reduced striatal activation during reward anticipation due to appetite-provoking cues in chronic schizophrenia: A fMRI study. *Schizophrenia Research, 134*(2–3), 151–157. https://doi.org/10.1016/j.schres.2011.11.027

Haber, S. N., & Behrens, T. E. J. (2014). The neural network underlying incentive-based learning: Implications for interpreting circuit disruptions in psychiatric disorders. *Neuron, 83*(5), 1019–1039.

Hall, G. B., Milne, A. M., & Macqueen, G. M. (2014). An fMRI study of reward circuitry in patients with minimal or extensive history of major depression. *European Archives of Psychiatry and Clinical Neuroscience, 264*(3), 187–198. https://doi.org/10.1007/s00406-013-0437-9

Hardin, M. G., Schroth, E., Pine, D. S., & Ernst, M. (2007). Incentive-related modulation of cognitive control in healthy, anxious, and depressed adolescents: Development and psychopathology related differences. *Journal of Child Psychology and Psychiatry, 48*(5), 446–454.

Hartmann, M. N., Hager, O. M., Reimann, A. V., Chumbley, J. R., Kirschner, M., Seifritz, E., … Kaiser, S. (2014). Apathy but not diminished expression in schizophrenia is associated with discounting of monetary rewards by physical effort. *Schizophrenia Bulletin*. https://doi.org/10.1093/schbul/sbu102

Hartmann-Riemer, M. N., Aschenbrenner, S., Bossert, M., Westermann, C., Seifritz, E., Tobler, P. N., … Kaiser, S. (2017). Deficits in reinforcement learning but no link to apathy in patients with schizophrenia. *Scientific Reports, 7*, 40352. https://doi.org/10.1038/srep40352

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2007). Towards an executive without a homunculus: Computational models of the prefrontal cortex/basal ganglia system. *Philosophical Transactions of Royal Society B: Biological Sciences, 362*(1485), 1601–1613.

Heerey, E. A., Bell-Warren, K. R., & Gold, J. M. (2008). Decision-making impairments in the context of intact reward sensitivity in schizophrenia. *Biological Psychiatry, 64*(1), 62–69.

Heerey, E. A., & Gold, J. M. (2007). Patients with schizophrenia demonstrate dissociation between affective experience and motivated behavior. *Journal of Abnormal Psychology, 116*(2), 268–278.

Hegedus, K. M., Szkaliczki, A., Gal, B. I., Ando, B., Janka, Z., & Almos, P. Z. (2018). Decision-making performance of depressed patients within 72 h following a suicide attempt. *Journal of Affective Disorders, 235*, 583–588. https://doi.org/10.1016/j.jad.2018.04.082

Henderson, D., Poppe, A. B., Barch, D. M., Carter, C. S., Gold, J. M., Ragland, J. D., … MacDonald, A. W., 3rd. (2012). Optimization of a goal maintenance task for use in clinical applications. *Schizophrenia Bulletin, 38*(1), 104–113. https://doi.org/10.1093/schbul/sbr172

Henriques, J. B., Glowacki, J. M., & Davidson, R. J. (1994). Reward fails to alter response bias in depression. *Journal of Abnormal Psychology, 103*, 460–466.

Hernaus, D., Gold, J. M., Waltz, J. A., & Frank, M. J. (2018). Impaired expected value computations coupled with overreliance on stimulus-response learning in schizophrenia. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. https://doi.org/10.1016/j.bpsc.2018.03.014

Hershenberg, R., Satterthwaite, T. D., Daldal, A., Katchmar, N., Moore, T. M., Kable, J. W., & Wolf, D. H. (2016). Diminished effort on a progressive ratio task in both unipolar and bipolar depression. *Journal of Affective Disorders, 196*, 97–100. https://doi.org/10.1016/j.jad.2016.02.003

Herzallah, M. M., Moustafa, A. A., Misk, A. J., Al-Dweib, L. H., Abdelrazeq, S. A., Myers, C. E., & Gluck, M. A. (2010). Depression impairs learning whereas anticholinergics impair transfer generalization in Parkinson patients tested on dopaminergic medications. *Cognitive and Behavioral Neurology, 23*(2), 98–105. https://doi.org/10.1097/WNN.0b013e3181df3048

Herzallah, M. M., Moustafa, A. A., Natsheh, J. Y., Danoun, O. A., Simon, J. R., Tayem, Y. I., … Gluck, M. A. (2013). Depression impairs learning, whereas the selective serotonin reuptake inhibitor, paroxetine, impairs generalization in patients with major depressive disorder. *Journal of Affective Disorders, 151*(2), 484–492. https://doi.org/10.1016/j.jad.2013.06.030

Hillman, K. L., & Bilkey, D. K. (2012). Neural encoding of competitive effort in the anterior cingulate cortex. *Nature Neuroscience, 15*(9), 1290–1297. https://doi.org/10.1038/nn.3187

Holroyd, C. B., & McClure, S. M. (2015). Hierarchical control over effortful behavior by rodent medial frontal cortex: A computational model. *Psychological Review, 122*(1), 54–83. https://doi.org/10.1037/a0038339

Hosking, J. G., Cocker, P. J., & Winstanley, C. A. (2015). Prefrontal cortical inactivations decrease willingness to expend cognitive effort on a rodent cost/benefit decision-making task. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhu321

Howes, O. D., Kambeitz, J., Kim, E., Stahl, D., Slifstein, M., Abi-Dargham, A., & Kapur, S. (2012). The nature of dopamine dysfunction in schizophrenia and what this means for treatment. *Archives of General Psychiatry, 69*(8), 776–786. https://doi.org/10.1001/archgenpsychiatry.2012.169

Howes, O. D., & Kapur, S. (2009). The dopamine hypothesis of schizophrenia: Version III—The final common pathway. *Schizophrenia Bulletin, 35*(3), 549–562. https://doi.org/10.1093/schbul/sbp006

Huang, J., Yang, X. H., Lan, Y., Zhu, C. Y., Liu, X. Q., Wang, Y. F., … Chan, R. C. (2016). Neural substrates of the impaired effort expenditure decision making in schizophrenia. *Neuropsychology, 30*(6), 685–696. https://doi.org/10.1037/neu0000284

Insel, C., Reinen, J., Weber, J., Wager, T. D., Jarskog, L. F., Shohamy, D., & Smith, E. E. (2014). Antipsychotic dose modulates behavioral and neural responses to feedback during reinforcement learning in schizophrenia. *Cognitive, Affective, & Behavioral Neuroscience, 14*(1), 189–201. https://doi.org/10.3758/s13415-014-0261-3

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., … Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *The American Journal of Psychiatry, 167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Jazbec, S., McClure, E., Hardin, M., Pine, D. S., & Ernst, M. (2005). Cognitive control under contingencies in anxious and depressed adolescents: An antisaccade task. *Biological Psychiatry, 58*(8), 632–639. https://doi.org/10.1016/j.biopsych.2005.04.010

Jazbec, S., Pantelis, C., Robbins, T., Weickert, T., Weinberger, D. R., & Goldberg, T. E. (2007). Intra-dimensional/extra-dimensional set-shifting performance in schizophrenia: Impact of distractors. *Schizophrenia Research, 89*(1–3), 339–349.

Johnstone, T., van Reekum, C. M., Urry, H. L., Kalin, N. H., & Davidson, R. J. (2007). Failure to regulate: Counterproductive recruitment of top-down prefrontal-subcortical circuitry in major depression. *The Journal of Neuroscience, 27*(33), 8877–8884.

Juckel, G., Friedel, E., Koslowski, M., Witthaus, H., Ozgurdal, S., Gudlowski, Y., … Schlagenhauf, F. (2012). Ventral striatal activation during reward processing in subjects with ultra-high risk for schizophrenia. *Neuropsychobiology, 66*(1), 50–56. https://doi.org/10.1159/000337130

Juckel, G., Schlagenhauf, F., Koslowski, M., Filonov, D., Wustenberg, T., Villringer, A., … Heinz, A. (2006). Dysfunction of ventral striatal reward prediction in schizophrenic patients treated with typical, not atypical, neuroleptics. *Psychopharmacology, 187*(2), 222–228.

Kamath, V., Lasutschinkow, P., Ishizuka, K., & Sawa, A. (2018). Olfactory functioning in first-episode psychosis. *Schizophrenia Bulletin, 44*(3), 672–680. https://doi.org/10.1093/schbul/sbx107

Keedwell, P. A., Andrew, C., Williams, S. C., Brammer, M. J., & Phillips, M. L. (2005). The neural correlates of anhedonia in major depressive disorder. *Biological Psychiatry, 58*(11), 843–853. https://doi.org/10.1016/j.biopsych.2005.05.019

Keren, H., O'Callaghan, G., Vidal-Ribas, P., Buzzell, G. A., Brotman, M. A., Leibenluft, E., … Stringaris, A. (2018). Reward processing in depression: A conceptual and meta-analytic review across fMRI and EEG studies. *American Journal of Psychiatry*. https://doi.org/10.1176/appi.ajp.2018.17101124

Kerestes, R., Segreti, A. M., Pan, L. A., Phillips, M. L., Birmaher, B., Brent, D. A., & Ladouceur, C. D. (2016). Altered neural function to happy faces in adolescents with and at risk for depression. *Journal of Affective Disorders, 192*, 143–152. https://doi.org/10.1016/j.jad.2015.12.013

Keri, S., Nagy, O., Kelemen, O., Myers, C. E., & Gluck, M. A. (2005). Dissociation between medial temporal lobe and basal ganglia memory systems in schizophrenia. *Schizophrenia Research, 77*(2–3), 321–328.

Kim, M. S., Kang, B. N., & Lim, J. Y. (2016). Decision-making deficits in patients with chronic schizophrenia: Iowa Gambling Task and Prospect Valence Learning model. *Neuropsychiatric Disease and Treatment, 12*, 1019–1027. https://doi.org/10.2147/NDT.S103821

Kim, Y. T., Lee, K. U., & Lee, S. J. (2009). Deficit in decision-making in chronic, stable schizophrenia: From a reward and punishment perspective. *Psychiatry Investigation, 6*(1), 26–33.

Kluge, A., Kirschner, M., Hager, O. M., Bischof, M., Habermeyer, B., Seifritz, E., … Kaiser, S. (2018). Combining actigraphy, ecological momentary assessment and neuroimaging to study apathy in patients with schizophrenia. *Schizophrenia Research, 195*, 176–182. https://doi.org/10.1016/j.schres.2017.09.034

Knutson, B., Bhanji, J. P., Cooney, R. E., Atlas, L. Y., & Gotlib, I. H. (2008). Neural responses to monetary incentives in major depression. *Biological Psychiatry, 63*(7), 686–692. https://doi.org/10.1016/j.biopsych.2007.07.023

Knutson, B., Fong, G. W., Adams, C. M., Varner, J. L., & Hommer, D. (2001). Dissociation of reward anticipation and outcome with event-related fMRI. *Neuroreport, 12*(17), 3683–3687.

Kring, A. M., & Barch, D. M. (2014). The motivation and pleasure dimension of negative symptoms: Neural substrates and behavioral outputs. *European Neuropsychopharmacology, 24*(5), 725–736. https://doi.org/10.1016/j.euroneuro.2013.06.007

Kring, A. M., & Moran, E. K. (2008). Emotional response deficits in schizophrenia: Insights from affective science. *Schizophrenia Bulletin, 34*(5), 819–834.

Kringelbach, M. L., & Berridge, K. C. (2017). The affective core of emotion: Linking pleasure, subjective well-being, and optimal metastability in the brain. *Emotion Review, 9*(3), 191–199. https://doi.org/10.1177/1754073916684558

Kujawa, A., Proudfit, G. H., & Klein, D. N. (2014). Neural reactivity to rewards and losses in offspring of mothers and fathers with histories of depressive and anxiety disorders. *Journal of Abnormal Psychology, 123*(2), 287–297. https://doi.org/10.1037/a0036285

Kumar, P., Goer, F., Murray, L., Dillon, D. G., Beltzer, M. L., Cohen, A. L., … Pizzagalli, D. A. (2018). Impaired reward prediction error encoding and striatal-midbrain connectivity in depression. *Neuropsychopharmacology, 43*(7), 1581–1588. https://doi.org/10.1038/s41386-018-0032-x

Kumar, P., Waiter, G., Ahearn, T., Milders, M., Reid, I., & Steele, J. D. (2008). Abnormal temporal difference reward-learning signals in major depression. *Brain: A Journal of Neurology, 131*(Pt 8), 2084–2093. https://doi.org/10.1093/brain/awn136

Kurniawan, I. T., Guitart-Masip, M., Dayan, P., & Dolan, R. J. (2013). Effort and valuation in the brain: the effects of anticipation and execution. *The Journal of Neuroscience, 33*(14), 6160–6169. https://doi.org/10.1523/JNEUROSCI.4777-12.2013

Lalanne, L., Ayranci, G., Kieffer, B. L., & Lutz, P. E. (2014). The kappa opioid receptor: from addiction to depression, and back. *Frontiers in Psychiatry, 5*, 170. https://doi.org/10.3389/fpsyt.2014.00170

Lesh, T. A., Niendam, T. A., Minzenberg, M. J., & Carter, C. S. (2011). Cognitive control deficits in schizophrenia: mechanisms and meaning. *Neuropsychopharmacology, 36*(1), 316–338. https://doi.org/10.1038/npp.2010.156

Lesh, T. A., Westphal, A. J., Niendam, T. A., Yoon, J. H., Minzenberg, M. J., Ragland, J. D., … Carter, C. S. (2013). Proactive and reactive cognitive control and dorsolateral prefrontal cortex dysfunction in first episode schizophrenia. *NeuroImage: Clinical, 2*, 590–599. https://doi.org/10.1016/j.nicl.2013.04.010

Liu, W. H., Roiser, J. P., Wang, L. Z., Zhu, Y. H., Huang, J., Neumann, D. L., … Chan, R. C. K. (2016). Anhedonia is associated with blunted reward sensitivity in first-degree relatives of patients with major depression. *Journal of Affective Disorders, 190*, 640–648. https://doi.org/10.1016/j.jad.2015.10.050

Liu, W. H., Wang, L. Z., Shang, H. R., Shen, Y., Li, Z., Cheung, E. F., & Chan, R. C. (2014). The influence of anhedonia on feedback negativity in major depressive disorder. *Neuropsychologia, 53*, 213–220. https://doi.org/10.1016/j.neuropsychologia.2013.11.023

Liverant, G. I., Sloan, D. M., Pizzagalli, D. A., Harte, C. B., Kamholz, B. W., Rosebrock, L. E., … Kaplan, G. B. (2014). Associations among smoking, anhedonia, and reward learning in depression. *Behavior Therapy, 45*(5), 651–663. https://doi.org/10.1016/j.beth.2014.02.004

Llerena, K., Wynn, J. K., Hajcak, G., Green, M. F., & Horan, W. P. (2016). Patterns and reliability of EEG during error monitoring for internal versus external feedback in schizophrenia. *International Journal of Psychophysiology, 105*, 39–46. https://doi.org/10.1016/j.ijpsycho.2016.04.012

Luking, K. R., Neiman, J. S., Luby, J. L., & Barch, D. M. (2017). Reduced hedonic capacity/approach motivation relates to blunted responsivity to gain and loss feedback in children. *Journal of Clinical Child and Adolescent Psychology, 46*(3), 450–462. https://doi.org/10.1080/15374416.2015.1012721

Luking, K. R., Pagliaccio, D., Luby, J. L., & Barch, D. M. (2016a). Depression risk predicts blunted neural responses to gains and enhanced responses to losses in healthy children. *Journal*

of the American Academy of Child and Adolescent Psychiatry, 55(4), 328–337. https://doi.org/10.1016/j.jaac.2016.01.007

Luking, K. R., Pagliaccio, D., Luby, J. L., & Barch, D. M. (2016b). Reward processing and risk for depression across development. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2016.04.002

MacDonald, A.W., III, Patzelt, E., Kurth-Nelson, Z., Barch, D. M., Carter, C. S., Gold, J. M., … Strauss, M. E. (in submission). Shared reversal learning impairments in schizophrenia and bipolar disorder reflect a failure to exploit rewards in computational model.

Maddox, W. T., Gorlick, M. A., Worthy, D. A., & Beevers, C. G. (2012). Depressive symptoms enhance loss-minimization, but attenuate gain-maximization in history-dependent decision-making. *Cognition, 125*(1), 118–124. https://doi.org/10.1016/j.cognition.2012.06.011

Maia, T. V. (2009). Reinforcement learning, conditioning, and the brain: Successes and challenges. *Cognitive, Affective, & Behavioral Neuroscience, 9*(4), 343–364. https://doi.org/10.3758/CABN.9.4.343

Mann, C. L., Footer, O., Chung, Y. S., Driscoll, L. L., & Barch, D. M. (2013). Spared and impaired aspects of motivated cognitive control in schizophrenia. *Journal of Abnormal Psychology, 122*(3), 745–755. https://doi.org/10.1037/a0033069

Martinelli, C., Rigoli, F., Dolan, R. J., & Shergill, S. S. (2018). Decreased value-sensitivity in schizophrenia. *Psychiatry Research, 259*, 295–301. https://doi.org/10.1016/j.psychres.2017.10.031

Massar, S. A., Libedinsky, C., Weiyan, C., Huettel, S. A., & Chee, M. W. (2015). Separate and overlapping brain areas encode subjective value during delay and effort discounting. *NeuroImage, 120*, 104–113. https://doi.org/10.1016/j.neuroimage.2015.06.080

McCabe, C., Cowen, P. J., & Harmer, C. J. (2009). Neural representation of reward in recovered depressed patients. *Psychopharmacology, 205*(4), 667–677. https://doi.org/10.1007/s00213-009-1573-9

McCabe, C., Woffindale, C., Harmer, C. J., & Cowen, P. J. (2012). Neural processing of reward and punishment in young people at increased familial risk of depression. *Biological Psychiatry, 72*(7), 588–594. https://doi.org/10.1016/j.biopsych.2012.04.034

McCarthy, J. M., Treadway, M. T., Bennett, M. E., & Blanchard, J. J. (2016). Inefficient effort allocation and negative symptoms in individuals with schizophrenia. *Schizophrenia Research, 170*(2–3), 278–284. https://doi.org/10.1016/j.schres.2015.12.017

McDermott, L. M., & Ebmeier, K. P. (2009). A meta-analysis of depression severity and cognitive function. *Journal of Affective Disorders, 119*(1–3), 1–8. https://doi.org/10.1016/j.jad.2009.04.022

Medic, N., Ziauddeen, H., Vestergaard, M. D., Henning, E., Schultz, W., Farooqi, I. S., & Fletcher, P. C. (2014). Dopamine modulates the neural representation of subjective value of food in hungry subjects. *The Journal of Neuroscience, 34*(50), 16856–16864. https://doi.org/10.1523/JNEUROSCI.2051-14.2014

Meyer, A., Bress, J. N., Hajcak, G., & Gibb, B. E. (2018). Maternal depression is related to reduced error-related brain activity in child and adolescent offspring. *Journal of Clinical Child and Adolescent Psychology, 47*(2), 324–335. https://doi.org/10.1080/15374416.2016.1138405

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 21*, 167–202.

Miller, J. M., Zanderigo, F., Purushothaman, P. D., DeLorenzo, C., Rubin-Falcone, H., Ogden, R. T., … Mann, J. J. (2018). Kappa opioid receptor binding in major depression: A pilot study. *Synapse*. https://doi.org/10.1002/syn.22042

Minami, S., Satoyoshi, H., Ide, S., Inoue, T., Yoshioka, M., & Minami, M. (2017). Suppression of reward-induced dopamine release in the nucleus accumbens in animal models of depression: Differential responses to drug treatment. *Neuroscience Letters, 650*, 72–76. https://doi.org/10.1016/j.neulet.2017.04.028

Minzenberg, M. J., Laird, A. R., Thelen, S., Carter, C. S., & Glahn, D. C. (2009). Meta-analysis of 41 functional neuroimaging studies of executive function in schizophrenia. *Archives of General Psychiatry, 66*(8), 811–822. https://doi.org/10.1001/archgenpsychiatry.2009.91

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience, 16*, 1936–1947.

Moran, E. K., Culbreth, A. J., & Barch, D. M. (2017). Ecological momentary assessment of negative symptoms in schizophrenia: Relationships to effort-based decision making and reinforcement learning. *Journal of Abnormal Psychology, 126*(1), 96–105. https://doi.org/10.1037/abn0000240

Moran, E. K., Culbreth, A. J., Kandala, S., & Barch, D. M. (in submission). Reward anticipation in schizophrenia: neural and psychological factors.

Morgan, J. K., Olino, T. M., McMakin, D. L., Ryan, N. D., & Forbes, E. E. (2013). Neural response to reward as a predictor of increases in depressive symptoms in adolescence. *Neurobiology of Disease, 52*, 66–74. https://doi.org/10.1016/j.nbd.2012.03.039

Morkl, S., Blesl, C., Jahanshahi, M., Painold, A., & Holl, A. K. (2016). Impaired probabilistic classification learning with feedback in patients with major depression. *Neurobiology of Learning and Memory, 127*, 48–55. https://doi.org/10.1016/j.nlm.2015.12.001

Morris, R. W., Cyrzon, C., Green, M. J., Le Pelley, M. E., & Balleine, B. W. (2018). Impairments in action-outcome learning in schizophrenia. *Translational Psychiatry, 8*(1), 54. https://doi.org/10.1038/s41398-018-0103-0

Morris, S. E., Holroyd, C. B., Mann-Wrobel, M. C., & Gold, J. M. (2011). Dissociation of response and feedback negativity in schizophrenia: Electrophysiological and computational evidence for a deficit in the representation of value. *Frontiers in Human Neuroscience, 5*, 123. https://doi.org/10.3389/fnhum.2011.00123

Mote, J., Minzenberg, M. J., Carter, C. S., & Kring, A. M. (2014). Deficits in anticipatory but not consummatory pleasure in people with recent-onset schizophrenia spectrum disorders. *Schizophrenia Research, 159*(1), 76–79. https://doi.org/10.1016/j.schres.2014.07.048

Mucci, A., Dima, D., Soricelli, A., Volpe, U., Bucci, P., Frangou, S., … Maj, M. (2015). Is avolition in schizophrenia associated with a deficit of dorsal caudate activity? A functional magnetic resonance imaging study during reward anticipation and feedback. *Psychological Medicine, 45*(8), 1765–1778. https://doi.org/10.1017/S0033291714002943

Murphy, N. P. (2015). Dynamic measurement of extracellular opioid activity: Status quo, challenges, and significance in rewarded behaviors. *ACS Chemical Neuroscience, 6*(1), 94–107. https://doi.org/10.1021/cn500295q

Must, A., Horvath, S., Nemeth, V. L., & Janka, Z. (2013). The Iowa Gambling Task in depression - What have we learned about sub-optimal decision-making strategies? *Frontiers in Psychology, 4*, 732. https://doi.org/10.3389/fpsyg.2013.00732

Nelson, B. D., Kessel, E. M., Klein, D. N., & Shankman, S. A. (2018). Depression symptom dimensions and asymmetrical frontal cortical activity while anticipating reward. *Psychophysiology, 55*(1). https://doi.org/10.1111/psyp.12892

Nelson, B. D., McGowan, S. K., Sarapas, C., Robison-Andrew, E. J., Altman, S. E., Campbell, M. L., … Shankman, S. A. (2013). Biomarkers of threat and reward sensitivity demonstrate unique associations with risk for psychopathology. *Journal of Abnormal Psychology, 122*(3), 662–671. https://doi.org/10.1037/a0033982

Nelson, B. D., Perlman, G., Klein, D. N., Kotov, R., & Hajcak, G. (2016). Blunted neural response to rewards as a prospective predictor of the development of depression in adolescent girls. *The American Journal of Psychiatry*. https://doi.org/10.1176/appi.ajp.2016.15121524

Nelson, B. D., Shankman, S. A., & Proudfit, G. H. (2014). Intolerance of uncertainty mediates reduced reward anticipation in major depressive disorder. *Journal of Affective Disorders, 158*, 108–113. https://doi.org/10.1016/j.jad.2014.02.014

Nestor, P. G., Choate, V., Niznikiewicz, M., Levitt, J. J., Shenton, M. E., & McCarley, R. W. (2014). Neuropsychology of reward learning and negative symptoms in schizophrenia. *Schizophrenia Research, 159*(2–3), 506–508. https://doi.org/10.1016/j.schres.2014.08.028

Nielsen, M. O., Rostrup, E., Broberg, B. V., Wulff, S., & Glenthoj, B. (2018). Negative symptoms and reward disturbances in schizophrenia before and after antipsychotic monotherapy. *Clinical EEG and Neuroscience, 49*(1), 36–45. https://doi.org/10.1177/1550059417744120

Nielsen, M. O., Rostrup, E., Wulff, S., Bak, N., Broberg, B. V., Lublin, H., … Glenthoj, B. (2012). Improvement of brain reward abnormalities by antipsychotic monotherapy in schizophrenia. *Archives of General Psychiatry*, 1–10. https://doi.org/10.1001/archgenpsychiatry.2012.847

Nielsen, M. O., Rostrup, E., Wulff, S., Bak, N., Lublin, H., Kapur, S., & Glenthoj, B. (2012). Alterations of the brain reward system in antipsychotic naive schizophrenia patients. *Biological Psychiatry, 71*(10), 898–905. https://doi.org/10.1016/j.biopsych.2012.02.007

O'Doherty, J. P. (2007). Lights, camembert, action! The role of human orbitofrontal cortex in encoding stimuli, rewards and choices. *Annals of the New York Academy of Sciences, 1121*, 254–272.

Olino, T. M., Silk, J. S., Osterritter, C., & Forbes, E. E. (2015). Social reward in youth at risk for depression: A preliminary investigation of subjective and neural differences. *Journal of Child and Adolescent Psychopharmacology, 25*(9), 711–721. https://doi.org/10.1089/cap.2014.0165

Otto, A. R., Skatova, A., Madlon-Kay, S., & Daw, N. D. (2015). Cognitive control predicts use of model-based reinforcement learning. *Journal of Cognitive Neuroscience, 27*(2), 319–333. https://doi.org/10.1162/jocn_a_00709

Padoa-Schioppa, C., & Cai, X. (2011). The orbitofrontal cortex and the computation of subjective value: Consolidated concepts and new perspectives. *Annals of the New York Academy of Sciences, 1239*, 130–137. https://doi.org/10.1111/j.1749-6632.2011.06262.x

Padoa-Schioppa, C., & Conen, K. E. (2017). Orbitofrontal cortex: A neural circuit for economic decisions. *Neuron, 96*(4), 736–754. https://doi.org/10.1016/j.neuron.2017.09.031

Paradiso, S., Andreasen, N. C., Crespo-Facorro, B., O'Leary, D. S., Watkins, G. L., Boles Ponto, L. L., & Hichwa, R. D. (2003). Emotions in unmedicated patients with schizophrenia during evaluation with positron emission tomography. *The American Journal of Psychiatry, 160*(10), 1775–1783.

Park, I. H., Lee, B. C., Kim, J. J., Kim, J. I., & Koo, M. S. (2017). Effort-based reinforcement processing and functional connectivity underlying amotivation in medicated patients with depression and schizophrenia. *The Journal of Neuroscience, 37*(16), 4370–4380. https://doi.org/10.1523/JNEUROSCI.2524-16.2017

Pechtel, P., Dutra, S. J., Goetz, E. L., & Pizzagalli, D. A. (2013). Blunted reward responsiveness in remitted depression. *Journal of Psychiatric Research, 47*(12), 1864–1869. https://doi.org/10.1016/j.jpsychires.2013.08.011

Pizzagalli, D. A., Iosifescu, D., Hallett, L. A., Ratner, K. G., & Fava, M. (2008). Reduced hedonic capacity in major depressive disorder: Evidence from a probabilistic reward task. *Journal of Psychiatric Research, 43*(1), 76–87. https://doi.org/10.1016/j.jpsychires.2008.03.001

Plailly, J., d'Amato, T., Saoud, M., & Royet, J. P. (2006). Left temporo-limbic and orbital dysfunction in schizophrenia during odor familiarity and hedonicity judgments. *NeuroImage, 29*(1), 302–313.

Prevost, C., Pessiglione, M., Metereau, E., Clery-Melin, M. L., & Dreher, J. C. (2010). Separate valuation subsystems for delay and effort decision costs. *The Journal of Neuroscience, 30*(42), 14080–14090. https://doi.org/10.1523/JNEUROSCI.2752-10.2010

Radua, J., Schmidt, A., Borgwardt, S., Heinz, A., Schlagenhauf, F., McGuire, P., & Fusar-Poli, P. (2015). Ventral striatal activation during reward processing in psychosis: A neurofunctional meta-analysis. *JAMA Psychiatry, 72*(12), 1243–1251. https://doi.org/10.1001/jamapsychiatry.2015.2196

Ragland, J. D., Laird, A. R., Ranganath, C., Blumenfeld, R. S., Gonzales, S. M., & Glahn, D. C. (2009). Prefrontal activation deficits during episodic memory in schizophrenia. *The American Journal of Psychiatry, 166*(8), 863–874.

Rassovsky, Y., Green, M. F., Nuechterlein, K. H., Breitmeyer, B., & Mintz, J. (2005). Modulation of attention during visual masking in schizophrenia. *The American Journal of Psychiatry, 162*(8), 1533–1535.

Reddy, L. F., Horan, W. P., Barch, D. M., Buchanan, R. W., Dunayevich, E., Gold, J. M., … Green, M. F. (2015). Effort-based decision-making paradigms for clinical trials in schizophrenia: Part 1-psychometric characteristics of 5 paradigms. *Schizophrenia Bulletin, 41*(5), 1045–1054. https://doi.org/10.1093/schbul/sbv089

Reddy, L. F., Waltz, J. A., Green, M. F., Wynn, J. K., & Horan, W. P. (2016). Probabilistic reversal learning in schizophrenia: Stability of deficits and potential causal mechanisms. *Schizophrenia Bulletin, 42*(4), 942–951. https://doi.org/10.1093/schbul/sbv226

Redlich, R., Dohm, K., Grotegerd, D., Opel, N., Zwitserlood, P., Heindel, W., … Dannlowski, U. (2015). Reward processing in unipolar and bipolar depression: A functional MRI study. *Neuropsychopharmacology, 40*(11), 2623–2631. https://doi.org/10.1038/npp.2015.110

Reinen, J., Smith, E. E., Insel, C., Kribs, R., Shohamy, D., Wager, T. D., & Jarskog, L. F. (2014). Patients with schizophrenia are impaired when learning in the context of pursuing rewards. *Schizophrenia Research, 152*(1), 309–310. https://doi.org/10.1016/j.schres.2013.11.012

Reinen, J. M., Van Snellenberg, J. X., Horga, G., Abi-Dargham, A., Daw, N. D., & Shohamy, D. (2016). Motivational context modulates prediction error response in schizophrenia. *Schizophrenia Bulletin*. https://doi.org/10.1093/schbul/sbw045

Robinson, O. J., Cools, R., Carlisi, C. O., Sahakian, B. J., & Drevets, W. C. (2012). Ventral striatum response during reward and punishment reversal learning in unmedicated major depressive disorder. *The American Journal of Psychiatry, 169*(2), 152–159.

Rock, P. L., Roiser, J. P., Riedel, W. J., & Blackwell, A. D. (2014). Cognitive impairment in depression: A systematic review and meta-analysis. *Psychological Medicine, 44*(10), 2029–2040. https://doi.org/10.1017/S0033291713002535

Rolls, E. T., Sienkiewicz, Z. J., & Yaxley, S. (1989). Hunger modulates the responses to gustatory stimuli of single neurons in the caudolateral orbitofrontal cortex of the macaque monkey. *The European Journal of Neuroscience, 1*(1), 53–60.

Rothkirch, M., Tonn, J., Kohler, S., & Sterzer, P. (2017). Neural mechanisms of reinforcement learning in unmedicated patients with major depressive disorder. *Brain, 140*(4), 1147–1157. https://doi.org/10.1093/brain/awx025

Rudebeck, P. H., Walton, M. E., Smyth, A. N., Bannerman, D. M., & Rushworth, M. F. (2006). Separate neural pathways process different decision costs. *Nature Neuroscience, 9*(9), 1161–1168.

Rushworth, M. F., Behrens, T. E., Rudebeck, P. H., & Walton, M. E. (2007). Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. *Trends in Cognitive Sciences, 11*(4), 168–176.

Rutledge, R. B., Moutoussis, M., Smittenaar, P., Zeidman, P., Taylor, T., Hrynkiewicz, L., … Dolan, R. J. (2017). Association of neural and emotional impacts of reward prediction errors with major depression. *JAMA Psychiatry, 74*(8), 790–797. https://doi.org/10.1001/jamapsychiatry.2017.1713

Rzepa, E., Fisk, J., & McCabe, C. (2017). Blunted neural response to anticipation, effort and consummation of reward and aversion in adolescents with depression symptomatology. *Journal of Psychopharmacology, 31*(3), 303–311. https://doi.org/10.1177/0269881116681416

Salamone, J. D., Correa, M., Farrar, A., & Mingote, S. M. (2007). Effort-related functions of nucleus accumbens dopamine and associated forebrain circuits. *Psychopharmacology, 191*(3), 461–482.

Salamone, J. D., Correa, M., Nunes, E. J., Randall, P. A., & Pardo, M. (2012). The behavioral pharmacology of effort-related choice behavior: Dopamine, adenosine and beyond. *Journal of the Experimental Analysis of Behavior, 97*(1), 125–146. https://doi.org/10.1901/jeab.2012.97-125

Salamone, J. D., Correa, M., Yohn, S., Lopez Cruz, L., San Miguel, N., & Alatorre, L. (2016). The pharmacology of effort-related choice behavior: Dopamine, depression, and individual differences. *Behavioural Processes, 127*, 3–17. https://doi.org/10.1016/j.beproc.2016.02.008

Satterthwaite, T. D., Kable, J. W., Vandekar, L., Katchmar, N., Bassett, D. S., Baldassano, C. F., … Wolf, D. H. (2015). Common and dissociable dysfunction of the reward system in bipolar and unipolar depression. *Neuropsychopharmacology, 40*(9), 2258–2268. https://doi.org/10.1038/npp.2015.75

Savitz, J. B., & Drevets, W. C. (2013). Neuroreceptor imaging in depression. *Neurobiology of Disease, 52*, 49–65. https://doi.org/10.1016/j.nbd.2012.06.001

Schlagenhauf, F., Huys, Q. J., Deserno, L., Rapp, M. A., Beck, A., Heinze, H. J., … Heinz, A. (2014). Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. *NeuroImage, 89*, 171–180. https://doi.org/10.1016/j.neuroimage.2013.11.034

Schlagenhauf, F., Sterzer, P., Schmack, K., Ballmaier, M., Rapp, M., Wrase, J., … Heinz, A. (2009). Reward feedback alterations in unmedicated schizophrenia patients: Relevance for delusions. *Biological Psychiatry, 65*(12), 1032–1039.

Schneider, F., Habel, U., Reske, M., Toni, I., Falkai, P., & Shah, N. J. (2007). Neural substrates of olfactory processing in schizophrenia patients and their healthy relatives. *Psychiatry Research, 155*(2), 103–112.

Schultz, W. (2007). Multiple dopamine functions at different time courses. *Annual Review of Neuroscience, 30*, 259–288.

Schultz, W. (2016a). Dopamine reward prediction error coding. *Dialogues in Clinical Neuroscience, 18*(1), 23–32.

Schultz, W. (2016b). Reward functions of the basal ganglia. *Journal of Neural Transmission (Vienna), 123*(7), 679–693. https://doi.org/10.1007/s00702-016-1510-0

Scinska, A., Sienkiewicz-Jarosz, H., Kuran, W., Ryglewicz, D., Rogowski, A., Wrobel, E., … Bienkowski, P. (2004). Depressive symptoms and taste reactivity in humans. *Physiology & Behavior, 82*(5), 899–904. https://doi.org/10.1016/j.physbeh.2004.07.012

Serper, M., Payne, E., Dill, C., Portillo, C., & Taliercio, J. (2017). Allocating effort and anticipating pleasure in schizophrenia: Relationship with real world functioning. *European Psychiatry, 46*, 57–64. https://doi.org/10.1016/j.eurpsy.2017.07.008

Shankman, S. A., Klein, D. N., Tenke, C. E., & Bruder, G. E. (2007). Reward sensitivity in depression: A biobehavioral study. *Journal of Abnormal Psychology, 116*(1), 95–104. https://doi.org/10.1037/0021-843X.116.1.95

Shankman, S. A., Nelson, B. D., Sarapas, C., Robison-Andrew, E. J., Campbell, M. L., Altman, S. E., … Gorka, S. M. (2013). A psychophysiological investigation of threat and reward sensitivity in individuals with panic disorder and/or major depressive disorder. *Journal of Abnormal Psychology, 122*(2), 322–338. https://doi.org/10.1037/a0030747

Sheline, Y. I., Barch, D. M., Price, J. L., Rundle, M. M., Vaishnavi, S. N., Snyder, A. Z., … Raichle, M. E. (2009). The default mode network and self-referential processes in depression. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 1942–1947.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron, 79*(2), 217–240. https://doi.org/10.1016/j.neuron.2013.07.007

Shenhav, A., Cohen, J. D., & Botvinick, M. M. (2016). Dorsal anterior cingulate cortex and the value of control. *Nature Neuroscience, 19*(10), 1286–1291. https://doi.org/10.1038/nn.4384

Sherdell, L., Waugh, C. E., & Gotlib, I. H. (2012). Anticipatory pleasure predicts motivation for reward in major depression. *Journal of Abnormal Psychology, 121*(1), 51–60. https://doi.org/10.1037/a0024945

Siegert, R. J., Weatherall, M., & Bell, E. M. (2008). Is implicit sequence learning impaired in schizophrenia? A meta-analysis. *Brain and Cognition, 67*(3), 351–359. https://doi.org/10.1016/j.bandc.2008.02.005

Smith, K. S., & Berridge, K. C. (2007). Opioid limbic circuit for reward: Interaction between hedonic hotspots of nucleus accumbens and ventral pallidum. *The Journal of Neuroscience, 27*(7), 1594–1605.

Smoski, M. J., Rittenberg, A., & Dichter, G. S. (2011). Major depressive disorder is characterized by greater reward network activation to monetary than pleasant image rewards. *Psychiatry Research, 194*(3), 263–270. https://doi.org/10.1016/j.pscychresns.2011.06.012

Somlai, Z., Moustafa, A. A., Keri, S., Myers, C. E., & Gluck, M. A. (2011). General functioning predicts reward and punishment learning in schizophrenia. *Schizophrenia Research, 127*(1–3), 131–136. https://doi.org/10.1016/j.schres.2010.07.028

Stepien, M., Manoliu, A., Kubli, R., Schneider, K., Tobler, P. N., Seifritz, E., … Kirschner, M. (2018). Investigating the association of ventral and dorsal striatal dysfunction during reward anticipation with negative symptoms in patients with schizophrenia and healthy individuals. *PLoS One, 13*(6), e0198215. https://doi.org/10.1371/journal.pone.0198215

Stoy, M., Schlagenhauf, F., Sterzer, P., Bermpohl, F., Hagele, C., Suchotzki, K., … Strohle, A. (2012). Hyporeactivity of ventral striatum towards incentive stimuli in unmedicated depressed patients normalizes after treatment with escitalopram. *Journal of Psychopharmacology, 26*(5), 677–688. https://doi.org/10.1177/0269881111416686

Strauss, G. P., Visser, K. F., Keller, W. R., Gold, J. M., & Buchanan, R. W. (2018). Anhedonia reflects impairment in making relative value judgments between positive and neutral stimuli in schizophrenia. *Schizophrenia Research*. https://doi.org/10.1016/j.schres.2018.02.016

Strauss, G. P., Whearty, K. M., Morra, L. F., Sullivan, S. K., Ossenfort, K. L., & Frost, K. H. (2016). Avolition in schizophrenia is associated with reduced willingness to expend effort for reward on a Progressive Ratio task. *Schizophrenia Research, 170*(1), 198–204. https://doi.org/10.1016/j.schres.2015.12.006

Stringaris, A., Vidal-Ribas Belil, P., Artiges, E., Lemaitre, H., Gollier-Briant, F., Wolke, S., … Consortium, Imagen. (2015). The brain's response to reward anticipation and depression in adolescence: Dimensionality, specificity, and longitudinal predictions in a community-based sample. *The American Journal of Psychiatry, 172*(12), 1215–1223. https://doi.org/10.1176/appi.ajp.2015.14101298

Subramaniam, K., Hooker, C. I., Biagianti, B., Fisher, M., Nagarajan, S., & Vinogradov, S. (2015). Neural signal during immediate reward anticipation in schizophrenia: Relationship to real-world motivation and function. *NeuroImage: Clinical, 9*, 153–163. https://doi.org/10.1016/j.nicl.2015.08.001

Suzuki, S., Cross, L., & O'Doherty, J. P. (2017). Elucidating the underlying components of food valuation in the human orbitofrontal cortex. *Nature Neuroscience, 20*(12), 1780–1786. https://doi.org/10.1038/s41593-017-0008-x

Takamura, M., Okamoto, Y., Okada, G., Toki, S., Yamamoto, T., Ichikawa, N., … Yamawaki, S. (2017). Patients with major depressive disorder exhibit reduced reward size coding in the striatum. *Progress in Neuropsychopharmacology and Biological Psychiatry, 79*(Pt B), 317–323. https://doi.org/10.1016/j.pnpbp.2017.07.006

Taylor, N., Hollis, J. P., Corcoran, S., Gross, R., Cuthbert, B., Swails, L. W., & Duncan, E. (2018). Impaired reward responsiveness in schizophrenia. *Schizophrenia Research*. https://doi.org/10.1016/j.schres.2018.02.057

Taylor, S. F., Phan, K. L., Britton, J. C., & Liberzon, I. (2005). Neural response to emotional salience in schizophrenia. *Neuropsychopharmacology, 30*(5), 984–995.

Tobia, M. J., Guo, R., Schwarze, U., Boehmer, W., Glascher, J., Finckh, B., … Sommer, T. (2014). Neural systems for choice and valuation with counterfactual learning signals. *NeuroImage, 89*, 57–69. https://doi.org/10.1016/j.neuroimage.2013.11.051

Treadway, M. T., Bossaller, N. A., Shelton, R. C., & Zald, D. H. (2012). Effort-based decision-making in major depressive disorder: A translational model of motivational anhedonia. *Journal of Abnormal Psychology, 121*(3), 553–558. https://doi.org/10.1037/a0028813

Treadway, M. T., Buckholtz, J. W., Cowan, R. L., Woodward, N. D., Li, R., Ansari, M. S., … Zald, D. H. (2012). Dopaminergic mechanisms of individual differences in human effort-based decision-making. *The Journal of Neuroscience, 32*(18), 6170–6176. https://doi.org/10.1523/JNEUROSCI.6459-11.2012

Treadway, M. T., Peterman, J. S., Zald, D. H., & Park, S. (2015). Impaired effort allocation in patients with schizophrenia. *Schizophrenia Research, 161*(2–3), 382–385. https://doi.org/10.1016/j.schres.2014.11.024

Tremeau, F., Antonius, D., Nolan, K., Butler, P., & Javitt, D. C. (2014). Immediate affective motivation is not impaired in schizophrenia. *Schizophrenia Research, 159*(1), 157–163. https://doi.org/10.1016/j.schres.2014.08.001

Trifilieff, P., Feng, B., Urizar, E., Winiger, V., Ward, R. D., Taylor, K. M., … Javitch, J. A. (2013). Increasing dopamine D2 receptor expression in the adult nucleus accumbens enhances motivation. *Molecular Psychiatry*. https://doi.org/10.1038/mp.2013.57

Turnbull, O. H., Evans, C. E., Kemish, K., Park, S., & Bowman, C. H. (2006). A novel set-shifting modification of the iowa gambling task: Flexible emotion-based learning in schizophrenia. *Neuropsychology, 20*(3), 290–298.

Ubl, B., Kuehner, C., Kirsch, P., Ruttorf, M., Diener, C., & Flor, H. (2015). Altered neural reward and loss processing and prediction error signalling in depression. *Social Cognitive and Affective Neuroscience*. https://doi.org/10.1093/scan/nsu158

Urban-Kowalczyk, M., Smigielski, J., & Kotlicka-Antczak, M. (2018). Overrated hedonic judgment of odors in patients with schizophrenia. *CNS Neuroscience & Therapeutics*. https://doi.org/10.1111/cns.12849

Vaidyanathan, U., Nelson, L. D., & Patrick, C. J. (2012). Clarifying domains of internalizing psychopathology using neurophysiology. *Psychological Medicine, 42*(3), 447–459. https://doi.org/10.1017/S0033291711001528

Vanes, L. D., Mouchlianitis, E., Collier, T., Averbeck, B. B., & Shergill, S. S. (2018). Differential neural reward mechanisms in treatment-responsive and treatment-resistant schizophrenia. *Psychological Medicine*, 1–10. https://doi.org/10.1017/S0033291718000041

Vassena, E., Holroyd, C. B., & Alexander, W. H. (2017). Computational models of anterior cingulate cortex: At the crossroads between prediction and effort. *Frontiers in Neuroscience, 11*, 316. https://doi.org/10.3389/fnins.2017.00316

Vrieze, E., Pizzagalli, D. A., Demyttenaere, K., Hompes, T., Sienaert, P., de Boer, P., … Claes, S. (2013). Reduced reward learning predicts outcome in major depressive disorder. *Biological Psychiatry, 73*(7), 639–645. https://doi.org/10.1016/j.biopsych.2012.10.014

Wallis, J. D. (2007). Orbitofrontal cortex and its contribution to decision-making. *Annual Review of Neuroscience, 30*, 31–56.

Walsh, A. E. L., Browning, M., Drevets, W. C., Furey, M., & Harmer, C. J. (2018). Dissociable temporal effects of bupropion on behavioural measures of emotional and reward processing in depression. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 373*(1742). https://doi.org/10.1098/rstb.2017.0030

Walter, H., Heckers, S., Kassubek, J., Erk, S., Frasch, K., & Abler, B. (2010). Further evidence for aberrant prefrontal salience coding in schizophrenia. *Frontiers in Behavioral Neuroscience, 3*, 62. https://doi.org/10.3389/neuro.08.062.2009

Walton, M. E., Bannerman, D. M., Alterescu, K., & Rushworth, M. F. (2003). Functional specialization within medial frontal cortex of the anterior cingulate for evaluating effort-related decisions. *The Journal of Neuroscience, 23*(16), 6475–6479.

Waltz, J. A., Brown, J. K., Gold, J. M., Ross, T. J., Salmeron, B. J., & Stein, E. A. (2015). Probing the dynamic updating of value in schizophrenia using a sensory-specific satiety paradigm. *Schizophrenia Bulletin, 41*(5), 1115–1122. https://doi.org/10.1093/schbul/sbv034

Waltz, J. A., Frank, M. J., Robinson, B. M., & Gold, J. M. (2007). Selective reinforcement learning deficits in schizophrenia support predictions from computational models of striatal-cortical dysfunction. *Biological Psychiatry, 62*, 756–764.

Waltz, J. A., & Gold, J. M. (2007). Probabilistic reversal learning impairments in schizophrenia: Further evidence of orbitofrontal dysfunction. *Schizophrenia Research, 93*(1–3), 296–303.

Waltz, J. A., & Gold, J. M. (2016). Motivational deficits in schizophrenia and the representation of expected value. *Current Topics in Behavioral Neurosciences, 27*, 375–410. https://doi.org/10.1007/7854_2015_385

Waltz, J. A., Kasanova, Z., Ross, T. J., Salmeron, B. J., McMahon, R. P., Gold, J. M., & Stein, E. A. (2013). The roles of reward, default, and executive control networks in set-shifting impairments in schizophrenia. *PLoS One, 8*(2), e57257. https://doi.org/10.1371/journal.pone.0057257

Waltz, J. A., Schweitzer, J. B., Gold, J. M., Kurup, P. K., Ross, T. J., Salmeron, B. J., … Stein, E. A. (2009). Patients with schizophrenia have a reduced neural response to both unpredictable and predictable primary reinforcers. *Neuropsychopharmacology, 34*(6), 1567–1577.

Waltz, J. A., Schweitzer, J. B., Ross, T. J., Kurup, P. K., Salmeron, B. J., Rose, E. J., … Stein, E. A. (2010). Abnormal responses to monetary outcomes in cortex, but not in the basal ganglia, in schizophrenia. *Neuropsychopharmacology, 35*(12), 2427–2439. https://doi.org/10.1038/npp.2010.126

Wang, J., Huang, J., Yang, X. H., Lui, S. S., Cheung, E. F., & Chan, R. C. (2015). Anhedonia in schizophrenia: Deficits in both motivation and hedonic capacity. *Schizophrenia Research, 168*(1–2), 465–474. https://doi.org/10.1016/j.schres.2015.06.019

Wang, K. S., Smith, D. V., & Delgado, M. R. (2016). Using fMRI to study reward processing in humans: Past, present, and future. *Journal of Neurophysiology, 115*(3), 1664–1678. https://doi.org/10.1152/jn.00333.2015

Wang, L., LaBar, K. S., Smoski, M., Rosenthal, M. Z., Dolcos, F., Lynch, T. R., … McCarthy, G. (2008). Prefrontal mechanisms for executive control over emotional distraction are altered in major depression. *Psychiatry Research, 163*(2), 143–155. https://doi.org/10.1016/j.pscychresns.2007.10.004

Weinberg, A., Liu, H., & Shankman, S. A. (2016). Blunted neural response to errors as a trait marker of melancholic depression. *Biological Psychology, 113*, 100–107. https://doi.org/10.1016/j.biopsycho.2015.11.012

White, D. M., Kraguljac, N. V., Reid, M. A., & Lahti, A. C. (2015). Contribution of substantia nigra glutamate to prediction error signals in schizophrenia: A combined magnetic resonance spectroscopy/functional imaging study. *NPJ Schizophrenia, 1*, 14001. https://doi.org/10.1038/npjschz.2014.1

Whitmer, A. J., Frank, M. J., & Gotlib, I. H. (2012). Sensitivity to reward and punishment in major depressive disorder: Effects of rumination and of single versus multiple experiences. *Cognition and Emotion, 26*(8), 1475–1485. https://doi.org/10.1080/02699931.2012.682973

Whitton, A. E., Kakani, P., Foti, D., Van't Veer, A., Haile, A., Crowley, D. J., & Pizzagalli, D. A. (2016). Blunted neural responses to reward in remitted major depression: A high-density event-related potential study. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 1*(1), 87–95. https://doi.org/10.1016/j.bpsc.2015.09.007

Whitton, A. E., Van't Veer, A., Kakani, P., Dillon, D. G., Ironside, M. L., Haile, A., … Pizzagalli, D. A. (2017). Acute stress impairs frontocingulate activation during error monitoring in remitted depression. *Psychoneuroendocrinology, 75*, 164–172. https://doi.org/10.1016/j.psyneuen.2016.10.007

Wise, T., Radua, J., Via, E., Cardoner, N., Abe, O., Adams, T. M., … Arnone, D. (2017). Common and distinct patterns of grey-matter volume alteration in major depression and bipolar disorder: Evidence from voxel-based meta-analysis. *Molecular Psychiatry, 22*(10), 1455–1463. https://doi.org/10.1038/mp.2016.72

Wolf, D. H., Satterthwaite, T. D., Kantrowitz, J. J., Katchmar, N., Vandekar, L., Elliott, M. A., & Ruparel, K. (2014). Amotivation in schizophrenia: Integrated assessment with behavioral, clinical, and imaging measures. *Schizophrenia Bulletin, 40*(6), 1328–1337. https://doi.org/10.1093/schbul/sbu026

Yan, C., Su, L., Wang, Y., Xu, T., Yin, D. Z., Fan, M. X., … Chan, R. C. (2016). Multivariate neural representations of value during reward anticipation and consummation in the human orbitofrontal cortex. *Scientific Reports, 6*, 29079. https://doi.org/10.1038/srep29079

Yang, X. H., Huang, J., Lan, Y., Zhu, C. Y., Liu, X. Q., Wang, Y. F., … Chan, R. C. (2016). Diminished caudate and superior temporal gyrus responses to effort-based decision making in patients with first-episode major depressive disorder. *Progress in Neuro-Psychopharmacology & Biological Psychiatry, 64*, 52–59. https://doi.org/10.1016/j.pnpbp.2015.07.006

Yang, X. H., Huang, J., Zhu, C. Y., Wang, Y. F., Cheung, E. F., Chan, R. C., & Xie, G. R. (2014). Motivational deficits in effort-based decision making in individuals with subsyndromal depression, first-episode and remitted depression patients. *Psychiatry Research, 220*(3), 874–882. https://doi.org/10.1016/j.psychres.2014.08.056

Zhang, L., Tang, J., Dong, Y., Ji, Y., Tao, R., Liang, Z., … Wang, K. (2015). Similarities and differences in decision-making impairments between autism spectrum disorder and schizophrenia. *Frontiers in Behavioral Neuroscience, 9*, 259. https://doi.org/10.3389/fnbeh.2015.00259

Zhang, R., Picchioni, M., Allen, P., & Toulopoulou, T. (2016). Working memory in unaffected relatives of patients with schizophrenia: A meta-analysis of functional magnetic resonance imaging studies. *Schizophrenia Bulletin, 42*(4), 1068–1077. https://doi.org/10.1093/schbul/sbv221

Zhang, W. N., Chang, S. H., Guo, L. Y., Zhang, K. L., & Wang, J. (2013). The neural correlates of reward-related processing in major depressive disorder: A meta-analysis of functional magnetic resonance imaging studies. *Journal of Affective Disorders, 151*(2), 531–539. https://doi.org/10.1016/j.jad.2013.06.039

Zou, L. Q., Zhou, H. Y., Lui, S. S. Y., Wang, Y., Wang, Y., Gan, J., … Chan, R. C. K. (2018). Olfactory identification deficit and its relationship with hedonic traits in patients with first-episode schizophrenia and individuals with schizotypy. *Progress in Neuro-Psychopharmacology & Biological Psychiatry, 83*, 137–141. https://doi.org/10.1016/j.pnpbp.2018.01.014

Chapter 6
Motivation: A Valuation Systems Perspective

Andero Uusberg, Gaurav Suri, Carol Dweck, and James J. Gross

Motivation: A Valuation Systems Perspective

The questions that keep behavioral scientists up at night often concern motivation or why people and other animals do what they do. Why do people behave in ways that harm them in the long run? Why don't all students try to learn and all adults engage in exercise? Why do people conform to some norms but break others? Motivation has been central to behavioral science since early theorists such as Sigmund Freud and Clark Hull used it as a foundation for constructing grand accounts of behavior. In the decades since their time, motivation has continued to fascinate researchers both as a focal interest (Dweck, 2017; Ryan, 2012; Shah & Gardner, 2008) and a pathway to understanding other phenomena such as the nervous system (Simpson & Balsam, 2016), emotion (Fox, Lapate, Shackman, & Davidson, 2018), cognition (Braver, 2016; Kreitler, 2013), development (Heckhausen, 2000), individual differences (Corr, DeYoung, & McNaughton, 2013), and social relations (Dunning, 2011). These diverse efforts to understand motivation have yielded a diversity of accounts that await attempts at integration. In this chapter, we offer one such attempt.

A. Uusberg
Department of Psychology, Stanford University, Stanford, CA, USA

Institute of Psychology, University of Tartu, Tartu, Estonia

G. Suri
Department of Psychology, San Francisco State University, San Francisco, CA, USA

C. Dweck · J. J. Gross (✉)
Department of Psychology, Stanford University, Stanford, CA, USA
e-mail: gross@stanford.edu

M. Neta, I. J. Haas (eds.), *Emotion in the Mind and Body*, Nebraska Symposium on Motivation 66, https://doi.org/10.1007/978-3-030-27473-3_6

Our starting point is the idea that understanding motivation involves understanding how behavior obtains its force and direction (Pezzulo, Rigoli, & Friston, 2018). For instance, take the behavior of queuing to buy a ticket for a concert. Motivational force relates to the quantitative aspects of this behavior, such as the time spent in the queue or the price paid for the ticket. Motivational direction relates to the qualitative aspects of this behavior, such as choosing a particular concert or ticket booth over alternatives. Where do these aspects of behavior come from? Our view is that they emerge from the complex dynamics that produce behavior—different mental processes acting and interacting in parallel (Cisek, 2012; Gross, 2015; Hunt & Hayden, 2017; Ochsner & Gross, 2014; Pessoa, 2018). Products of complex dynamics tend to have emergent properties—features that characterize the product but not necessarily any of the individual processes that give rise to it. For example, political will is an emergent property of a society that cannot be found in its entirety within any individual or institution. We view the defining features of motivation—force and direction of behavior—as similarly emergent properties of behavior that need not exist in their entirety anywhere else in the mind. The force and direction of queuing for a ticket simply emerge from a combination of perceptions, beliefs, expectations, plans, feelings, habits, and other mental processes.

In this chapter, we trace the emergence of different motivational phenomena from the mental system that shape behavior. In the first section, we offer a simplified sketch of the systems that give rise to behavior and thereby motivation. Specifically, we introduce the notion of a *valuation system* that shapes behavior by solving two adaptive problems. *Perception loops* within valuation systems solve the problem of understanding the world by matching models of the world to sensory evidence. *Action loops* within valuation systems solve the problem of acting effectively on the world by matching models of means to models of ends. Both loops rely on different versions of hierarchical feedback control, the principle of reducing gaps between pairs of representations by iteratively altering one of them.

In the second section of the paper, we suggest that distributed valuation systems give rise to different forms of motivational force and direction that can be placed along a gradient of complexity, revealing three broad levels. The first *inherent motivation* level consists of the *predictability* and *competence* motives arising from the gaps that perception and action loops seek to minimize. The second *intentional motivation* level consists of *goal commitment* arising from sufficiently realistic and valuable goals and *goal pursuit* arising from synchronization of valuation systems into a behavioral feedback control cycle. The third *identity motivation* level consists of goals about goals, or *identity* as well as pursuit of pursuits, or *self-regulation* that emerges from further synchronization of intentional motivation. These emergent motivational phenomena are often reflected in awareness as feelings that modulate the operation of distributed valuation systems and provide a teaching signal. The valuation system perspective integrates insights from motivation theories in a novel way and demonstrates how complex motivational phenomena can be characterized as emerging from basic perception and action processes.

Valuation Systems

To trace how motivation emerges during behavior, we begin with a functional analysis of the valuation systems that produce behavior. By valuation system, we mean any mental system that represents the world and prompts action to help an individual to transition toward more valued states of the world. The mind can be viewed as a collection of different valuation systems, many of which are active and interactive most of the time (Gross, 2015). For instance, evolutionarily older systems involved in producing automatic behavior are complemented by evolutionarily younger systems producing flexible behavior (Evans & Stanovich, 2013; Rangel, Camerer, & Montague, 2008). Likewise, more specialized systems involved in dealing with particular challenges are complemented by more domain-general systems (Cosmides & Tooby, 2013).

To characterize the broad set of different valuation systems in common terms, we turn to a functional analysis. Grounded in an understanding of the problems that a set of systems can solve, a functional analysis seeks to identify general operating principles of these systems on a computational level, overlooking, initially at least, algorithmic and implementational details (Marr, 1982). For instance, a functional analysis of braking systems would reveal that all braking systems address the problem of how to slow a vehicle by converting kinetic energy into another type of energy. These insights characterize braking systems irrespective of their underlying algorithms (e.g., friction or regeneration) and implementations (e.g., steel or carbon fiber), thereby providing a common set of concepts for thinking about different braking systems. Our aim is to find a comparable common set of concepts for thinking about different valuation systems.

As with any functional analysis, we start by asking what problems valuation systems address. Broadly, these systems produce behavior that helps an individual to approach rewarding and to avoid punishing configurations of the internal and external environment. To do this, the valuation systems need to solve two basic problems—the perception problem of building a serviceable map of the world while relying only on fragmented sensory input and the action problem of finding situation-specific means to desired ends.

The *perception problem* arises because the mind lacks direct knowledge of the world. It receives information through an array of sensors that transform isolated features of the internal and external environment into streams of noisy data. For instance, single features of fruits, such as their size, color, or location, may all fail to reliably distinguish edible from inedible fruits. In order to act adaptively, valuation systems need some understanding of the structure of the world, such as the objects of edible and inedible fruits. Solving the perception problem therefore requires extracting the adaptively relevant structure of the world.

The *action problem* arises because an action that is appropriate in one place or time may not be appropriate in another place or time. For example, just because looking near a tree for food worked well last time does not mean it will work well this time. Trees do not carry fruit all of the time, and not all trees carry edible fruit.

Solving the action problem thus requires flexibly producing different actions in different situations. This is because it would be difficult to solve this problem by relying solely on rigid links between stimuli and actions (e.g., reflexes) or between needs and actions (e.g., instincts). Solving the action problem therefore requires acting in accordance with the structure of the world.

Formulating the perception and action problems helps to identify the operating principles that valuation systems use to solve these problems. In the sections that follow, we argue that valuation systems solve both problems by combining *hierarchical mental models* that represent the structure of the world by conjoining simpler models into increasingly elaborate ones with *hierarchical feedback control* processes that minimize gaps between pairs of models by altering one of them.

Hierarchical Mental Models

Mental representations are neural patterns that stand in for different pieces of information in some computation (Pouget, Dayan, & Zemel, 2000). Some mental representations are mental models that stand in for multimodal states that the world can take. The term *world* is used broadly here, to denote the environment both outside and inside the individual, and the term *state* is used to denote a multimodal configuration of the internal and external environment. States of the world can therefore include places like a grocery store, beings like a cashier, or objects like an apple. They can also include bodily states such as hunger and mental states such as a plan to get some apples.

Most mental models rely on hierarchical abstraction, whereby more elaborate models are formed by conjoining a number of less elaborate models (Fig. 6.1; Ballard, 2017; Simon, 1962). The least elaborate mental models represent embodied experiences produced by the sensory-motor repertoire of the individual (Binder et al., 2016). Embodied models on a lower layer help define less embodied semantic models on a higher layer such as "food" and "paying." As hierarchical abstraction progresses, it yields increasingly elaborate mental models including schemata, scenarios, and narratives (Baldassano, Hasson, & Norman, 2018; Binder, 2016). Elaborate models can denote whole situations or events that relate places, beings, objects, as well as mental and bodily states into a single comprehensive representation such as "grocery shopping" (Radvansky & Zacks, 2011). Abstraction hierarchies are implemented throughout the brain (Ballard, 2017; Fuster, 2017) and can be algorithmically expressed as multilayered neural networks (Lake, Ullman, Tenenbaum, & Gershman, 2017; McClelland & Rumelhart, 1981).

A key feature of mental models is their reusability. For example, there is considerable overlap between the neural patterns involved in perceiving and imagining equivalent stimuli (Lacey & Lawson, 2013) as well as between performing and imagining equivalent actions (Jeannerod, 2001; O'Shea & Moran, 2017). The same mental model can thus be used to denote a state of the world as it is experienced here and now for one computation and to denote an equivalent state of the world as it is

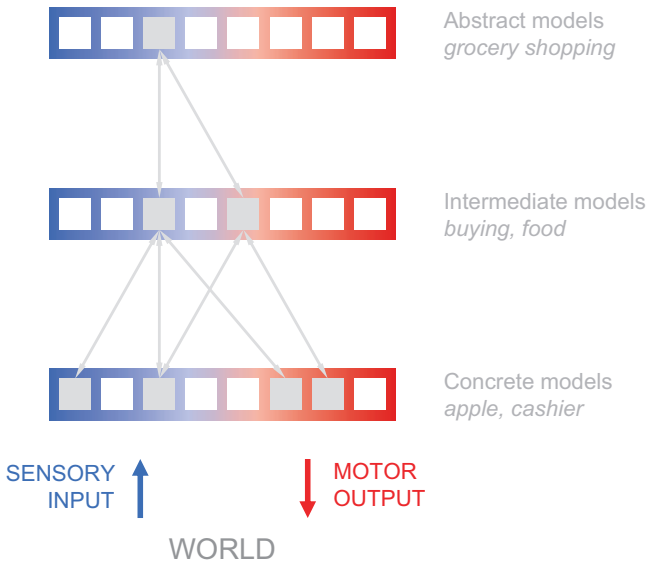

Fig. 6.1 Mental models formed through hierarchical abstraction. Mental models (squares in each row) are neural patterns denoting states that the world can take. They are formed through hierarchical abstraction whereby patterns on a lower layer denoting experienced states of the world (e.g., an apple) are linked to patterns on a higher layer denoting more abstract states of the world (e.g., grocery shopping)

mentally simulated within a different computation (Hesslow, 2012). For instance, seeing an apple within reach and wanting an apple that has yet to be found can involve the same mental model of an apple. Mental models can be reused for different purposes, including recalling how the world was, mentalizing how it might seem from another perspective, and, crucially for motivation, predicting how it might be in the near or distant future (Hesslow, 2012; Moulton & Kosslyn, 2009; Mullally & Maguire, 2014). Mental models are activated as *predictions* that denote states of the world that are probable given information arriving from—and stored knowledge about—the world. Some predictions concern imminent sensory input given how the world is believed, but not yet sensed, to be here and now (Huang & Rao, 2011; Kersten, Mamassian, & Yuille, 2004; Kok & de Lange, 2015). Other predictions concern sensory input from the states that the world is expected to take in the future (Gershman, 2018; Moulton & Kosslyn, 2009; Toomela, 2016). We assume that most mental models can be reused for either of these two versions of prediction (de Lange, Heilbron, & Kok, 2018).

Mental models play a key role in solving both the perception and action problems. They help to solve the perception problem by replacing the fragmented and variable sensory information arriving from the world with a coherent and stable perceived reality furnished by mental models. A crocodile is perceived to have sharp teeth even if its mouth is closed, because the actual sensory information about teeth, which is vulnerable to occlusion, is replaced by a mental model (Kersten et al., 2004).

However, a model is only as useful as its match to the world. It would be decidedly unhelpful to mentally model a swimming crocodile as a floating log. Solving the perception problem thus requires not only possessing mental models but also choosing the right ones to represent a given state of the world. This suggests that the mind has a way to keep track of the probability that a prediction it has made really corresponds to reality. In functional terms, the mind can be said to have a *tagging* system that captures perceptual *certainty* (Petty, Briñol, & DeMarree, 2007). For instance, the mental model of an apple will have a stronger certainty tag when it is used to perceive a graspable apple than when it is used to desire an as yet unseen apple. Certainty tagging is a functional construct that can be implemented in the brain using different neural codes (Ma & Jazayeri, 2014). The idea that activated mental models have variable certainty aligns with evidence that neural representations are often probabilistic and that awareness is often accompanied by variable degrees of certainty or confidence (Grimaldi, Lau, & Basso, 2015; Pouget, Drugowitsch, & Kepecs, 2016). There is further evidence that perceptual decisions involve accumulation of evidence in favor of competing representations until one crosses a threshold (Gold & Shadlen, 2007). This suggests that strengthening certainty tags above some threshold is what turns the tagged mental model from a prediction into part of perceived reality.

In addition to helping solve the perception problem, mental models also help to solve the action problem of choosing one of several possible means, such as pushing or pulling the door, to pursue an end, such as to enter a room. Mental models help here by representing means and ends in a common modality of *future states of the world*. Ends such as entering a room are mental models of future states that the individual is *inclined to approach or avoid* (c.f. Elliot & Fryer, 2008; Kruglanski et al., 2002; Tolman, 1925). Means are mental models of future states that are likely to result from *preforming some action*, such as a door being pushed open (Gershman, 2018; Hamilton, Grafton, & Hamilton, 2007; Hommel, Müsseler, Aschersleben, & Prinz, 2001; Ridderinkhof, 2014).

In addition to representing means and ends in a common domain, however, solving the action problem also requires choosing means that are appropriate for an end in a given context. This suggests that the mind has a way to keep track of the probability that a means would lead to a desired end. We propose, again in functional terms, that this probability is captured by bipolar *valence tags*. Specifically, a valence tag of a mental model represents the extent to which the state of the world denoted by that model (i.e., a means) would make a desired state (i.e., an end) more or less likely. The idea that activated mental models have variable valence tags aligns with evidence that most mental representations have an evaluative property of goodness vs. badness for the individual (Bargh, Chaiken, Govender, & Pratto, 1992; Carruthers, 2018; Cunningham, Zelazo, Packer, & Van Bavel, 2007; Man, Nohlen, Melo, & Cunningham, 2017).

Akin to how certainty tags above some threshold determine which mental models are perceived to be real, we argue that valence tags above some threshold determine which mental models function as action tendencies or a future state that valuation systems seek to make more likely (for positive valence tags) or less likely (for negative

valence tags) through action. This view aligns with evidence that actions are initiated in the brain not primarily as representations of motion paths or muscle movements but instead as representations of states of the world that muscle movements should produce (Adams, Shipp, & Friston, 2013; Colton, Bach, Whalley, & Mitchell, 2018; Todorov, 2004). For instance, the action tendency to grasp an apple is encoded as a valence-tagged model of the world where the apple is already in hand. Under favorable conditions, the tendency can be enacted through muscle movements believed to bring about this end state.

Action tendencies can range from very broad, such as to approach or to avoid a tagged state of the world (Krieglmeyer, De Houwer, & Deutsch, 2013; Phaf, Mohr, Rotteveel, & Wicherts, 2014), to very specific, such as to produce or refrain from a small movement. Broad action tendencies that merely indicate whether a state should be approached or avoided are usually thought of as ends. More specific action tendencies that indicate more fine-grained courses of action are usually thought of as means. However, our perspective suggests that both ends such as being in a room and means such as the door becoming open through pushing or pulling ultimately belong to the same class of action tendencies—valence-tagged predictions that valuation systems seek to make more or less likely to exist.

Our functional analysis thus far suggests that valuation systems involve mental models with variable certainty and valence tags. The strength of these tags, which can vary independently across different models as well as for the same model across different times, determine whether a model functions as a prediction, as part of perceptual reality, or as an action tendency (Fig. 6.2). To illustrate, consider a person

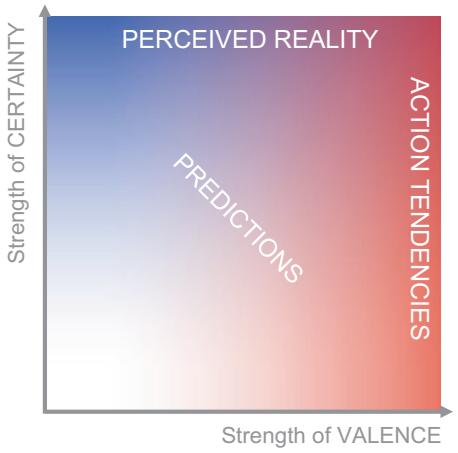

Fig. 6.2 Different functions of mental models based on variable certainty and valence tags. Many mental models are activated as *predictions* about what the situation might contain now or in the future. Valuation systems compare these predictions against sensory evidence and strengthen the certainty tags of the most accurate models, turning them into *perceived reality*. Valuation systems also compare means-like predictions to end-like predictions and strengthen the valence tags of the most effective means, turning them into *action tendencies*

entering a café. As she steps into the café, her abstract schema of a café as well as information arriving at her senses combine to activate a number of mental models. At this early stage, most of these models are *predictions* about what the room is believed but not yet confirmed to contain (e.g., there should be tea for sale here). The predictions also concern what is believed to happen at a café in the future, either via action (e.g., getting tea by ordering it) or otherwise (e.g., people will be talking). The initially weak certainty tags of such predictions are updated as more sensory evidence is accumulated, leading some predictions to be tagged certain enough to become part of perceived reality (e.g., I now smell and see tea on sale here). Meanwhile, valence tags will be transferred from broad end-like action tendencies (e.g., drink something) to increasingly specific means-like action tendencies (e.g., order a cup of green tea).

Hierarchical Feedback Control

Armed with the idea of mental models with variable certainty and valence tags, we can now ask how valuation systems activate mental models and update their certainty and valence tags. Mental models can be activated by bottom-up and top-down information flows within abstraction hierarchies (Fig. 6.1; Bar, 2007; de Lange et al., 2018; Lamme & Roelfsema, 2000). On the one hand, coarse sensory input rapidly spreads across abstraction layers where it activates various mental models of what could be causing the sensed input. For instance, from a distance, a shop front on a street could activate the models of a café, a restaurant, and a bakery. On the other hand, as each activated abstract model activates its less abstract constituent models, a parallel top-down stream of model activation ensues. For instance, the café model activates the models of people sitting at tables, while the bakery model activates the models of people queuing at the counter. As a result of the parallel bottom-up and top-down activation flows, there are usually a large number of activated mental models at any given time. Most of these models function as predictions about what might be going on in that moment as well as in the future.

The next step toward solving the perception and action problems involves updating the certainty and valence tags of the activated predictions so that only the most accurate models become parts of perceived reality and only the most desirable models become action tendencies. We suggest that both tasks can be accomplished by variations of the computational principle of *hierarchical feedback control* (Clark, 2013; Friston, 2010; Seth, 2015). Feedback control involves iteratively producing outputs that reduce a gap between an input and a target. For example, a guitar can be tuned by playing a note on one string (input), comparing it to the same note played on another string (target), and changing the tension of one of the two strings (output) until the gap between the strings is sufficiently reduced. Given that either of the strings could be tuned to reduce the gap, there are two kinds of feedback control—ascending and descending (see Fig. 6.3).

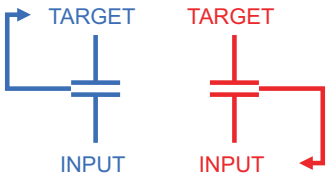

Fig. 6.3 Ascending and descending feedback control. Feedback control involves detecting a gap between an input and a target and seeking to reduce it with an output that changes the target (ascending feedback control) or the input (descending feedback control)

Ascending feedback control loops consider their lower-level input as the reference value and change their target until it matches the input. This form of feedback control helps solve the perception problem by matching mental models to sensory input. Descending feedback control loops, by contrast, consider their higher-level target as the reference value and change their input until it matches the target. This form of feedback control helps solve the action problem by matching means to ends. The computational principle of feedback control has a long history in behavioral science (Ashby, 1954; Maxwell, 1868; Miller, Galanter, & Pribram, 1960; Powers, 1973; von Uexküll, 1926; Wiener, 1948) as well as compelling algorithmic and implementation expressions including Bayesian inference (Friston, 2010; Gershman, 2019), optimal feedback control (Scott, 2004; Todorov, 2004), and reinforcement learning (Glimcher, 2011; Lee, Seo, & Jung, 2012). It is therefore a promising candidate for a functional description of the common operating principles of different valuation systems (Carver & Scheier, 2011; Gross, 2015; Pezzulo & Cisek, 2016; Seth, 2015; Stagner, 1977; Sterling, 2012).

Ascending and descending feedback control help to solve the perception and action problems, respectively, when they operate between layers of abstraction hierarchies populated by mental models (see Figs. 6.4 and 6.5). Ascending feedback control operating between abstraction layers forms *perception loops* that use bottom-up evidence to assess the accuracy of top-down predictions (Chanes & Barrett, 2016; Clark, 2013; Friston, 2010; Henson & Gagnepain, 2010; Huang & Rao, 2011). Imagine a person taking a first sip from a cup of tea she has just ordered in a café. The action of ordering the tea has activated the mental model of "hot tea" as the best guess of what her cup contains. This prediction in turn generates a top-down cascade of increasingly specific further predictions about the sensations that a cup of tea should cause, such as "hotness." Meanwhile, imagine that the drink in her cup is actually iced tea, producing the sensory observation of "coldness." As predictions such as hotness cascade downward and evidence such as coldness cascade upward along abstraction hierarchies, perception loops can harness the gaps between these information flows to solve the perception problem using ascending feedback control. Specifically, a perception loop takes top-down predictions (e.g., hotness) as its targets, compares them to the bottom-up sensory evidence (e.g., coldness) as its input, and updates the certainty tags of the predictions as its output (e.g., weaken the certainty tag of "hot tea," strengthen that of "iced tea").

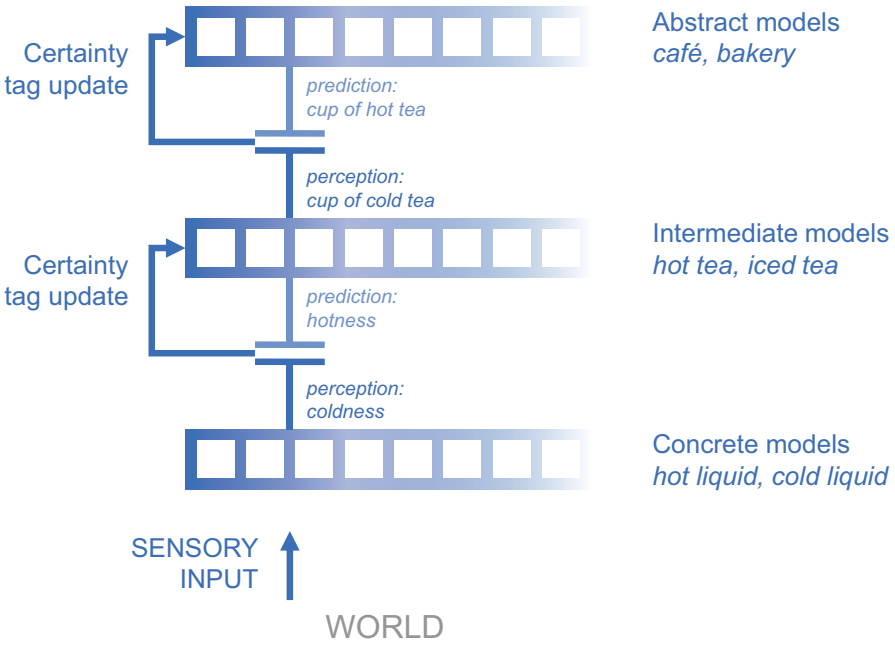

Fig. 6.4 Perception loops using ascending feedback control between successive abstraction layers to match predictions to sensory evidence. Top-down flow of information corresponds to increasingly specific sensory predictions. Bottom-up flow of information corresponds to increasingly abstract sensory evidence. Ascending feedback control loops operating between pairs of layers use gaps between predictions and evidence to update certainty tags of predictions on the upper layer

This process can be repeated until all perceptual gaps are sufficiently minimized (see Fig. 6.4).

Perception loops minimize gaps between many pairs of abstraction layers in parallel. For instance, the target layer of the previous example, where the models "hot tea" and "iced tea" reside, is simultaneously the input layer to a more abstract feedback loop whose target layer contains a schema representing how cafés work. The higher loop takes the evidence produced by the lower loop that the cup might contain iced tea as its input and compares it to predictions such as "receiving the hot tea that was ordered" produced by the schema. It detects a gap and converts it into a change to the broader schema, for instance by inferring that the barista must have misunderstood the original order to mean iced tea. Iterative and hierarchically parallel ascending feedback control can therefore underlie increasingly complex perceptual and cognitive phenomena from perception to categorization, attribution, judgment, and so forth (Clark, 2013; Friston, 2010; Seth, 2015).

Mirroring how ascending feedback control loops address the perception problem, descending feedback control loops address the action problem of selecting situation-specific means to valence-tagged ends. Action loops work with predictions that represent how the world ought to be in the future (i.e., ends) and how it

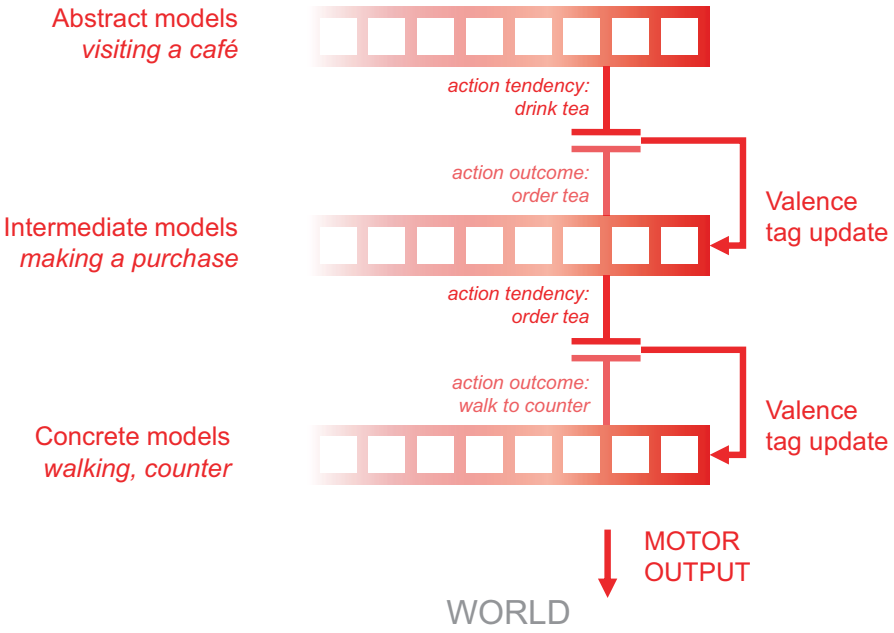

Fig. 6.5 Action loops using descending feedback control between successive abstraction layers to match means to ends. Top-down flow of information corresponds to increasingly specific action tendencies. Bottom-up flow of information corresponds to increasingly abstract action outcomes. Descending feedback control loops operating between pairs of layers use gaps between action tendencies and action outcomes to update valence tags of action outcomes on the lower layer

would be if different action tendencies were enacted (i.e., means). The computational task for the action loop is to strengthen the action tendencies that promise to be most effective means to an end in a given situation. This can be done by running descending feedback control loops between hierarchical layers of mental models (Adams et al., 2013; Shadmehr, Smith, & Krakauer, 2010; Todorov, 2004). The target positions of such loops are occupied by ends, such as the broad action tendency to "drink tea" that might be activated when a person enters a café (Fig. 6.5). The input to such loops consists of the expected outcomes of specific actions afforded by the situation, such as ordering different beverages from the barista. The action loop can now detect gaps between the end state (drink tea) and the predicted action outcomes (getting the ordered tea vs. getting the ordered coffee) and update the valence tags of the actions that yield the smallest gap (strengthen the positive tag for ordering tea, weaken the positive tag for ordering coffee).

Action loops minimize gaps between many abstraction layers in parallel. This is helpful for implementing relatively abstract action tendencies such as "drink tea" that can require different combinations of specific means depending on the characteristics of a situation, such as whether orders are taken at the table or at the counter in a particular café. Once a relatively abstract action loop has valence tagged an

action outcome such as "order tea" as an effective means to the end of "drink tea," a less abstract action loop can treat "order tea" as its end state and find that it would be served well by a means such as being over at the counter. An even less abstract loop may then valence-tag walking as a suitable means toward the end of being at the counter and so forth. Conversely, an end such as "drink tea" may itself have become an action tendency within an action loop serving a more abstract end such as adhering to the social convention of ordering something in a café. In effect, descending feedback control extends the valence tags from more to less abstract predictions until a way to change the world is found (Fishbach, Shah, & Kruglanski, 2004). This operating principle allows action loops to flexibly identify effective courses of action to strive for end states across different and changing situations.

Perception and action processes are deeply interwoven (Hamilton et al., 2007; Hommel et al., 2001; Ridderinkhof, 2014). For instance, perception makes use of simulated action outcomes to infer how different states of the world might have come about (Hesslow, 2012). Similarly, action makes use of perceived action outcomes to fine-tune motor control (Todorov, 2004). We therefore view each *valuation system* as a collection of functionally coupled perception and action loops (Fig. 6.6). A primary manifestation of perception-action coupling within a valuation system is the emergence of action affordances or perceived opportunities for action a situation offers (Cisek, 2007; Gibson, 1954). In functional terms, action affordances are a series of predictions that are deemed reasonably probable by perception loops and are also linked into a means-ends chain by action loops. For instance, the

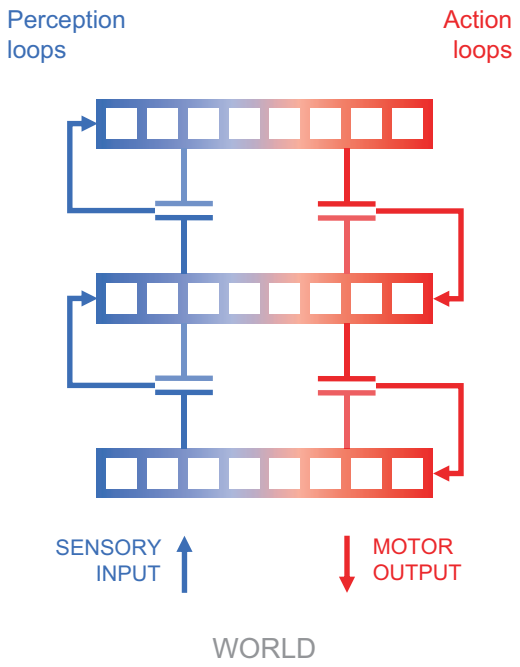

Fig. 6.6 A valuation system. A set of functionally coupled perception and action loops can be thought of as a valuation system. The system operates with a commonly accessible pool of mental models (squares in each row) activated across different layers of abstraction hierarchies. Action affordances emerge from valuation systems as perception loops activate models of states that may follow the current one and action loops organize these predictions into means-ends chains

model of drinking tea functions as an action affordance if it is deemed a probable occurrence in a café by a perception loop and is also related to action tendencies such as ordering tea and walking to the counter by an action loop. Detection of action affordances thus requires perception loops to predict different action outcomes and action loops to organize them into effective means-end chains.

We have now defined a single valuation system, consisting of coupled perception and action loops that minimize gaps between mental models to solve key adaptive problems. This sketch of the complex dynamics underlying behavior remains incomplete, however, as we also need to consider that behavior usually emerges from several valuation systems acting and interacting in parallel. The existence of many different valuation systems may reflect the evolution of the brain as an expanding set of fairly compartmentalized solutions to fairly circumscribed problems, in addition to a suite of shared and domain-general cognitive resources (Cosmides & Tooby, 2013; Pinker, 1999). Rather than being inefficient, this setup may in fact provide flexibility and robustness to behavior control (Sterling, 2012). Overt behavior may therefore be best thought of as a distributed consensus between valuation systems focusing on different features of the world as well as on different kinds of end states (Cisek, 2012; Hunt & Hayden, 2017; O'Doherty, 2014; Vickery, Chun, & Lee, 2011). This principle is illustrated by functional specialization in the prefrontal cortex between regions evaluating information from different sources such as exteroceptive and interoceptive senses, visceral and skeletal motor systems, episodic simulation, and metacognitive representations of actions, emotions, and the self (Dixon, Thiruchselvam, Todd, & Christoff, 2017).

Given the existence of different valuation systems, how can their contributions be integrated without producing contradictory behavior, such as someone reaching simultaneously for an apple and a chocolate bar, and failing to grasp either? One possibility is that behavioral consistency emerges from competitions between mental models. Both perceptual and action decisions appear to involve sequential accumulation of "evidence" in favor of alternatives until one crosses a threshold and emerges as a discrete winner (Bogacz, 2007; Gold & Shadlen, 2007; Ratcliff, Smith, Brown, & McKoon, 2016; Yoo & Hayden, 2018). Within our perspective, this implies that discrete perception of some models as real emerges from sequential accumulation of certainty tags and discrete commitment to act emerges from sequential accumulation of valence tags. For instance, a decision to grasp an apple over a chocolate bar can ensue when the valence tag on the action tendency to grasp an apple reaches a decision threshold sooner than the valence tag on the tendency to grasp the chocolate bar. Notably, competitions can occur at different abstraction layers in parallel. For instance, in parallel with the competition between grasping action tendencies, another competition may have occurred on a higher layer of abstraction between end-like tendencies such as eating something tasty or eating something healthy. As valuation systems facilitate both ascending and descending information flows, the competitions on different layers influence each-other. For instance, as the tendency to eat in a healthy manner is strengthened, it will function as one source of the evidence that can tip the competition between grasping actions in favor of grasping for the apple rather than the chocolate bar.

The Emergence of Motivation

We have now considered how valuation systems are formed when hierarchical mental models are combined with feedback control operating in perception and action loops. From our valuation systems perspective, it is these dynamic interactions within and between valuation systems that give rise to the emergent motivational properties of force and direction.

Complex dynamics can give rise to emergent properties and often do so across many levels of increasing complexity. To take an example from the physical domain, some properties of water—such as adhesion to other molecules—emerge as soon as hydrogen and oxygen atoms form a water molecule. By contrast, other properties of water—such as the orderly structure of ice crystals—emerge from interactions involving larger numbers of water molecules. In a similar fashion, behavior can be characterized by different kinds of force and direction that emerge along a gradient of complexity.

In the sections that follow, we identify motivational phenomena that emerge at three levels of complexity along this gradient. Each level corresponds to a broad section of the gradient, and transitions between the levels are gradual. On the first *inherent motivation* level, predictability and competence motives emerge from aggregated gap reduction imperatives of the perception and action loops, respectively. On the second *intentional motivation* level, goal commitment and goal pursuit cycles emerge from synchronized valuation systems. On the third *identity motivation* level, identity and self-regulation emerge from synchronized goal pursuit cycles.

As we consider each of these emergent phenomena, we will argue that they can become reflected in conscious awareness as affective feelings that orchestrate system-wide responses and facilitate learning from experience (Carver & Scheier, 1990; Chang & Jolly, 2018; Lang & Bradley, 2010; Pessoa, 2018; Weiner, 1985). We consider feelings to be affective when they contain the evaluative property of goodness vs badness. This suggests that all affective feelings reflect the valence tags of relevant mental models, as they are retrieved and updated. However, not all valence tags are reflected in feelings as valence tags can also be retrieved and updated outside conscious awareness.

From the perspective of understanding motivation, affective feelings have two important functions. First, they can modulate several distributed valuation systems at once. For instance, affective feelings prioritize relevant world states within mental competitions (Frijda, 2009), constrict and broaden the scope of information processing (Gable & Harmon-Jones, 2010), and make certain action families more or less prepotent (Frijda, 1987). Affective feelings are therefore one way in which emergent motivational phenomena can influence the processes they emerge from. The second function of affective feelings is to produce a learnable piece of information. As a conscious reflection of an otherwise hidden process, an affective feeling makes motivational phenomena part of the world that can be explained by mental models (Barrett, 2017; Seth, 2013). As such models are stored in memory, they can

influence the operation of valuation systems upon future encounters of similar situations. For instance, a memory trace of relaxation brought about by a cup of tea can strengthen an action tendency to have another cup of tea in the future.

Inherent Motivation: Predictability and Competence

The first novel feature to emerge from a constellation of valuation systems involves an aggregation of the gap reduction imperatives within individual feedback control loops into the motives of predictability and competence (c.f. Mineka & Hendersen, 1985). Feedback control loops generate elemental motivational force by transforming otherwise inert differences between mental models into changes to certainty and valence tags. Individually, each change generated in this way may fall short of being a consistent form of motivation, as it may not become manifest in behavior. Collectively, however, the outputs of all active feedback control loops give rise to the emergent motives of predictability and competence.

The predictability motive emerges from the aggregate imperatives to minimize gaps in perception loops. This motive manifests as a desire to understand the world, over and above any desire to influence it. Constructs that overlap with the predictability motive include epistemic motivation (De Dreu, Nijstad, & van Knippenberg, 2008) and the needs for optimal predictability (Dweck, 2017), for confidence (Cialdini & Goldstein, 2004), for cognition (Cacioppo & Petty, 1982), for closure (Kruglanski & Webster, 1996), and for understanding (Stevens & Fiske, 1995). Our perspective suggests that these constructs relate to an imperative to minimize perceptual gaps by finding mental models that explain information arriving from the world. Sometimes, sufficiently accurate models can simply be retrieved from memory. This in itself can be motivating as indicated by the allure of quizzes and crossword puzzles. At other times, new models need to be constructed by combining new information with information that is already known. The predictability motive therefore also contributes to behaviors that facilitate the development of mental models such as strategic observation and intuitive experimentation (Gopnik & Schulz, 2007).

The competence motive emerges from the aggregate imperative to minimize gaps in action loops. This motive manifests as a desire to be able to impact the world over and above any ensuing rewards and punishments (Abramson, Seligman, & Teasdale, 1978; Bandura, 1977; Leotti, Iyengar, & Ochsner, 2010; Skinner, 1996). Constructs that overlap with the competence motive include the needs for competence (Deci & Ryan, 2000; Dweck, 2017), for achievement (McClelland, Atkinson, Clark, & Lowell, 1953), for control (Burger & Cooper, 1979), and for effectance (Stevens & Fiske, 1995; White, 1959). Our perspective suggests that these constructs relate to an imperative to minimize action gaps by finding effective means to various ends. Over short time scales, this can be accomplished without overt action, by computations within valuation systems that organize scattered predictions into coherent means-ends chains, or action affordances. This in itself can be motivating

as indicated by the aversion people feel to situations where their freedom to act is restricted. Over longer time scales, minimizing action gaps requires overt action and feedback to acquire and hone new skills. The competence motive therefore contributes to behaviors that facilitate skill acquisition such as play and exploration (Pellegrini, 2009).

Interestingly, people's preferences for predictability and competence appear to taper off above some optimal level (Dweck, 2017). For instance, people tend to like music in which they can predict many but not all changes in melody and rhythm (Eerola, 2016). Similarly, people tend to enjoy games in which they have a good but not perfect control over winning (Abuhamdeh & Csikszentmihalyi, 2012). It appears that neither complete predictability nor complete competence is necessarily desirable. One explanation for this is that the predictability and competence motives are satisfied not only by the state of minimized perception and action gaps but also by the progress in minimizing them. Focusing only on the size of perception and action gaps, and not on their dynamics, can be short-sighted as it may preclude the individual from exploring new environments and acquiring new skills. For instance, playing one level of a multilevel computer game over and over again would soon provide minimal perception and action gaps as event sequences and action outcomes become fully known. However, sticking to one level would preclude the player from discovering new environments and acquiring new skills. Players' general eagerness to progress to new levels suggests that people are motivated by progressive decreases in perception and action gaps not only by their low levels. The nonlinearity of the predictability and competence motives may therefore help maintain a balance between exploiting and exploring the environment (Cohen, McClure, & Yu, 2007; Friston et al., 2015).

As they emerge from distributed valuation systems, predictability and competence motives can give rise to affective feelings. Conscious reflections of the predictability motive include the feelings of surprise, confusion and curiosity that have been associated with directing cognitive resources toward understanding (D'Mello, Lehman, Pekrun, & Graesser, 2014; Loewenstein, 1994; Silvia, 2008; Wessel, Danielmeier, Morton, & Ullsperger, 2012). Our perspective suggests that these feelings reflect to-be-minimized gaps within perception loops. Surprise, elicited by unexpected events, should correspond to perception gaps caused by sensory evidence contradicting recent predictions about the future. Confusion and curiosity, by contrast, should correspond to perception gaps caused by sensory evidence contradicting predictions about the present, i.e., difficulties in finding mental models that would explain the current state of the world.

Conscious reflections of the competence motive may include the feelings of frustration and boredom that have been associated with regulation of effort and exploration (Geana, Wilson, Daw, & Cohen, 2016; Louro, Pieters, & Zeelenberg, 2007; Westgate & Wilson, 2018). Our perspective relates these feelings to gaps within action loops. Frustration should arise from gaps remaining within action loops because of difficulties in detecting feasible action affordances or means-end chains that would take the individual from the current state of affairs toward some end state. Boredom, by contrast, should arise when the gaps in action loops

are minimized to such a high degree that the individual runs the risk of missing opportunities to learn new skills, i.e., of sacrificing exploration to exploitation (Geana et al., 2016).

Predictability and competence motives form the first level of motivation to emerge from distributed valuation systems. This level is relatively low on the gradient of complexity as predictability and competence motives arise from simple aggregation of the gap reduction imperatives within perception and action loops. Over and above predictability and competence, people prefer to understand some things more than others and to be competent in some activities more than in others. We argue that these motives result from the more complex forms of motivation relating to goals and identity, which we will consider in the next two subsections.

Intentional Motivation: From Goal Commitment to Goal Pursuit

The motivational phenomena emerging on the second level along the gradient of complexity range from goal commitment to goal pursuit. Goal commitment, or incentive salience (Berridge, 2018), is what distinguishes the few goals that dominate behavior at a given time from the many other potential goals or ends that action loops also consider. Goal pursuit is the relatively coherent and persistent behavior aimed at reducing goal gaps between the current and desired states of the world (Moskowitz & Grant, 2009). In this section, we suggest that synchronous certainty and valence tagging across several valuation systems give rise to committed goals and goal pursuit cycles.

At any given time, people are committed to pursue only a subset of activated action tendencies—the future states with valence tags suggesting they should be approached or avoided (Elliot & Fryer, 2008; Klein, Wesson, Hollenbeck, & Alge, 1999). For example, a person in a café might exhibit action tendencies to "talk to people" as well as to "read the news" but become committed to only one of these goals. What determines which one? Expectancy-value accounts of motivation suggest that people generally commit to end states that are sufficiently valuable as well as sufficiently probable (Atkinson, 1957; Eccles & Wigfield, 2002; Hull, 1932; Steel & König, 2006; Weiner, 1985). Expressed in terms of our perspective, committed goals are therefore predictions with sufficiently strong certainty and valence tags. For a prediction such as "talk to people" to emerge as a goal, it thus needs to be considered sufficiently probable by perception loops as well as a sufficiently feasible means toward some end by action loops. Crucially, the more ends a prediction serves, the stronger its valence tag can be. For instance, "talk to people" may win commitment over "read the news" because even as both action tendencies are feasible means to the end of "avoid boredom," only talking to people is also a feasible means to the end of "find a companion." We therefore suggest that predictions generally become goals through synchronized consideration by several valuation systems.

The emergence of a goal can in turn amplify the synchronization between different valuation systems. Goal commitment is often accompanied by substantial prioritization of goal-relevant perception and action at the expense of alternatives (Landhäußer & Keller, 2012; Shah, Friedman, & Kruglanski, 2002). Our framework explains this by the synchronizing impact a prediction with strong certainty and valence tags can have on perception and action loops. As a prediction with a strong certainty tag, a committed goal activates a number of goal-relevant predictions within perception loops. For instance, someone committed to drinking tea may be imagining what drinking tea feels like, thinking about where to get tea, and recalling a recent article about the health effects of drinking tea. As a prediction with a strong valence tag, a committed goal also generates further goal-relevant action tendencies within action loops. For instance, the person wanting tea may imagine walking to one café, driving to another, and preparing tea at the office. The simultaneous impacts on perception and action loops manifest in goal-relevant information becoming more easily detected, more thoroughly processed, and more difficult to ignore, often at the expense of models that are less relevant for the goal, contributing to the related phenomena of motivated attention (Pessoa, 2015; Vuilleumier, 2015) and goal shielding (Shah et al., 2002).

Another consequence of increased synchrony between perception and action loops is the reliable emergence of a previously unavailable signal of *goal gap*. Goal gap represents the distance between how the world is perceived to be and how it is desired to be according to the goal (Chang & Jolly, 2018; Elliot & Fryer, 2008; Kruglanski et al., 2002). This signal is distinct from both the perception and action gaps that are computed within valuation systems. A goal gap compares the world as it *is* according to the most certain models to how it *should* be according to the committed goal. By contrast, a perception gap compares the world as it *might* be according to various predictions to how it *is* according to sensory evidence, and an action gap compares the world as it *would* be in some end state to how it *would* be owing to some action. A goal gap is an important additional piece of information that complements the value (valence tag) and expectancy (certainty tag) associated with a goal. The goal gap indicates how much more work and time might be needed before a goal can be attained. Integration of recent goal gap changes can further function as a speedometer indicating whether success in goal pursuit is accelerating or decelerating. These pieces of information are known to be pivotal to the force and direction with which people strive for goals (Carver & Scheier, 2011; Chang & Jolly, 2018; Louro et al., 2007).

The final unique property to emerge from distributed valuation systems on the second level of complexity is the goal pursuit cycle that implements descending feedback control to minimize goal gaps (Fig. 6.7a). Recall that descending feedback control, which is also operative within action loops, involves iteratively changing an input to minimize a gap between the input and a target. Within the goal pursuit cycle, the target position is occupied by the goal, the input position by the current state of world, and the output position by a desired change to the world. As it iterates, the goal pursuit cycle seeks to minimize the goal gap by changing the world. This function emerges from the operation of distributed valuation systems in the

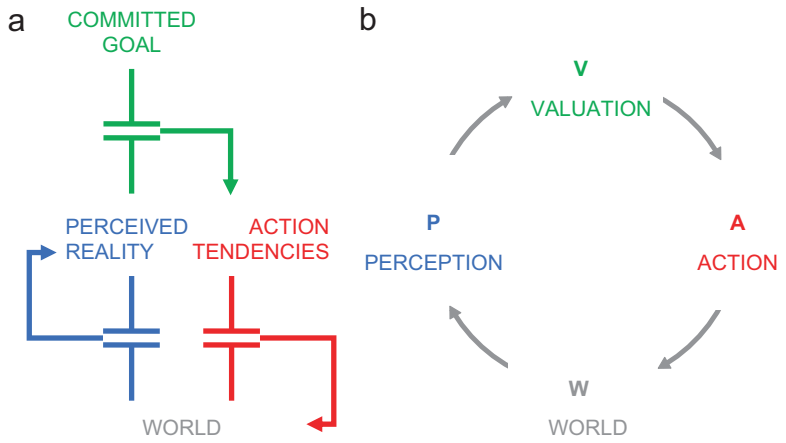

Fig. 6.7 Emergence of goal pursuit. (**a**) Goal pursuit as an emergent feedback control cycle that takes the perceived reality produced by perception loops as input, compares it to the committed goal as its target, and uses action loops to change the world as its output. (**b**) A schematic rendering of the processes in (**a**) that focuses on four iterative steps of the goal pursuit cycle

sense that the goal pursuit cycle relies on perception loops for its input and action loops for its output. The current state of world, which is compared to the goal, is produced by the collective operation of perception loops. Likewise, the change to the world that the goal pursuit cycle outputs is produced by the collective operation of action loops that translate the desired change to the world into action tendencies by valence tagging increasingly specific predictions. The goal pursuit feedback control cycle is therefore an emergent process that recapitulates the structure of descending feedback control.

The goal pursuit process can be redrawn as a simpler cycle consisting of four key steps of World, Perception, Valuation, and Action (Fig. 6.7b). Consider for instance someone committed to a goal to assemble a piece of furniture such as a shelf. The World step of the goal pursuit cycle denotes the current state of the world with a disassembled shelf. At the Perception step, mental models are found to capture goal-relevant information such as pieces of the shelf and affordances for connecting them to each other. At the Valuation step, the perceived disassembled shelf is compared to the committed goal of assembled shelf, and the gaps between the two are detected. At the Action step, action tendencies intended to reduce the goal gap are generated and, as long as the pursuit of the given goal remains a priority, enacted. Next, all steps of the loop are repeated to adjust behavior to the outcomes of the actions and other changes in the world. The loop generally iterates until the goal gap has been minimized, unless people are also motivated to maintain the absence of the gap (Ecker & Gilead, 2018). The loop can also disintegrate when the goal loses its committed status (Carver & Scheier, 2005).

These intentional motivational phenomena can be reflected in awareness through their contributions to achievement emotions such as hope and anxiety, contentment

and disappointment, or relief and despair (Harley, Pekrun, Taxer, & Gross, 2019; Pekrun, 2006; Weiner, 1985). This is because emotions rely on appraisal processes that represent the relationship between a situation and goals (Moors, 2010; Smith & Lazarus, 1993), leading to loosely orchestrated changes in the mind and the body (Barrett, Mesquita, Ochsner, & Gross, 2007; Moors, Ellsworth, Scherer, & Frijda, 2013; Mulligan & Scherer, 2012). In terms of our framework, the appraised relationships between a situation and goals overlaps with the goal gaps computed at the Valuation step of the feedback control goal pursuit cycle (Chang & Jolly, 2018; Moors, Boddez, & De Houwer, 2017; Uusberg, Taxer, Yih, Uusberg, & Gross, 2019). In particular, goal gaps are closely aligned with the appraisal of goal congruence that is strongly associated with the valence of affective feelings (Scherer, Dan, & Flykt, 2006). Specifically, positive affect is generated when the world is helpful for goals and negative affect is generated when the world is unhelpful for goals. The helpfulness assessment may also take into account the rate of goal progress, leading to positive affect when a goal is getting closer and negative affect when it is not (Carver & Scheier, 1990). Other important appraisal dimensions such as accountability and coping potential can be thought of as abstract features of the mental models that valuation systems have applied to explain the situation.

We have now seen how intentional motivation ranging from goal commitment to goal pursuit can emerge from distributed valuation systems. The synchronized combination of valence and certainty tags produces committed goals and goal pursuit feedback control cycles. This cycle focuses perception loops on the extraction of goal-relevant information and action loops on the implementation of desired changes to the world. One consequence of the emergence of goal pursuit is the temporal and cross-situational durability of the impact a committed goal has on behavior. For instance, someone who has already spent some time queuing for a concert ticket may be more resistant to giving up than someone who has not yet begun. However, the temporal durability of some goals exceeds what can be explained by intentional motivation alone, suggesting a role for a third level of motivation to emerge along the gradient of complexity that we discuss next.

Identity Motivation: From Self to Self-Regulation

The motivational phenomena to emerge on the identity motivation level include identity, or a valued sense of self (Berkman, Livingston, & Kahn, 2017), and self-regulation, or biasing of behavioral impulses serving more imminent goals in favor of pursuits of more distant goals (Berkman et al., 2017; Kotabe & Hofmann, 2015; O'Leary, Uusberg, & Gross, 2017). We propose that these motivational phenomena emerge from distributed valuation systems on the third level along the gradient of complexity. In this section, we will argue that identity and self-regulation can be seen as meta-level versions of goal commitment and goal pursuit. Specifically, we view identity as a commitment to attain certain goals and self-regulation as a feedback control pursuit of certain goal pursuits.

Identity as well as self-regulation revolve around highly abstract mental models that denote the self. Mental models of the self can be viewed as conjunctions of various other models that represent self-related information such as personal characteristics, social roles, long-term goals, and personal narratives (Dweck, 2013; Gillihan & Farah, 2005; McAdams, 2013). These self-related mental models arise within perception loops to help make sense of what is going on inside and outside of the person. Self-models with a sufficiently good match to evidence populate a person's self-awareness. Over time, some self-models can obtain persistent certainty tags and become part of perceived reality irrespective of momentary evidence, underlying a person's self-concept.

Self-models can also function as committed goals or end states that an individual seeks to turn into reality. We refer to such goals as identity. Our perspective suggests that self-models amount to identity the same way any mental model becomes a goal—by sufficiently strong valence and certainty tags. Identity includes parts of the self-concept that are persistently tagged with positive or negative valence, giving rise to the phenomenon of self-esteem (Mann, Hosman, Schaalma, & de Vries, 2004). We suggest that the need for self-coherence or the self-verification motive can be understood as a commitment to positively valenced aspects of the self-concepts (Dweck, 2017; Leary, 2007; Swann, 1982). A related but distinct component of identity is the ideal self (Higgins, 1987) which can be viewed as a set of self-models that are strongly valence tagged, insufficiently certainty tagged to already belong to the self-concept, but sufficiently certainty tagged to emerge as a committed goal. The striving for this aspect of identity overlaps with motivational phenomena such as self-enhancement and self-protection (Alicke & Sedikides, 2009; Leary, 2007).

We suggest that identity can synchronize valuation systems the same way all committed goals do and should therefore produce goal shielding and goal gaps. This prediction aligns with findings that self-related information is prioritized in various information processing stages, indicating that identity can indeed produce goal shielding (Alexopoulos, Muller, Ric, & Marendaz, 2012). Identity can also produce goal gaps, or representations of the distance between the self as it is perceived to be and how it is desired to be. For instance, people can have a strong sense of being incongruent with their self-concept (Swann, 1982) and not living up to their ideal selves (Alicke & Sedikides, 2009).

The unique feature to emerge on the third level of the gradient of complexity is recursiveness or meta-level nature of identity and self-regulation. Identity, a type of goal, can be thought of as a goal about other goals. The intentional-level goals are often in competition as people regularly juggle different pursuits in parallel. For instance, a meeting with a colleague can involve working on several agenda items, maintaining the relationship, handling of phone notifications, and dealing with bodily signals such as thirst. The number of parallel goals people care about increases with the temporal window of analysis, as we move from a moment to a day, to a week, to a year, or to the foreseeable future. Identity can be seen as one mechanism through which some goals will become prioritized over others (Berkman et al., 2017). For instance, when the meeting described above grows overwhelming,

a person who identifies with being an efficient manager but not with being a nice person may sacrifice the goal of managing the relationship in service of managing the agenda. The person's identity has therefore functioned as a goal to prioritize one goal over another.

The recursiveness of identity motivation is also visible when self-regulation is viewed as an identity pursuit gap reduction feedback loop. Self-regulation is what is needed to stop oneself from consuming pleasant substances that are harmful in the long run or to sacrifice activities with a short-term payoff such as watching TV to activities with a long-term payoff such as exercising. A common element in these situations, and a defining feature of self-regulation, is the competition between the pursuit of shorter-term goals and the pursuit of longer-term goals (Berkman et al., 2017; Duckworth, Gendler, & Gross, 2016; Kotabe & Hofmann, 2015; Van Tongeren et al., 2018). The goals or end states that self-regulation seeks to alter thus amount not to any state in the external environment but to the state of competition between different goal pursuits within the individual.

Viewing self-regulation as a form of goal pursuit suggests that self-regulation can also be analyzed as a feedback control process involving the World, Perception, Valuation, and Action steps (see Fig. 6.8). Self-regulation as a feedback control goal pursuit cycle seeks to reduce the gap between some component of identity and the perceived self. A key difference between regular goal pursuit emerging on the intentional motivation level and the meta-level goal pursuit of self-regulation is the nature of the world that these cycles seek to change. Whereas goal pursuit seeks to change the state of environment, both external and internal, self-regulation seeks to change the state of other goal pursuits. For instance, consider someone trying to overcome a craving for a tasty burger in favor of a healthy and environmentally friendly salad. Self-regulation is needed in this situation not for actually ordering the salad,

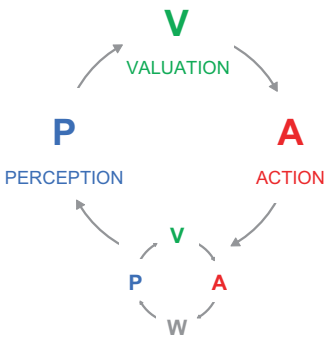

Fig. 6.8 Self-regulation as identity pursuit. Self-regulation is a feedback control goal pursuit cycle that seeks to minimize a gap between an aspect of identity and perceived state of the self. It represents the state of relevant ongoing goal pursuits at the Perception step, evaluates them in relation to identity at Valuation step, and launches regulation strategies at the Action step

which would be trivial for action loops within an intentional-level goal pursuit cycle. Self-control is needed in this situation to shift the balance among the motivational processes emerging on the intentional level to commit to the salad instead of the burger. Thus, the world that self-regulation seeks to change is the state of other goal pursuits (Fig. 6.8).

The focus on other goals is then propagated to the remaining steps of self-regulation goal pursuit cycle (Gross, 2015; O'Leary et al., 2017). The Perception step of self-regulation involves perception loops using interoceptive and other evidence to populate self-awareness with appropriate mental models, such as the concept of craving for the burger (Barrett, 2017; Seth, 2013). At the Valuation step, gaps are detected between the perceived state of ongoing goal pursuits and aspects of identity such as being a healthy and ethical person. Finally, the Action step of self-regulation includes overt and covert action that are directed at changing the state of ongoing goal pursuits, such as deliberately focusing on the negative consequences of eating the burger with the aim to reappraise its allure (Duckworth et al., 2016; Lazarus, 1993; O'Leary et al., 2017). Self-regulation can go through multiple iterations before the identity gap is minimized.

Identity motivation can also give rise to unique emotional episodes. One class of emotions emerging on this level are self-conscious emotions such as pride, shame, and guilt (Leary, 2007; Tracy & Robins, 2004). The appraisal process underlying these emotions assesses the congruence between the situation and aspects of identity, such as one's social standing. Another unique class of emotion to emerge on this level are meta-emotions or emotions arising in response to another emotion. For instance, people can feel negatively about an emotion they experience, such as anxiety, if they have appraised this emotion to be incongruent with a relevant goal such as giving a good presentation (Tamir, 2015). The affective feelings in response to internal states are important triggers of self-regulatory processes such as emotion regulation (Gross, 2015).

We have drawn parallels between committed goals and identity and between goal pursuit and self-regulation. In fact, the structure of the feedback control goal pursuit process can also help us understand the action tendencies produced by the higher-order self-regulatory process (Gross, 1998, 2015). As the goal of identity pursuit is to alter the state of concurrent goal pursuits, it can in principle alter each of the four phases of goal pursuit. First, the self-regulatory loop can alter or modify the world states that goal pursuit processes take as their input. For instance, someone wishing to avoid eating too many sweets may remove sweets from their home. Second, self-regulation can interfere with the Perception step of goal pursuit by directing attention away from thinking about sweets. Third, the self-regulatory loop can interfere with the Valuation step of goal pursuit, for instance by thinking about how a recent meal already provided a sweet experience, thereby making the goal gaps seem smaller. Finally, the self-regulatory loop can launch actions that directly target the tendencies produced by goal pursuit processes, such as suppressing the urge to get some sweets.

Conclusion

In this chapter, we have presented a valuation systems perspective on motivation. This account relies on a functional analysis of valuation systems that combine mental models with hierarchical feedback control to solve the perception and action problems associated with producing adaptive behavior in a dynamic, rapidly changing world. The perception problem is solved by perception loops that populate perceptual reality with predictions that do the best job of explaining sensory evidence. The action problem is solved by action loops that generate action tendencies by identifying the means that do the best job of approximating end states. Motivational force and direction emerge from the dynamic interactions within and between valuation systems at three broad levels along a gradient of complexity. Inherent motives of predictability and competence arise from aggregated gaps within perception and action loops, respectively. Intentional motivation arises as predictions with sufficient certainty and valence tags become committed goals that synchronize valuation systems and give rise to a goal pursuit cycle that uses descending feedback control to minimize goal gaps. Identity motivation arises from further synchronization of valuation systems into identity, or goal about goals, and self-regulation, or feedback control of goal pursuits. Each of these levels can also give rise to affective feelings that can regulate distributed valuation systems and function as teaching signals.

Motivation as viewed from the valuation systems perspective has three broad characteristics. First, motivation is *emergent*. There is no stage in the unfolding of behavior at which the motive to act is fully formed and then merely implemented. Instead, action affordances detected by valuation systems are converted to action tendencies across several competing valuation systems. Second, motivation is *constructive* as it arises neither from the environment nor the person in isolation but from an active negotiation between the two within valuation systems. Our perception of the world, of our own goals, and of afforded actions relies on the mental models that perception systems have generated over time and in the moment. This suggests that the mental models we bring to a situation have a substantial impact on the motivation we experience (Dweck, 2017). Third, motivation is *allostatic*. While homeostatic control seeks to maintain a fixed state of a system, allostatic control seeks to flexibly adjust the state of the system in anticipation of changes in the world (Barrett, 2017; Sterling, 2012; Toomela, 2016). Action loops enact allostatic control by guiding behavior toward predictive mental models across multiple layers of complexity. Taken together, we hope these ideas help move us toward an integrative perspective on motivation.

References

Abramson, L. Y., Seligman, M. E., & Teasdale, J. D. (1978). Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology, 87*, 49–74.

Abuhamdeh, S., & Csikszentmihalyi, M. (2012). The importance of challenge for the enjoyment of intrinsically motivated, goal-directed activities. *Personality and Social Psychology Bulletin, 38*, 317–330. https://doi.org/10.1177/0146167211427147

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: Active inference in the motor system. *Brain Structure and Function, 218*, 611–643. https://doi.org/10.1007/s00429-012-0475-5

Alexopoulos, T., Muller, D., Ric, F., & Marendaz, C. (2012). I, me, mine: Automatic attentional capture by self-related stimuli. *European Journal of Social Psychology, 42*, 770–779. https://doi.org/10.1002/ejsp.1882

Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology, 20*, 1–48. https://doi.org/10.1080/10463280802613866

Ashby, W. R. (1954). *Design for a brain*. New York, NY: John Wiley & Sons Inc.

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review, 64*, 359–372. https://doi.org/10.1037/h0043445

Baldassano, C., Hasson, U., & Norman, K. A. (2018). Representation of real-world event schemas during narrative perception. *Journal of Neuroscience, 38*, 9689–9699. https://doi.org/10.1523/JNEUROSCI.0251-18.2018

Ballard, D. H. (2017). *Brain computation as hierarchical abstraction*. Cambridge, MA: The MIT Press.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences, 11*, 280–289. https://doi.org/10.1016/j.tics.2007.05.005

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology, 62*, 893–912. https://doi.org/10.1037/0022-3514.62.6.893

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience, 12*, 1–23. https://doi.org/10.1093/scan/nsw154

Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annual Review of Psychology, 58*, 373–403. https://doi.org/10.1146/annurev.psych.58.110405.085709

Berkman, E. T., Livingston, J. L., & Kahn, L. E. (2017). Finding the "self" in self-regulation: The identity-value model. *Psychological Inquiry, 28*, 77–98. https://doi.org/10.1080/1047840X.2017.1323463

Berridge, K. C. (2018). Evolving concepts of emotion and motivation. *Frontiers in Psychology, 9*, 1647. https://doi.org/10.3389/fpsyg.2018.01647

Binder, J. R. (2016). In defense of abstract conceptual representations. *Psychonomic Bulletin & Review, 23*, 1096–1108. https://doi.org/10.3758/s13423-015-0909-1

Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cognitive Neuropsychology, 33*, 130–174. https://doi.org/10.1080/02643294.2016.1147426

Bogacz, R. (2007). Optimal decision-making theories: Linking neurobiology with behaviour. *Trends in Cognitive Sciences, 11*, 118–125. https://doi.org/10.1016/j.tics.2006.12.006

Braver, T. S. (Ed.). (2016). *Motivation and cognitive control*. New York, NY: Routledge.

Burger, J. M., & Cooper, H. M. (1979). The desirability of control. *Motivation and Emotion, 3*, 381–393. https://doi.org/10.1007/BF00994052

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*, 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Carruthers, P. (2018). Valence and value. *Philosophy and Phenomenological Research, 97*, 658–680. https://doi.org/10.1111/phpr.12395

Carver, C. S., & Scheier, M. F. (1990). Origins and functions of positive and negative affect: A control-process view. *Psychological Review, 97*, 19–35. https://doi.org/10.1037/0033-295X.97.1.19

Carver, C. S., & Scheier, M. F. (2005). Engagement, disengagement, coping, and catastrophe. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 527–547). New York, NY: Guilford Press.

Carver, C. S., & Scheier, M. F. (2011). Self-regulation of action and affect. In K. D. Vohs & R. F. Baumeister (Eds.), *Handbook of self-regulation: Research, theory, and applications* (2nd ed., pp. 13–39). New York, NY: Guilford Press.

Chanes, L., & Barrett, L. F. (2016). Redefining the role of limbic areas in cortical processing. *Trends in Cognitive Sciences, 20*, 96–106. https://doi.org/10.1016/j.tics.2015.11.005

Chang, L. J., & Jolly, E. (2018). Emotions as computational signals of goal error. In A. S. Fox, R. C. Lapate, A. J. Shackman, & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (2nd ed., p. 21). New York, NY: Oxford University Press.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology, 55*, 591–621. https://doi.org/10.1146/annurev.psych.55.090902.142015

Cisek, P. (2007). Cortical mechanisms of action selection: The affordance competition hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*, 1585–1599. https://doi.org/10.1098/rstb.2007.2054

Cisek, P. (2012). Making decisions through a distributed consensus. *Current Opinion in Neurobiology, 22*, 927–936. https://doi.org/10.1016/j.conb.2012.05.007

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences, 36*, 181–204. https://doi.org/10.1017/S0140525X12000477

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*, 933–942. https://doi.org/10.1098/rstb.2007.2098

Colton, J., Bach, P., Whalley, B., & Mitchell, C. (2018). Intention insertion: Activating an action's perceptual consequences is sufficient to induce non-willed motor behavior. *Journal of Experimental Psychology: General, 147*, 1256–1263. https://doi.org/10.1037/xge0000435

Corr, P. J., DeYoung, C. G., & McNaughton, N. (2013). Motivation and personality: A neuropsychological perspective. *Social and Personality Psychology Compass, 7*, 158–175. https://doi.org/10.1111/spc3.12016

Cosmides, L., & Tooby, J. (2013). Evolutionary psychology: New perspectives on cognition and motivation. *Annual Review of Psychology, 64*, 201–229. https://doi.org/10.1146/annurev.psych.121208.131628

Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The iterative reprocessing model: A multilevel framework for attitudes and evaluation. *Social Cognition, 25*, 736–760. https://doi.org/10.1521/soco.2007.25.5.736

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction, 29*, 153–170. https://doi.org/10.1016/j.learninstruc.2012.05.003

De Dreu, C. K. W., Nijstad, B. A., & van Knippenberg, D. (2008). Motivated information processing in group judgment and decision making. *Personality and Social Psychology Review, 12*, 22–49. https://doi.org/10.1177/1088868307304092

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences, 22*, 764–779. https://doi.org/10.1016/j.tics.2018.06.002

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*, 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Dixon, M. L., Thiruchselvam, R., Todd, R., & Christoff, K. (2017). Emotion and the prefrontal cortex: An integrative review. *Psychological Bulletin, 143*, 1033–1081. https://doi.org/10.1037/bul0000096

Duckworth, A. L., Gendler, T. S., & Gross, J. J. (2016). Situational strategies for self-control. *Perspectives on Psychological Science, 11*, 35–55. https://doi.org/10.1177/1745691615623247

Dunning, D. (Ed.). (2011). *Social motivation*. New York, NY: Psychology Press.

Dweck, C. S. (Ed.). (2013). *Self-theories: Their role in motivation, personality, and development* (2nd ed.). New York, NY: Psychology Press.

Dweck, C. S. (2017). From needs to goals and representations: Foundations for a unified theory of motivation, personality, and development. *Psychological Review, 124*, 689–719. https://doi.org/10.1037/rev0000082

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*, 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Ecker, Y., & Gilead, M. (2018). Goal-directed allostasis: The unique challenge of keeping things as they are and strategies to overcome it. *Perspectives on Psychological Science, 13*, 618–633. https://doi.org/10.1177/1745691618769847

Eerola, T. (2016). Expectancy-violation and information-theoretic models of melodic complexity. *Empirical Musicology Review, 11*, 2–17. https://doi.org/10.18061/emr.v11i1.4836

Elliot, A. J., & Fryer, J. W. (2008). The goal construct in psychology. In J. Y. Shah & W. L. Gardner (Eds.), *Handbook of motivation science*. New York, NY: Guilford Press.

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*, 223–241. https://doi.org/10.1177/1745691612460685

Fishbach, A., Shah, J. Y., & Kruglanski, A. W. (2004). Emotional transfer in goal systems. *Journal of Experimental Social Psychology, 40*, 723–738. https://doi.org/10.1016/j.jesp.2004.04.001

Fox, A. S., Lapate, R. C., Shackman, A. J., & Davidson, R. J. (2018). *The nature of emotion: Fundamental questions* (2nd ed.). New York, NY: Oxford University Press.

Frijda, N. H. (1987). Emotion, cognitive structure, and action tendency. *Cognition and Emotion, 1*, 115–143. https://doi.org/10.1080/02699938708408043

Frijda, N. H. (2009). Emotion experience and its varieties. *Emotion Review, 1*, 264–271. https://doi.org/10.1177/1754073909103595

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11*, 127–138. https://doi.org/10.1038/nrn2787

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience, 6*, 187–214. https://doi.org/10.1080/17588928.2015.1020053

Fuster, J. M. (2017). Prefrontal cortex in decision-making: The perception–action cycle. In J.-C. Dreher & L. Tremblay (Eds.), *Decision neuroscience* (pp. 95–105). San Diego, CA: Academic Press. https://doi.org/10.1016/B978-0-12-805308-9.00008-7

Gable, P., & Harmon-Jones, E. (2010). The motivational dimensional model of affect: Implications for breadth of attention, memory, and cognitive categorisation. *Cognition & Emotion, 24*, 322–337. https://doi.org/10.1080/02699930903378305

Geana, A., Wilson, R. C., Daw, N., & Cohen, J. D. (2016). Boredom, information-seeking and exploration. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 1751–1756).

Gershman, S. J. (2018). The successor representation: Its computational logic and neural substrates. *Journal of Neuroscience, 38*, 7193–7200. https://doi.org/10.1523/JNEUROSCI.0151-18.2018

Gershman, S. J. (2019). What does the free energy principle tell us about the brain? *ArXiv: 1901.07945 [q-Bio]*. Retrieved from http://arxiv.org/abs/1901.07945

Gibson, J. J. (1954). The visual perception of objective motion and subjective movement. *Psychological Review, 61*, 304–314.

Gillihan, S. J., & Farah, M. J. (2005). Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. *Psychological Bulletin, 131*, 76–97.

Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences, 108*, 15647–15654. https://doi.org/10.1073/pnas.1014269108

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience, 30*, 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038

Gopnik, A., & Schulz, L. (Eds.). (2007). *Causal learning*. New York, NY: Oxford University Press.

Grimaldi, P., Lau, H., & Basso, M. A. (2015). There are things that we know that we know, and there are things that we do not know we do not know: Confidence in decision-making. *Neuroscience & Biobehavioral Reviews, 55*, 88–97. https://doi.org/10.1016/j.neubiorev.2015.04.006

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*, 271–299.

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry, 26*, 1–26. https://doi.org/10.1080/1047840X.2014.940781

Hamilton, A. F., Grafton, S. T., & Hamilton, A. (2007). The motor hierarchy: From kinematics to goals and intentions. In P. Haggard, Y. Rossetti, & M. Kawato (Eds.), *Sensorimotor foundations of higher cognition* (pp. 381–408). New York, NY: Oxford University Press.

Harley, J. M., Pekrun, R., Taxer, J. L., & Gross, J. J. (2019). Emotion regulation in achievement situations: An integrated model. *Educational Psychologist., 54*, 1–21.

Heckhausen, J. (Ed.). (2000). *Motivational psychology of human development: Developing motivation and motivating development*. Amsterdam, The Netherlands: Elsevier.

Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus, 20*, 1315–1326. https://doi.org/10.1002/hipo.20857

Hesslow, G. (2012). The current status of the simulation theory of cognition. *Brain Research, 1428*, 71–79. https://doi.org/10.1016/j.brainres.2011.06.026

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review, 94*, 319–340. https://doi.org/10.1037/0033-295X.94.3.319

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*, 849–878.

Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*, 580–593. https://doi.org/10.1002/wcs.142

Hull, C. L. (1932). The goal-gradient hypothesis and maze learning. *Psychological Review, 39*, 25–43. https://doi.org/10.1037/h0072640

Hunt, L. T., & Hayden, B. Y. (2017). A distributed, hierarchical and recurrent framework for reward-based choice. *Nature Reviews Neuroscience, 18*, 172–182.

Jeannerod, M. (2001). Neural simulation of action: A unifying mechanism for motor cognition. *NeuroImage, 14*, S103–S109. https://doi.org/10.1006/nimg.2001.0832

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology, 55*, 271–304. https://doi.org/10.1146/annurev.psych.55.090902.142005

Klein, H. J., Wesson, M. J., Hollenbeck, J. R., & Alge, B. J. (1999). Goal commitment and the goal-setting process: Conceptual clarification and empirical synthesis. *Journal of Applied Psychology, 84*, 885. https://doi.org/10.1037/0021-9010.84.6.885

Kok, P., & de Lange, F. P. (2015). Predictive coding in sensory cortex. In B. U. Forstmann & E.-J. Wagenmakers (Eds.), *An introduction to model-based cognitive neuroscience* (pp. 221–244). New York, NY: Springer. https://doi.org/10.1007/978-1-4939-2236-9_11

Kotabe, H. P., & Hofmann, W. (2015). On integrating the components of self-control. *Perspectives on Psychological Science, 10*, 618–638. https://doi.org/10.1177/1745691615593382

Kreitler, S. (Ed.). (2013). *Cognition and motivation: Forging an interdisciplinary perspective*. New York, NY: Cambridge University Press.

Krieglmeyer, R., De Houwer, J., & Deutsch, R. (2013). On the nature of automatically triggered approach–avoidance behavior. *Emotion Review, 5*, 280–284. https://doi.org/10.1177/1754073913477501

Kruglanski, A. W., Shah, J. Y., Fishbach, A., Friedman, R., Chun, W. Y., & Sleeth-Keppler, D. (2002). A theory of goal systems. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 34, pp. 331–378). San Diego, CA: Academic Press.

Kruglanski, A. W., & Webster, D. M. (1996). Motivated closing of the mind: "seizing" and " freezing". *Psychological Review, 103*, 263–283.

Lacey, S., & Lawson, R. (Eds.). (2013). *Multisensory imagery*. New York, NY: Springer.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences, 40*, E253. https://doi.org/10.1017/S0140525X16001837

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23*, 571–579. https://doi.org/10.1016/S0166-2236(00)01657-X

Landhäußer, A., & Keller, J. (2012). Flow and its affective, cognitive, and performance-related consequences. In S. Engeser (Ed.), *Advances in flow research* (pp. 65–85). New York, NY: Springer. https://doi.org/10.1007/978-1-4614-2359-1_4

Lang, P. J., & Bradley, M. M. (2010). Emotion and the motivational brain. *Biological Psychology, 84*, 437–450. https://doi.org/10.1016/j.biopsycho.2009.10.007

Lazarus, R. S. (1993). Coping theory and research: Past, present, and future. *Psychosomatic Medicine, 55*, 234–247.

Leary, M. R. (2007). Motivational and emotional aspects of the self. *Annual Review of Psychology, 58*, 317–344. https://doi.org/10.1146/annurev.psych.58.110405.085658

Lee, D., Seo, H., & Jung, M. W. (2012). Neural basis of reinforcement learning and decision making. *Annual Review of Neuroscience, 35*, 287–308.

Leotti, L. A., Iyengar, S. S., & Ochsner, K. N. (2010). Born to choose: The origins and value of the need for control. *Trends in Cognitive Sciences, 14*, 457–463. https://doi.org/10.1016/j.tics.2010.08.001

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin, 116*, 75–98.

Louro, M. J., Pieters, R., & Zeelenberg, M. (2007). Dynamics of multiple-goal pursuit. *Journal of Personality and Social Psychology, 93*, 174–193. https://doi.org/10.1037/0022-3514.93.2.174

Ma, W. J., & Jazayeri, M. (2014). Neural coding of uncertainty and probability. *Annual Review of Neuroscience, 37*, 205–220. https://doi.org/10.1146/annurev-neuro-071013-014017

Man, V., Nohlen, H. U., Melo, H., & Cunningham, W. A. (2017). Hierarchical brain systems support multiple representations of valence and mixed affect. *Emotion Review, 9*, 124–132.

Mann, M., Hosman, C. M. H., Schaalma, H. P., & de Vries, N. K. (2004). Self-esteem in a broad-spectrum approach for mental health promotion. *Health Education Research, 19*, 357–372. https://doi.org/10.1093/her/cyg041

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Cambridge, MA: The MIT Press.

Maxwell, J. C. (1868). On governors. *Proceedings of the Royal Society of London, 16*, 270–283.

McAdams, D. P. (2013). The psychological self as actor, agent, and author. *Perspectives on Psychological Science, 8*, 272–295. https://doi.org/10.1177/1745691612464657

McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. (1953). *The achievement motive*. East Norwalk, CT: Appleton-Century-Crofts. https://doi.org/10.1037/11144-000

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88*, 375–407. https://doi.org/10.1037/0033-295X.88.5.375

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. New York, NY: Henry Holt.

Mineka, S., & Hendersen, R. W. (1985). Controllability and predictability in acquired motivation. *Annual Review of Psychology, 36*, 495–529.

Moors, A. (2010). Automatic constructive appraisal as a candidate cause of emotion. *Emotion Review, 2*, 139–156. https://doi.org/10.1177/1754073909351755

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*, 310–318. https://doi.org/10.1177/1754073916669595

Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and future development. *Emotion Review, 5*, 119–124. https://doi.org/10.1177/1754073912468165

Moskowitz, G. B., & Grant, H. (Eds.). (2009). *The psychology of goals*. New York, NY: Guilford Press.

Moulton, S. T., & Kosslyn, S. M. (2009). Imagining predictions: Mental imagery as mental emulation. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*, 1273–1280. https://doi.org/10.1098/rstb.2008.0314

Mullally, S. L., & Maguire, E. A. (2014). Memory, imagination, and predicting the future: A common brain mechanism? *The Neuroscientist, 20*, 220–234. https://doi.org/10.1177/1073858413495091

Mulligan, K., & Scherer, K. R. (2012). Toward a working definition of emotion. *Emotion Review, 4*, 345–357. https://doi.org/10.1177/1754073912445818

O'Doherty, J. P. (2014). The problem with value. *Neuroscience & Biobehavioral Reviews, 43*, 259–268. https://doi.org/10.1016/j.neubiorev.2014.03.027

O'Leary, D., Uusberg, A., & Gross, J. J. (2017). Identity and self-control: Linking identity-value and process models of self-control. *Psychological Inquiry, 28*, 132–138. https://doi.org/10.1080/1047840X.2017.1337404

O'Shea, H., & Moran, A. (2017). Does motor simulation theory explain the cognitive mechanisms underlying motor imagery? A critical review. *Frontiers in Human Neuroscience, 11*, 72. https://doi.org/10.3389/fnhum.2017.00072

Ochsner, K. N., & Gross, J. J. (2014). The neural bases of emotion and emotion regulation: A valuation perspective. In J. J. Gross (Ed.), *Handbook of emotion regulation* (2nd ed., pp. 23–42). New York, NY: Guilford Press.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*, 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pellegrini, A. (2009). *The role of play in human development*. New York, NY: Oxford University Press.

Pessoa, L. (2015). Multiple influences of reward on perception and attention. *Visual Cognition, 23*, 272–290. https://doi.org/10.1080/13506285.2014.974729

Pessoa, L. (2018). Emotion and the interactive brain: Insights from comparative neuroanatomy and complex systems. *Emotion Review, 10*, 204–216. https://doi.org/10.1177/1754073918765675

Petty, R. E., Briñol, P., & DeMarree, K. G. (2007). The meta–cognitive model (MCM) of attitudes: Implications for attitude measurement, change, and strength. *Social Cognition, 25*, 657–686. https://doi.org/10.1521/soco.2007.25.5.657

Pezzulo, G., & Cisek, P. (2016). Navigating the affordance landscape: Feedback control as a process model of behavior and cognition. *Trends in Cognitive Sciences, 20*, 414–424. https://doi.org/10.1016/j.tics.2016.03.013

Pezzulo, G., Rigoli, F., & Friston, K. J. (2018). Hierarchical active inference: A theory of motivated control. *Trends in Cognitive Sciences, 22*, 294–306. https://doi.org/10.1016/j.tics.2018.01.009

Phaf, H., Mohr, S. E., Rotteveel, M., & Wicherts, J. M. (2014). Approach, avoidance, and affect: A meta-analysis of approach-avoidance tendencies in manual reaction time tasks. *Frontiers in Psychology, 5*, 378. https://doi.org/10.3389/fpsyg.2014.00378

Pinker, S. (1999). How the mind works. *Annals of the New York Academy of Sciences, 882*, 119–127. https://doi.org/10.1111/j.1749-6632.1999.tb08538.x

Pouget, A., Dayan, P., & Zemel, R. (2000). Information processing with population codes. *Nature Reviews Neuroscience, 1*, 125–132. https://doi.org/10.1038/35039062

Pouget, A., Drugowitsch, J., & Kepecs, A. (2016). Confidence and certainty: Distinct probabilistic quantities for different goals. *Nature Neuroscience, 19*, 366–374. https://doi.org/10.1038/nn.4240

Powers, W. T. (1973). Feedback: Beyond behaviorism. *Science, 179*, 351–356. https://doi.org/10.1126/science.179.4071.351

Radvansky, G. A., & Zacks, J. M. (2011). Event perception. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*, 608–620. https://doi.org/10.1002/wcs.133

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience, 9*, 545–556. https://doi.org/10.1038/nrn2357

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences, 20*, 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Ridderinkhof, K. R. (2014). Neurocognitive mechanisms of perception–action coordination: A review and theoretical integration. *Neuroscience & Biobehavioral Reviews, 46*, 3–29. https://doi.org/10.1016/j.neubiorev.2014.05.008

Ryan, R. M. (2012). *The Oxford handbook of human motivation*. New York, NY: Oxford University Press.

Scherer, K., Dan, E., & Flykt, A. (2006). What determines a feeling's position in affective space? A case for appraisal. *Cognition and Emotion, 20*, 92–113. https://doi.org/10.1080/02699930500305016

Scott, S. H. (2004). Optimal feedback control and the neural basis of volitional motor control. *Nature Reviews Neuroscience, 5*, 532–546. https://doi.org/10.1038/nrn1427

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences, 17*, 565–573. https://doi.org/10.1016/j.tics.2013.09.007

Seth, A. K. (2015). The cybernetic Bayesian brain: From interoceptive inference to sensorimotor contingencies. In T. Metzinger & J. M. Windt (Eds.), *Open MIND*. Frankfurt, Germany: MIND Group. Retrieved from http://www.open-mind.net/DOI?isbn=9783958570108

Shadmehr, R., Smith, M. A., & Krakauer, J. W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annual Review of Neuroscience, 33*, 89–108. https://doi.org/10.1146/annurev-neuro-060909-153135

Shah, J. Y., Friedman, R., & Kruglanski, A. W. (2002). Forgetting all else: On the antecedents and consequences of goal shielding. *Journal of Personality and Social Psychology, 83*, 1261–1280. https://doi.org/10.1037/0022-3514.83.6.1261

Shah, J. Y., & Gardner, W. L. (2008). *Handbook of motivation science*. New York, NY: Guilford Press.

Silvia, P. J. (2008). Interest - The curious emotion. *Current Directions in Psychological Science, 17*, 57–60. https://doi.org/10.1111/j.1467-8721.2008.00548.x

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society, 106*, 467–482.

Simpson, E. H., & Balsam, P. D. (Eds.). (2016). *Behavioral neuroscience of motivation*. Cham, Switzerland: Springer.

Skinner, E. A. (1996). A guide to constructs of control. *Journal of Personality and Social Psychology, 71*, 549–570.

Smith, C. A., & Lazarus, R. S. (1993). Appraisal components, core relational themes, and the emotions. *Cognition & Emotion, 7*, 233–269.

Stagner, R. (1977). Homeostasis, discrepancy, dissonance. *Motivation and Emotion, 1*, 103–138. https://doi.org/10.1007/BF00998515

Steel, P., & König, C. J. (2006). Integrating theories of motivation. *Academy of Management Review, 31*, 889–913.

Sterling, P. (2012). Allostasis: A model of predictive regulation. *Physiology & Behavior, 106*, 5–15. https://doi.org/10.1016/j.physbeh.2011.06.004

Stevens, L. E., & Fiske, S. T. (1995). Motivation and cognition in social life: A social survival perspective. *Social Cognition, 13*, 189–214. https://doi.org/10.1521/soco.1995.13.3.189

Swann, W. B. (1982). The self. *Science, 218*, 782–782. https://doi.org/10.1126/science.218.4574.782

Tamir, M. (2015). Why do people regulate their emotions? A taxonomy of motives in emotion regulation. *Personality and Social Psychology Review, 20*, 199–222. https://doi.org/10.1177/1088868315586325

Todorov, E. (2004). Optimality principles in sensorimotor control. *Nature Neuroscience, 7*, 907–915. https://doi.org/10.1038/nn1309

Tolman, E. C. (1925). Purpose and cognition: The determiners of animal learning. *Psychological Review, 32*, 285–297.

Toomela, A. (2016). The ways of scientific anticipation: From guesses to probabilities and from there to certainty. In M. Nadin (Ed.), *Anticipation across disciplines* (Vol. 29). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-22599-9

Tracy, J. L., & Robins, R. W. (2004). Putting the self into self-conscious emotions: A theoretical model. *Psychological Inquiry, 15*, 103–125. https://doi.org/10.1207/s15327965pli1502_01

Uusberg, A., Taxer, J. L., Yih, J., Uusberg, H., & Gross, J. J. (2019). Reappraising reappraisal. *Emotion Review*. https://doi.org/10.1177/1754073919862617.

Van Tongeren, D. R., DeWall, C. N., Green, J. D., Cairo, A. H., Davis, D. E., & Hook, J. N. (2018). Self-regulation facilitates meaning in life. *Review of General Psychology, 22*, 95–106. https://doi.org/10.1037/gpr0000121

Vickery, T. J., Chun, M. M., & Lee, D. (2011). Ubiquity and specificity of reinforcement signals throughout the human brain. *Neuron, 72*, 166–177. https://doi.org/10.1016/j.neuron.2011.08.011

von Uexküll, J. (1926). *Theoretical biology*. (D. L. Mackinnon, Trans.). London, England and New York, NY: K. Paul, Trench, Trubner & Co. Ltd. and Harcourt, Brace & Company.

Vuilleumier, P. (2015). Affective and motivational control of vision. *Current Opinion in Neurology, 28*, 29–35. https://doi.org/10.1097/WCO.0000000000000159

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92*, 548–573. https://doi.org/10.1037/0033-295X.92.4.548

Wessel, J. R., Danielmeier, C., Morton, J. B., & Ullsperger, M. (2012). Surprise and error: Common neuronal architecture for the processing of errors and novelty. *Journal of Neuroscience, 32*, 7528–7537. https://doi.org/10.1523/JNEUROSCI.6352-11.2012

Westgate, E. C., & Wilson, T. D. (2018). Boring thoughts and bored minds: The MAC model of boredom and cognitive engagement. *Psychological Review, 125*, 689–713. https://doi.org/10.1037/rev0000097

White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review, 66*, 297–333.

Wiener, N. (1948). *Cybernetics: Control and communication in the animal and the machine*. New York, NY: Wiley.

Yoo, S. B. M., & Hayden, B. Y. (2018). Economic choice as an untangling of options into actions. *Neuron, 99*, 434–447. https://doi.org/10.1016/j.neuron.2018.06.038

Chapter 7
Toward a Deep Science of Affect and Motivation

Brian Knutson and Tara Srirangarajan

Scientists of the mind have long sought to marry their models with mechanism. For instance, the innovators of neural network models of cognitive processing advised that a thorough understanding of something at one level of analysis also requires understanding at adjacent levels of analysis (Rumelhart, McClelland, & PDP Research Group, 1987). Linking levels of analysis represents the core of the "deep science" approach we advocate below. While such an approach is challenging and often represents a road less traveled in research, it may also offer unique advantages. For instance, linking levels of analysis may provide the most direct route from the scientific goals of observation and explanation to those of prediction and control (Watson, 1913).

This review enlists a deep science approach to reconnect affect and motivation by linking them to a neural level of analysis. The first section looks to the past to define components within levels of analysis and propose a framework for linking levels of analysis. The second and third sections describe current evidence linking neural activity to anticipatory affect and motivated behavior. The fourth section highlights future extensions to other levels of analysis and opportunities for exploration.

Past Foundations

Theories about links between affect and motivation are at least as old as the field of experimental psychology, yet their connection remains unclear (Berridge, 2004). Over time, research on affect and motivation has diverged into separate fields of inquiry, and their connections have been lost or forgotten. Reconnecting affect and

B. Knutson (✉) · T. Srirangarajan
Department of Psychology, Stanford University, Stanford, CA, USA
e-mail: knutson@stanford.edu

M. Neta, I. J. Haas (eds.), *Emotion in the Mind and Body*, Nebraska Symposium on Motivation 66, https://doi.org/10.1007/978-3-030-27473-3_7

motivation requires both definitions of these concepts as well as a framework for linking them.

Defining Affect Scientific definitions of affect can be traced to the first experimental psychologist, Wilhelm Wundt, who wrote: "In this manifold of feelings…it is nevertheless possible to distinguish certain different chief directions, including certain affective opposites of predominant character" (Wundt, 1897). Underlying the variety of emotional experiences, Wundt proposed dimensions running from positive to negative, aroused to subdued, and strained to relaxed. Remarkably, research over the following century repeatedly supported Wundt's early suspicions. For instance, studies of diverse emotional stimuli, including words used to describe emotional experience, emotional facial expressions, and responses to various sensory stimuli (e.g., sounds, smells, tastes) have consistently revealed that two independent dimensions can account for over half of their covariance. These independent dimensions have been called valence (running from positive to negative) and arousal (running from high to low) (Lang, Bradley, & Cuthbert, 1990; Osgood, Suci, & Tannenbaum, 1957; Russell, 1980).

Affective dimensions of valence and arousal have the potential to modulate sensory input as well as motor output. Subsequent theorists noted that a quarter turn (45° rotation) of the valence and arousal dimensions yielded continua which might descriptively be labeled "positive arousal" and "negative arousal" (Thayer, 1989; Watson & Tellegen, 1985). Functionally, the arousal component of these rotated dimensions should recruit attention and behavior, while the valence component might direct elicited attention or behavior toward or away from stimuli under consideration (Watson, Wiese, Vaidya, & Tellegen, 1999). The rotated dimensions therefore imply that positive arousal and negative arousal might not only sharpen sensory processing of opportunities or threats, but also could prepare relevant approach or avoidance behaviors, respectively. These dimensions might also evoke distinct affective experiences—with positive arousal eliciting feelings like energy, excitement, and confidence but negative arousal eliciting feelings like tension, anxiety, and irritability. Thus, affective dimensions describe covariance in subjective responses across a range of stimuli rather than to an isolated stimulus (e.g., words, faces, smells). Further, the fact that these affective dimensions can be assessed not only with verbal reports, but also with nonverbal expressive behavior (e.g., facial expression) and peripheral physiology (e.g., skin conductance, heart rate) (Lang, Greenwald, & Bradley, 1993) implies that conscious awareness or symbolic representation is not necessary for affect to modulate perception or behavior (Zajonc, 1980).

Beyond valence and arousal, Wundt proposed a third dimension running from tension to release, which was associated with the passage of time. In the context of motivation, tension versus release might represent affective changes that occur before versus after goal attainment (since behavioral approach and avoidance require both arousal and action). Consistent with Wundt's third dimension, we have proposed that "anticipatory affect" involves increases in positive arousal and/or

negative arousal, which then primes appetitive and/or aversive motivational states that facilitate movement toward or away from stimuli (Knutson & Greer, 2008).

The notion that affect occurs not only in response to significant outcomes, but also in anticipation of them, draws upon more recent theories which imply that arousal can influence both optimal (Bechara, Tranel, Damasio, & Damasio, 1996) and suboptimal risky choice (Loewenstein, Weber, Hsee, & Welch, 2001). These theories, however, typically invoke general arousal without also specifying valence, and so do not clarify when arousal should promote approach or avoidance behavior. The Anticipatory Affect Model sharpens these accounts by positing that positive arousal promotes approach, while negative arousal instead promotes avoidance (Knutson & Greer, 2008). Notably, anticipatory affect can be distinguished from "anticipated affect"—which refers to cognitive predictions about how one will feel in the future after an outcome has occurred, rather than how one feels immediately during anticipation of the outcome (Wilson & Gilbert, 2003). Anticipatory affect instead increases before uncertain goal outcomes occur. In this review, we focus on anticipatory affect, as defined by the independent dimensions of positive arousal and negative arousal, to cleanly link affect to motivation (see the third ring from the center of Fig. 7.1).

Defining Motivation Behaviorally, motivation (derived from the Latin "movere," meaning "to move") can simply be defined as an energization or amplification of ongoing activity. Psychological definitions for motivation, however, have ranged from broad to specific (Berridge, 2004). A broad definition might simply distinguish between different levels of motivation, which might correlate with changes in a state of general arousal. Narrower definitions typically refer to drives to fulfill specific

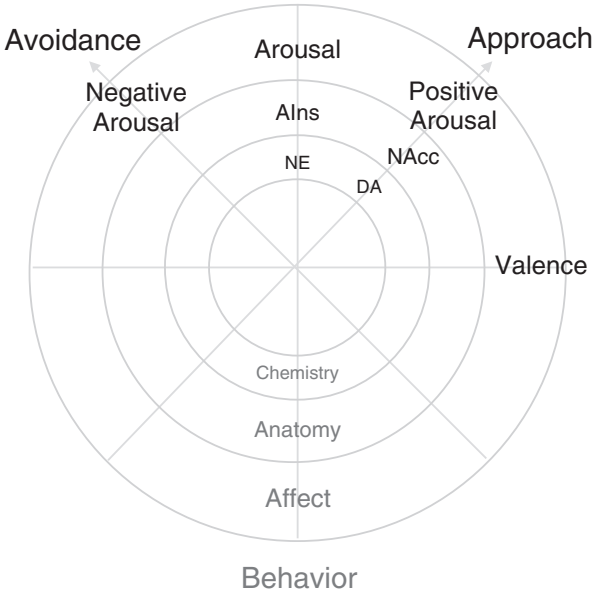

Fig. 7.1 Linking levels of analysis. Concentric circles represent levels of analysis extending from molecular (inner) to molar (outer). These levels depict neurochemistry (*DA* DopAmine, *NE* NorEpinephrine), neuroanatomy (*NAcc* Nucleus Accumbens, *AIns* Anterior Insula), anticipatory affect, and motivated behavior (adapted from Knutson, Katovich, & Suri, 2014)

unmet needs (i.e., which might compensate for a lack of specific necessities like food, water, oxygen, etc.). Between these broad and narrow definitions lies an intermediate definition describing motivations to approach potential opportunities or to avoid possible threats (Craig, 1918). These appetitive and aversive motivations further imply subsequent "consummatory" states capable of terminating motivated behavior after acquisition of an opportunity or avoidance of a threat.

Linking Levels of Analysis At the turn of the twenty-first century, growing computational power and availability of behavioral data (e.g., on the Internet) ushered in a new era of social science—transforming the earlier problem of too little data into a new challenge of too much data. In response, teams of researchers combined efforts to comprehensively map out different levels of analysis—including genetics, epigenetics, metabolics, neural connectivity, and other domains (sometimes applying the "-omics" suffix in the process). A primary goal of these projects typically involved comprehensively mapping all components ("nodes") and connections ("edges") within a given level of analysis (e.g., mapping out all the neurons and their connections in a worm; Bargmann, 2012). After a given level of analysis had been thoroughly characterized, researchers assumed that the acquired knowledge could inform research at other levels of analysis. Based on the goal of comprehensively characterizing all components and connections within a given level of analysis, these approaches might collectively be characterized as "broad science" (Knutson, 2016). In contrast to these "broad science" approaches, however, "deep science" approaches might instead seek to first identify critical components in adjacent levels of analysis and then to connect them across levels of analysis (e.g., demonstrating that optogenetic stimulation of midbrain dopamine neurons in rats can increase striatal Functional Magnetic Resonance Imaging (FMRI) activity and approach behavior (Ferenczi et al., 2016)).

Although broad and deep scientific approaches differ in their initial aims, they might serve complementary and synergistic functions. For example, the Research Domain Criteria (RDoC) framework endorsed by the National Institute of Mental Health (Insel et al., 2010) is both horizontally defined by different functional systems, and vertically defined by different levels of analysis (ranging from micro to macro; see Table 7.1). Broad science versus deep science approaches, however, invoke different potential costs and benefits. While broad science approaches require expertise and instrumentation at a single level of analysis, deep science approaches require expertise and instrumentation across two or more levels of analysis. Thus, while broad science approaches might accumulate findings faster within a given level of analysis, deep science approaches might more rapidly link components across levels of analysis.

The deep science goal of linking levels of analysis first requires identifying adjacent levels of analysis and relevant components within them to connect (Cacioppo & Berntson, 1992). A popular three-level scheme proposed by neuroscientist David Marr included: (1) a computational level, describing the goal of a computation; (2) an algorithmic level, describing relevant representations and rules for transforming

Table 7.1 Broad (rows) versus deep (columns) science approaches in the National Institute of Mental Health Research Domain Criteria (adapted from Insel et al., 2010)

	Functional Domains				
Levels of Analysis	*Positive Valence Systems*	*Negative Valence Systems*	*Cognitive Systems*	*Social Process Systems*	*Arousal / Regulatory Systems*
Genes					
Molecules					
Cells					
Circuits					
Behavior					
Self-reports					

them; and (3) an implementational level, describing the machinery supporting the algorithm (Marr, 1982). Though logically and causally connected, Marr noted that these three levels were only "loosely related," allowing some phenomena to be explored at only one level of analysis. He also suggested that many phenomena could be addressed by analyzing higher computational or algorithmic levels before the lower implementational level. Consequently, theorists often interpreted Marr's suggestions in a way that justified focusing exclusively on higher functional levels of analysis (but not lower physical levels), thus pursuing broad but not deep scientific aims.

Although originally applied to visual processing, Marr's scheme might also extend to affective processing—but only after some modifications. First, the three levels could be more transparently relabeled (from bottom to top) as "physiology," "process," and "purpose." This relabeling might reaffirm the implicit aim of using lower-level neurophysiology to constrain higher-level algorithms and computations. Second, the lower level (of physiology) might offer a more promising starting point than the middle (of process) or higher (of purpose) levels of analysis, as causal influences are likely to flow first and fastest up from physiology to process to purpose. Additionally, while the physiological level is necessarily constrained by the design of nature, the purpose level is only constrained by the bounds of human imagination. Third, the ultimate purpose of vision likely differs from that of affect. For instance, meeting the visual computational goal of object identification (originally specified by Marr) might require a series of algorithms capable of identifying features, textures, shapes, objects, and so forth, which are implemented by a "ventral visual" cortical processing stream (DiCarlo, Zoccolan, & Rust, 2012). By contrast, the affective purpose of approaching opportunities while avoiding threats might require processes that weigh potential gains against potential losses, and

which are physiologically modulated by ascending monoaminergic projections to critical subcortical targets (Knutson & Greer, 2008).

These overarching differences in purpose imply that linking brain to affect to motivation may ultimately require shifting from an "information processor" metaphor (e.g., in the case of processing visual objects) to a "hedonic sharpener" metaphor (e.g., in the case of processing affect; see Table 7.2). Specifically, the goal of affective circuits is not necessarily to accurately convey information, but rather, to efficiently assess potential gains and losses in order to facilitate rapid action capable of promoting or preserving inclusive fitness. This overarching goal of pursuing positive feelings versus informational accuracy might lead to divergent outcomes over time. But information processing and hedonic sharpening purposes need not necessarily conflict, and might also sequentially and synergistically align.

Once relevant concepts have been identified to connect across levels, evaluating potential links raises a further challenge of measuring relevant concepts at matching resolution. Starting from the physiological level of brain activity, two primary resolution criteria include space (e.g., the size of the brain circuit under consideration) and time (e.g., its speed of operation). For instance, linking monoaminergic activity to anticipatory affect requires consideration of the spatial constraint that neurons carrying these neurotransmitters project to small subcortical regions mere millimeters in diameter, as well as the temporal constraint that the firing of these neurons and subsequent release of neurotransmitters in projection targets varies on a second-to-second basis (Robinson, Venton, Heien, & Wightman, 2003). These constraints imply that neural measures should offer millimeter subcortical spatial resolution as well as second-to-second temporal resolution, while measures of affect should match a similar timescale. Methods that measure concepts with matching resolution could therefore best allow researchers to test new links across levels. Indeed, rapid advances since the turn of the twenty-first century in the discovery of neural

Table 7.2 Comparison of levels of analysis for processing visual objects versus anticipatory affect (modified from Marr, 1982)

Vision: "Information Processor"	**Affect**: "Hedonic Sharpener"
Computation: Classify objects	**Purpose**: Approach potential gains while avoiding losses
Algorithm: Identify features, shapes, categories	**Process**: Identify and weigh potential gains against losses
Implementation: Ventral visual cortical stream	**Physiology**: Midbrain monoaminergic projections to subcortical targets

mechanisms that drive behavior might have resulted from the rise of neuroimaging methods like Functional Magnetic Resonance Imaging (FMRI) and neural manipulation methods like optogenetics—which feature overlapping spatial (on the order of millimeters) and temporal (on the order of sub-seconds) resolution (Sejnowski, Churchland, & Movshon, 2014). A deep science approach could therefore not only inform the selection of concepts but also of matching methods capable of linking those concepts across levels of analysis.

Leveling Up from Physiology to Process: Linking FMRI Activity and Anticipatory Affect

Which brain circuits are recruited during the anticipation of good and bad outcomes? Based on the adapted levels of analysis approach described above, one might begin by linking physiology to process. But where in the haystack of the brain should researchers begin to search for the needles of activity that can connect neural activity to anticipatory affect? Over a century of affective neuroscience studies involving animal models could guide the search for relevant neural circuits, while technical developments offer newer methods with matching resolution for linking physiology to process in humans.

Midway through the twentieth century, comparative researchers discovered that electrical and chemical stimulation of specific brain circuits could unconditionally elicit approach or avoidance behavior (Panksepp, 1998). Dramatic examples included "self-stimulation," in which animals would work to increase or decrease electrical or chemical stimulation of their own brain, often to the exclusion of all other incentives—including food, drink, and sex (Olds, 1955; Olds & Milner, 1954). Subsequent research revealed that most circuits that support self-stimulation lie below the neocortex in deeper subcortical or allocortical regions. For instance, electrical stimulation of regions along the ascending trajectory of midbrain dopamine neurons (i.e., projecting from the Ventral Tegmental Area (VTA) to the Lateral Hypothalamus (LH), Ventral Striatum (VS; including the Nucleus Accumbens, NAcc), and Orbital and Medial Prefrontal Cortex (OFC and MPFC)) can unconditionally elicit approach behavior (Olds & Fobes, 1981). Electrical stimulation of other brain regions (i.e., descending from the Anterior Insula (AIns) and BasoLateral Amygdala (BLAmy) through the Stria Terminalis (ST) to the Medial Hypothalamus (MHyp) and PeriAqueductal Gray (PAG)) can instead unconditionally elicit avoidance behavior (Hess, 1958). Since electrical stimulation of these circuits unconditionally evokes approach or avoidance behavior, they might provide reasonable initial starting points for linking brain activity to anticipatory affect in humans (Knutson & Greer, 2008; Schultz, Dayan, & Montague, 1997).

Linking activity in these circuits to anticipatory affect in humans might next require noninvasive neuroimaging methods capable of resolving activity at millimeter deep spatial resolution and second-to-second temporal resolution. FMRI, developed

in the early 1990s, first offered this combination of spatial and temporal resolution (Bandettini, Wong, Hinks, Tikofsky, & Hyde, 1992; Kwong et al., 1992). Early FMRI studies attempted to localize neural activity associated with parametrically varying sensory stimuli (e.g., responses in primary visual cortex to checkerboards flickering at different frequencies) and motor responses (e.g., responses in primary motor cortex to finger tapping at varying tempos; Engel et al., 1994; Rao et al., 1995). Inspired by sensorimotor localization studies, researchers subsequently sought to localize neural activity related to more abstract psychological phenomena, including affect and valuation. While previous research using other neuroimaging methods had explored neural responses to positive and negative emotional stimuli (e.g., standardized sets of affective pictures), many could not control for confounds related to variation in sensory input, motor output, arousal, or expectancy due to limited temporal (e.g., Positron Emission Tomography or PET) or spatial (e.g., ElectroEncephaloGraphy or EEG) resolution.

The spatiotemporal resolution of FMRI allowed researchers to control for some of these confounds by precisely timing the presentation of positive and negative cues and outcomes, and by synchronizing task presentation to image acquisition. Further, although many comparative studies were conducted with primary rewards (e.g., juice) and punishments (e.g., shocks), primary incentives proved difficult to directly compare or scale. Thus, FMRI researchers began to use money as a flexible but controllable incentive that could be inverted, scaled, cued, and delivered to humans (Delgado et al., 2000; Elliott, Friston, & Dolan, 2000; Knutson, Westdorp, Kaiser, & Hommer, 2000; O'Doherty, Kringelbach, Rolls, Hornak, & Andrews, 2001). For instance, using a Monetary Incentive Delay (or "MID") task, researchers could distinguish neural responses during anticipation of uncertain monetary gains and losses from responses to actual monetary gain and loss outcomes (Knutson, Adams, Fong, & Hommer, 2001; Knutson, Fong, Adams, Varner, & Hommer, 2001; Knutson, Fong, Bennett, Adams, & Hommer, 2003). Beginning in the early 2000s, these FMRI studies using monetary incentives began to yield robust and replicable results. Specifically, while anticipation of increasing gains proportionally increased activity in the ventral striatal NAcc, dorsal striatal medial caudate, and AIns, anticipation of increasing losses proportionally increased activity only in the medial caudate and AIns (Knutson et al., 2003). Gain outcomes, on the other hand, increased activity in the MPFC and ventral striatal putamen (Delgado et al., 2000), whereas loss outcomes tended to increase activity in the AIns (Knutson et al., 2003).

Initial localization of neural responses during incentive anticipation with event-related FMRI raised further questions about the scope and limits of these findings, which were subsequently addressed by research. First, NAcc activity during anticipation of secondary (or learned) monetary gains and AIns activity during anticipation of monetary losses also generalized to anticipation of primary (or unlearned) gustatory gains and losses (e.g., tasting sweet juice vs. salty tea; O'Doherty, Deichmann, Critchley, & Dolan, 2002), suggesting that anticipatory activity does not depend on the sensory modality of outcomes. Second, NAcc activity during anticipation of gains and AIns activity during anticipation of losses did not depend

on a subsequent motor response requirement (Ramnani, Elliott, Athwal, & Passingham, 2004). This activity could be augmented by anticipating a motor response, however, particularly in dorsal striatal regions including the medial caudate (Tricomi, Delgado, & Fiez, 2004). Third, NAcc activity during anticipation of gains and AIns activity during anticipation of losses could be elicited by subliminally presented cues, suggesting that it does not require conscious awareness (Pessiglione et al., 2008). Fourth, NAcc activity during anticipation of gains could augment other types of subsequent behavior, including memory (Adcock, Thangavel, Whitfield-Gabrieli, Knutson, & Gabrieli, 2006) and effort (Pessiglione et al., 2007), implying that anticipatory activity has the capacity to modulate a broad range of outputs. Fifth, adding other attributes to cues during anticipation of gains and losses (e.g., probability, delay) tended to increase MPFC activity as well, consistent with the notion that the MPFC plays a role in value integration (Knutson, Taylor, Kaufman, Peterson, & Glover, 2005). Together, these findings suggest that neural activity during anticipation of gains and losses is robust, can be elicited by a flexible spectrum of cues, and can potentiate a broad range of responses.

Two decades and hundreds of studies later, these patterns of anticipatory activity have been largely confirmed by several meta-analytic reviews of FMRI studies of incentive processing (Bartra, McGuire, & Kable, 2013; Clithero & Rangel, 2013; Diekhof, Kaps, Falkai, & Gruber, 2012; Knutson & Greer, 2008; Liu, Hairston, Schrier, & Fan, 2011; Sescousse, Caldú, Segura, & Dreher, 2013). Moreover, when self-reported affective responses to incentive cues are probed, the anticipation of monetary gain proportionally increases positive arousal, whereas the anticipation of monetary loss proportionally increases negative arousal (Cooper & Knutson, 2008). Finally, individual differences in NAcc responses to large gain cues correlate with cue-elicited positive (but not negative) arousal, whereas individual differences in medial caudate and AIns responses to large loss cues correlate with cue-elicited negative arousal as well as positive arousal (Samanez-Larkin et al., 2007). Together, these findings suggest that anticipation of gain elicits proportional activity in the NAcc and correlated positive arousal, whereas anticipation of loss elicits proportional activity in the AIns and medial caudate and correlated general arousal—linking brain activity to anticipatory affect (see also: Kruschwitz et al., 2018; Kühn & Gallinat, 2012).

Unexpectedly, this pattern of findings appeared more robustly for anticipated gain than for anticipated loss. Whereas gain anticipation clearly increases NAcc, medial caudate, and AIns activity, loss anticipation also seems to increase medial caudate and AIns activity. So, while NAcc activity aligns well with positive arousal, AIns and medial caudate activity appear to more closely align with general arousal. Despite this apparent absence of a full dissociation, given the relative difference in regions' alignment with valence, researchers should still be able to use activity in the NAcc to infer positive arousal, and relative activity in the AIns versus the NAcc to infer negative arousal (Knutson et al., 2014; Fig. 7.1). Together, these findings could help to resolve a debate about whether NAcc activity correlates with the experience of affective valence or salience (Berridge & Robinson, 1998; Zink, Pagnoni,

Martin-Skurski, Chappelow, & Berns, 2004) by suggesting that it is associated with both positivity and arousal—and that the experience of anticipatory affect is likely to be fleeting (Cooper & Knutson, 2008; Litt, Plassmann, Shiv, & Rangel, 2011).

Leveling Up from Process to Purpose: Linking Anticipatory Affect and Incentive Motivation

After establishing links from brain activity to anticipatory affect, could additional links extend to motivated behavior? By 2005, researchers began to realize that FMRI methods could not only clarify how sensory input influences brain activity, but could also elucidate whether some of that brain activity predicts motor output. Research accordingly shifted from the scientific goal of explanation to that of prediction. Specifically, researchers began to examine whether activity in circuits associated with anticipatory affect could predict upcoming motivated behavior. According to an Anticipatory Affect Model inspired by localization findings, if risky propositions are framed as choices that require balancing uncertain gains against uncertain losses, NAcc activity should promote approach and risk-seeking, whereas AIns activity should instead promote avoidance and risk-aversion (Knutson & Greer, 2008; see Fig. 7.2). Subsequent studies investigating whether anticipatory affective activity could predict behavior involved diverse scenarios such as gambling, purchasing, and social interaction.

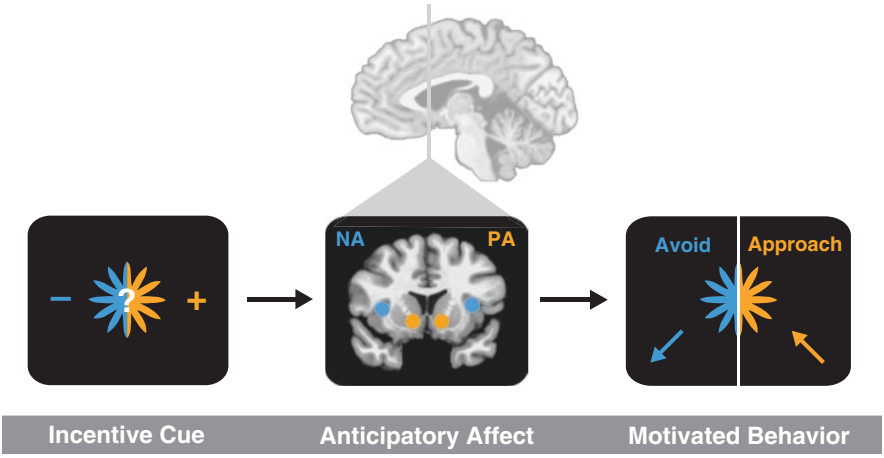

Fig. 7.2 Anticipatory affect model. An incentive cue for an uncertain future outcome initially elicits activity in at least two brain regions (NAcc = orange and AIns = blue), which may correlate with positive arousal and negative arousal, respectively. The balance of activity in these regions then promotes either approach toward or avoidance of the cued outcome (adapted from Knutson & Greer, 2008)

Early prediction studies focused on financial risk-taking. In an initial study of risk-taking in the context of financial investing, increased NAcc activity predicted both optimal and suboptimal risk-seeking choices, whereas increased AIns activity predicted both optimal and suboptimal risk-averse choices (Kuhnen & Knutson, 2005). Other research indicated that activity in these circuits could predict acceptance versus rejection of risky gambles, respectively (Canessa et al., 2013; Hampton & O'Doherty, 2007; Knutson, Wimmer, Kuhnen, & Winkielman, 2008). Some evidence linked these predictions to affect rather than numerical calculation, since both positive arousal and NAcc activity could account for commonly observed but apparently inconsistent preferences for positively skewed (or lottery-like) gambles, unlike traditional finance theory (e.g., mean-variance accounts; Leong, Pestilli, Wu, Samanez-Larkin, & Knutson, 2016; Wu, Bossaerts, & Knutson, 2011). Further, incidental affective stimuli may alter risky choice by changing activity in these circuits. On the one hand, presenting incidental but attractive pictures before gambles evoked positive arousal and increased risk-taking, an effect partially mediated by increased NAcc activity (Knutson et al., 2008). On the other hand, the threat of shock reduced risk-taking in the case of gambles, partially as a function of increasing AIns activity (Engelmann, Meyer, Fehr, & Ruff, 2015). Further, resting NAcc activity prior to gamble presentation could predict subsequent risk-taking (Huang, Soon, Mullette-Gillman, & Hsieh, 2014). Thus, these findings not only confirm that NAcc and AIns activity increase during risk anticipation (Preuschoff, Quartz, & Bossaerts, 2008), but further demonstrate that activity in these circuits differentially predicts choices to approach or avoid those risks (Wu, Sacchet, & Knutson, 2012), consistent with financial risk analyses that model mean and variance as distinct but oppositely weighted terms (Knutson & Huettel, 2015).

Other prediction studies explored people's choices to purchase consumer products. Early research suggested that increased NAcc activity in response to products and increased MPFC but decreased AIns activity in response to associated prices could predict choices to purchase seconds later (Karmarkar, Shiv, & Knutson, 2015; Knutson et al., 2008; Knutson, Rick, Wimmer, Prelec, & Loewenstein, 2007). Subsequent research indicated that brain activity could predict even more distant choices, since mere exposure to products without a choice prompt similarly elicited NAcc and MPFC responses that predicted later choices made outside the scanner (Levy, Lazzaro, Rutledge, & Glimcher, 2011; Smith, Douglas Bernheim, Camerer, & Rangel, 2014). Further, full attention was not necessary, since NAcc, MPFC, and AIns responses to products presented in the context of focused versus distracting tasks equally predicted later choices (Tusche, Bode, & Haynes, 2010). Together, these findings linked anticipatory affect to motivated choice, and further suggested an ongoing implicit influence (Zajonc, 1980). Other studies broadened the range of stimuli under consideration, demonstrating that increased NAcc and MPFC (and sometimes decreased AIns) activity in response to faces, places, pictures, and music could predict subjects' later preferences for those stimuli over other options or money (Lebreton, Jorge, Michel, Thirion, & Pessiglione, 2009; Salimpoor et al., 2013; Smith et al., 2010). Results from another study even suggested that students' NAcc responses to pictures of food and erotica could predict

those individuals' weight gain and sexual activity, respectively, several months later (Demos, Heatherton, & Kelley, 2012). Accordingly, reviews of this expanding literature have concluded that NAcc, MPFC, and AIns (negative) responses to varied stimuli can predict later choice behavior (Knutson & Karmarkar, 2014; Levy & Glimcher, 2012).

A third body of research investigated social interaction—often in the context of quantifiable and controllable exchange tasks adapted from Game Theory (Sanfey, 2007). With respect to cooperative behavior, increased NAcc activity predicted increased cooperation with strangers in a Prisoner's Dilemma Game (Rilling et al., 2002), as well as increased reciprocation in a Trust Game (King-Casas et al., 2005). Increased NAcc activity and self-reported positive arousal also predicted choices to give resources to strangers and charities in tasks similar to a Dictator Game (Genevsky, Västfjäll, Slovic, & Knutson, 2013; Harbaugh, Mayr, & Burghart, 2007; Krueger et al., 2007; Park, Blevins, Knutson, & Tsai, 2017). With respect to competitive behavior, however, increased AIns activity in response to unreciprocated cooperation predicted subsequent defection in the Prisoner's Dilemma Game (Rilling, Sanfey, Aronson, Nystrom, & Cohen, 2004). Increased AIns activity also predicted rejection of unfair offers, even at personal cost, in the Ultimatum Game (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). Although self-reported affect was not assessed in many of these dynamic interaction studies, several lines of evidence implicated anticipatory affect promoting acceptance or rejection of social offers. For instance, the presence of MPFC lesions is associated with increased rejection of unfair offers in the Ultimatum Game (Koenigs & Tranel, 2007). Further, induction of negative affect also increased rejection of unfair offers in the Ultimatum Game, and this effect was mediated by increased AIns activity (Harlé & Sanfey, 2007). Thus, as summarized in reviews, NAcc activity and positive arousal can foster cooperation, whereas AIns activity and negative arousal may instead promote competition in the context of social interaction (Knutson & Wimmer, 2007; Ruff & Fehr, 2014; Sanfey, 2007).

These collected findings are consistent with the prediction that neural activity associated with anticipatory affect can predict risky choice (Knutson & Greer, 2008). Specifically, when confronting diverse scenarios (e.g., financial risk, consumer products, and social interactions), NAcc activity predicts choices to approach, whereas AIns activity predicts choices to avoid. While activity in these circuits typically changes on a second-to-second basis, presenting incidental but affect-inducing stimuli immediately before choice can perturb ongoing activity in these circuits, which then appears to alter the upcoming choice. Further, activity in these circuits predicts both consistent and inconsistent choices, implying that anticipatory affect contributes to rational as well as irrational choices. Thus, these findings link both brain activity and anticipatory affect to motivated behavior.

Anticipatory affect can be further situated within a comparative anatomical framework that describes frontal and subcortical circuits as connecting in an "ascending spiral" pattern (Haber & Knutson, 2010). This Affect-Integration-Motivation (AIM) framework (Samanez-Larkin & Knutson, 2015) specifies anatomical, chemical, and functional physiology capable of supporting the processing

of: (1) anticipatory affect (midbrain dopamine connections to NAcc, midbrain norepinephrine connections to AIns, and glutamatergic connections from AIns to NAcc); (2) value integration (connections of NAcc and AIns indirectly to the MPFC and then back again to the ventral striatum); and (3) incentivized motivation (partially overlapping ascending loops through the dorsal striatum and medial wall of the frontal cortex to the motor cortex). The AIM framework thus presents a componential, sequential, and hierarchical scheme for predicting and testing links from brain activity to anticipatory affect to motivated behavior (Fig. 7.3).

Future Directions

Summary

Remarkable advances since the turn of the twenty-first century have illuminated how brain activity can support anticipatory affect and motivated behavior in humans. These advances likely arose not only from conceptual advances in acknowledging the influence of anticipatory affect in motivating subsequent behavior (Bechara et al., 1996; Finucane, Alhakami, Slovic, & Johnson, 2000; Knutson & Greer, 2008; Loewenstein et al., 2001), but even more from the technical innovation of methods for measuring brain activity immediately prior to behavioral responses.

Rapidly accumulating evidence has begun to link previously disparate levels of analysis (see Fig. 7.1). Initial findings linked brain activity to anticipatory affect, as NAcc activity increases during anticipation of diverse gains (including but not limited to monetary outcomes) and correlates with self-reported positive arousal, but AIns activity increases during anticipation of both losses and gains and correlates with self-reported general or negative arousal. Subsequent findings linked anticipatory affect to motivated behavior, as NAcc activity and positive arousal predict motivated approach toward diverse stimuli (e.g., financial risks, consumer products, social interaction), but AIns activity and negative arousal predict motivated avoidance of those same stimuli.

Together, these links across levels of analysis lay the groundwork for specifying testable causal predictions. On the one hand, dopamine release (and the resulting rate of postsynaptic agonism of D1 receptors) should increase NAcc FMRI activity, positive arousal, and subsequent behavioral approach toward stimuli under consideration (Ferenczi et al., 2016; Knutson & Gibbs, 2007). On the other hand (and more speculatively), norepinephrine release (and the resulting rate of postsynaptic agonism of AD1B receptors) should increase AIns FMRI activity, general or negative arousal, and subsequent behavioral avoidance of stimuli under consideration. The balance of activity in these circuits should predict choices to approach or avoid risky propositions, which feature uncertain gains as well as losses (Knutson et al., 2014; Knutson & Greer, 2008). If both circuits are similarly activated, other neural mechanisms (e.g., descending from the MPFC) may be necessary to resolve differences and thereby facilitate choice (Samanez-Larkin & Knutson, 2015).

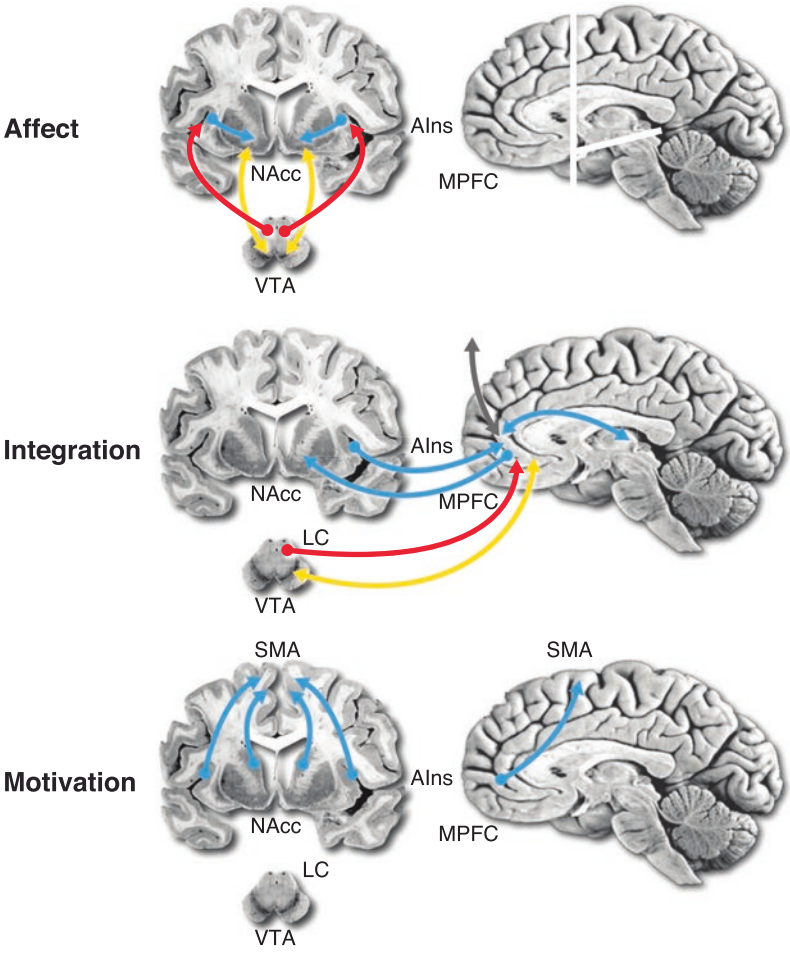

Fig. 7.3 The Affect-Integration-Motivation (AIM) framework. According to the AIM framework, three hierarchical and sequential processes can precede and promote choice. Brain regions involved in these processes are: (top) *Affect* processes associated with: Ventral Tegmental Area (VTA) DopAmine (DA; yellow) neurons projecting to the Nucleus Accumbens (NAcc); Locus Coeruleus (LC) NoradrenalinE (NE; red) neurons projecting to the Anterior Insula (AIns); and AIns glutamatergic (blue) neurons projecting to the NAcc, which potentiate anticipation of gain and loss (white lines on the right indicate the plane of sections depicted on the left); (middle) *Integration* processes associated with: VTA dopamine neurons and LC noradrenaline neurons which also project to the Medial PreFrontal Cortex (MPFC). Additionally, the NAcc indirectly projects to the MPFC via GABAergic connections to the pallidum (not depicted) and glutamatergic projections from the thalamus. The AIns also projects to the MPFC, presumably via glutamatergic connections. Finally, MPFC glutamatergic neurons project directly back to the NAcc (and adjacent Ventral Striatum), facilitating integration of value and other relevant input (for instance, arriving from the medial temporal and lateral frontal cortical regions); (bottom) *Motivation* processes are associated with dorsal striatal and insular glutamatergic neurons that project to the Supplementary Motor Area (SMA), potentiating motor action (adapted from Samanez-Larkin & Knutson, 2015)

Implications

A deep science approach need not restrict itself to only three levels of analysis—once links have been established from brain activity to anticipatory affect to motivated behavior, this approach could extend to include additional lower (e.g., neurochemistry) and higher (e.g., group behavior) levels of analysis (see Fig. 7.1).

Leveling Down Links might extend up from an even lower level to connect changes in neurochemistry to FMRI activity in predicted circuits. New comparative methods make causal tests of these links possible. For instance, optogenetic tools now allow researchers to transfect specific neurons with viruses that induce their genetic machinery to express light-sensitive ion channels. These transfected neurons can then be precisely controlled with light via implanted fiber optic probes (Witten et al., 2011). Based on the proposed levels of analysis scheme (see Fig. 7.1), dopamine firing should increase FMRI activity in the ventral striatum, including the NAcc (Knutson & Gibbs, 2007). In fact, research has indicated that in awake rats, phasic optogenetic stimulation of midbrain dopamine neuron firing at a frequency similar to that elicited by reward cues (i.e., 2 s of 20 Hz stimulation) robustly increased FMRI activity in both the ventral and dorsal striatum. Moreover, the magnitude of increased FMRI activity in the ventral striatum (including the NAcc) predicted how intensely rats would work to self-administer that same stimulation (Ferenczi et al., 2016; Fig. 7.4). This robust causal link from optogenetic stimulation of midbrain dopamine neurons to increased striatal FMRI activity has been independently replicated in other laboratories (Decot et al., 2017; Lohani, Poplawsky, Kim, & Moghaddam, 2017). By using tools with matching resolutions, researchers could causally demonstrate that optogenetically stimulating the firing of midbrain dopamine neurons increases NAcc FMRI activity, which further predicts approach behavior. Additional evidence for this link showed that: (1) optogenetically inhibiting midbrain dopamine neuron firing slightly decreased striatal FMRI activity; (2) blocking postsynaptic dopamine receptors blunted this effect; and (3) optogenetically enhancing MPFC input to the striatum also blunted this effect. Together, these findings establish causal links from an even lower level by demonstrating that selective optogenetic stimulation of midbrain dopamine firing can increase NAcc FMRI activity and associated approach behavior. Future research might explore the effects of norepinephrine firing in the AIns in a similar manner.

Leveling Up Links could further extend to an even higher level to connect individual behavior to aggregate behavior. Data from the motivated behavior level might be used to forecast aggregate choice. In the case of "neuroforecasting," researchers have used brain activity in smaller scanned groups to forecast the choices of other larger groups of people outside the laboratory (e.g., in markets on the internet; Knutson & Genevsky, 2018). Growing evidence suggests that sampled FMRI activity can forecast market demand for a diverse array of online products. Specifically, sampled NAcc activity has been used to forecast music sales (Berns & Moore, 2012), the impact of advertisements (Venkatraman et al., 2015), purchases of food

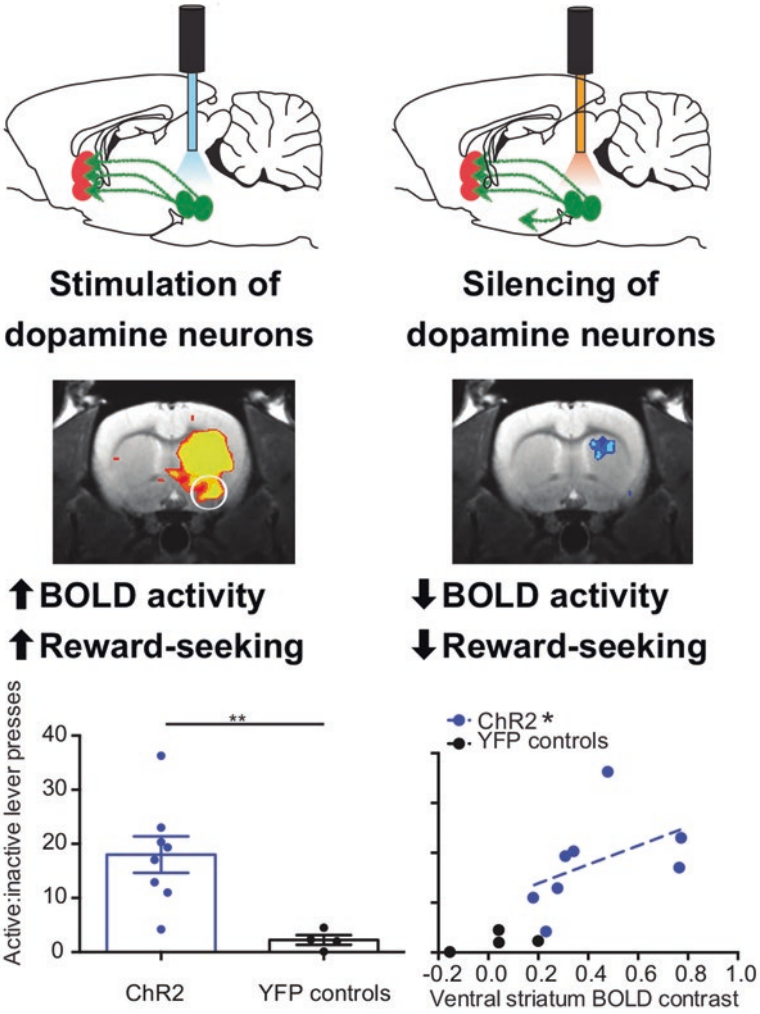

Fig. 7.4 Linking midbrain dopamine neuron firing to NAcc FMRI activity. Optogenetic stimulation of midbrain dopamine neurons increases striatal FMRI activity (top left; white circle indicates ventral striatum), whereas optogenetic silencing of these neurons mildly diminishes striatal FMRI activity (top right). Only transfected rats work to self-administer optogenetic midbrain dopamine stimulation (bottom left); and rats with increased ventral striatal activity from optogenetic midbrain dopamine neuron stimulation also work more intensely to self-administer that stimulation (bottom right) (adapted from Ferenczi et al., 2016)

(Kühn, Strelow, & Gallinat, 2016), the spread of news stories on social media platforms (Scholz et al., 2017), and the success of microlending appeals (Genevsky & Knutson, 2015) as well as crowdfunding appeals (Genevsky, Yoon, & Knutson, 2017). Researchers have additionally used group MPFC activity to forecast aggregate responses to smoking cessation appeals (Falk, Berkman, & Lieberman, 2012)

and news articles (Scholz et al., 2017). Remarkably, in some cases, neural activity can forecast market behavior even when individual self-report and behavior cannot—potentially supporting a "partial scaling" account in which neural activity in circuits associated with anticipatory affect affords better forecasts than activity in other circuits or even individual choice itself (Knutson & Genevsky, 2018). Together, these findings suggest that sampled neural activity can forecast aggregate choice. Further, in some cases, neural measures might augment or even outperform more traditional behavioral measures.

Leaping Levels The linking levels account implies movement from one level up to the next adjacent level in the same direction. In many cases, however, links bridging more than one level have been established. For instance, much of the research reviewed above links neural activity directly to motivated behavior without assessing intermediate anticipatory affect. While not inconsistent with the spatial logic of predictions implied by the linking levels account (Fig 7.1), these findings raise the possibility that intermediate measures could be refined either conceptually or technically (e.g., substituting momentary implicit measures of affective experience for retrospective explicit measures) to better match adjacent levels. In a more extreme example from neuroforecasting, sampled brain activity forecasts aggregate choice, even when sampled self-reported affect and choice do not. These findings may imply that some lower-level components can reveal "hidden information" about higher-level components (Ariely & Berns, 2010), and possibly, that concepts at intermediate levels need further refinement (e.g., mixed incentives may induce ambivalent affective responses). Thus, linking components across levels of analysis may provide clues for future conceptual and technical refinement of relevant measures.

Recursive Influence Unlike functional accounts that start from higher levels of analysis, the current approach builds from lower levels of analysis. Regardless of initial priorities, however, causality likely flows down as well as up the levels of analysis—but not in the same manner. Specifically, downward links might involve distinct processes which operate at longer timescales. For instance, approach behavior only requires neural firing to change on a second-to-second timescale (i.e., dopamine agonism of the postsynaptic receptor opens ion channels which change the membrane potential of the postsynaptic neuron, causing it to fire). Reward learning, however, requires genetic transcription to modify neural membranes and alter receptor expression, which necessarily unfolds over a longer timescale on the order of hours (Hyman, Malenka, & Nestler, 2006). Thus, reward learning might reciprocally influence reward anticipation, but only at this longer timescale after upward and downward causal influences have cycled through the system. By implication, then, tracking recursive causation from higher to lower levels might require distinct methods featuring different spatial and temporal resolutions. Studying reciprocal links across levels of analysis (both upwards and downwards) might ultimately enhance scientific understanding of how components at different levels interact over time, both with respect to negative feedback mechanisms typical of homeostatic

regulation (e.g., the cycle of food appetite, consumption, and satiety), as well as positive feedback loops that sometimes arise in the context of pathological dysregulation (e.g., escalating addiction to stimulants).

Limits

The deep science approach prioritizes depth over breadth, and so has associated costs as well as benefits. Critically, researchers need to first identify and extend from sparse nodes that can support robust, reliable, and ideally causal links across levels. This might come at the cost of conceptual richness associated with characterizing all the connections within a single level of analysis. The initial sparsity of the deep science approach, however, hopefully leaves gaps open for more extensive exploration later.

Emotion Emotion is notably absent from the levels of analysis framework presented so far. While Wundt believed that neural mechanisms drove both affect and emotion, he also stated that affective qualities infused all emotions but that emotions required a higher and more complex level of description. He did not, however, specify exactly how affect might link to emotion (Wundt, 1897). Following these historical claims and more recent arguments (Russell & Barrett, 1999), we also suspect that broad dimensions of affect underpin more specific categorical emotions. One intriguing possibility is that different movements through affective space (or "affect dynamics") might imply more categorical emotional states (Kirkland & Cunningham, 2012; Nielsen, Knutson, & Carstensen, 2008). While elegant measures of affect dynamics have been used to describe changes in experience at longer timescales of hours or days (Kuppens, 2015; Kuppens, Oravecz, & Tuerlinckx, 2010), a challenging but tantalizing line of future research might attempt to map affect dynamics at the more rapid timescale of seconds—which might most closely match the neural and affective measures described above (Knutson et al., 2014).

Connecting affect dynamics to emotion at matching temporal resolution might in turn demonstrate that affective qualities and their dynamics underlie different categorical emotions. For instance, starting from an affective baseline state, movement up and to the right might imply excitement, to the right happiness, down and to the right calmness, down and to the left sadness, to the left anger, and up and to the left anxiety (all predictions which would require verification with empirical data). Linking neural and affective levels of analysis might provide a framework for charting out these affect dynamics, which could be tested for specific mapping to temporally precise probes of emotional experience (see also Kirkland & Cunningham, 2011). Further avenues for exploration might include individual differences in affect dynamics and their relationship to emotional traits as well as psychiatric symptom profiles (Davidson, 2015). If affect dynamic probes can yield reliable and valid results, they might be used to assess the impact of various interventions (ranging

from psychological to pharmacological). Thus, affect dynamic probes might eventually improve the accuracy of diagnoses as well as the tracking of changes in psychiatric symptoms.

Self-Awareness Some theorists have asserted that affective experience requires self-aware reflection, and possibly verbal representation (e.g., Barrett, Mesquita, Ochsner, & Gross, 2007; LeDoux, 2012). Based on the lack of a strong association between brain activity and self-reported emotional experience in earlier neuroimaging studies, these theorists have argued that subcortical neural circuits implicated in anticipatory affect cannot generate emotional experience in humans. The proposition that affective experience requires self-reflective awareness is interesting because studies of lesioned patients (e.g., Stuss, Gow, & Hetherington, 1992) as well as neuroimaging research on healthy individuals (e.g., Northoff et al., 2006) have implicated the prefrontal cortex in self-reflective awareness. Current evidence linking brain activity to affective experience, however, contradicts these assertions by demonstrating that when measures with matching resolution are employed, subcortical brain activity can correlate with self-reported affective experience (i.e., NAcc activity with positive arousal, and AIns activity with general arousal; Knutson & Greer, 2008). Associations of subcortical activity with self-reported affect, however, are often fragile and not large. Future research might profitably explore where, when, and in whom neural activity most robustly correlates with affective experience. Assuming the use of measures with matching resolution, one surprising implication of the linking levels approach is that when brain activity and self-report fail to converge, brain activity may provide a better index of affective experience and associated behavioral tendencies than does self-reported experience. For instance, in stimulant users, NAcc responses to drug cues can predict relapse months later, even when self-reported affect cannot (MacNiven et al., 2018).

Contributions

Philosophical Tractability Demonstrations of causal influence across levels of analysis can refute at least two contrasting views of mental function. The first view, dualism, presumes that body (or brain) and mind exist on separate and mostly unconnected levels of analysis (e.g., Descartes, 1641). Demonstrating that perturbation of neural activity can alter affective experience or motivated behavior suggests that although components exist at different levels of analysis and can be measured separately, components at one level are connected to and can causally influence components at another level. The second view, reductionism (Nagel, 2007), implies that all higher levels of analysis can be reduced to lower levels of analysis. The separation of levels with respect to distinct components, temporally resolved sequential responses, and probabilistic causal influence implies that different levels can still be related. The present view further makes room for a type of "expansionism,"

since components at lower levels can influence those at a higher level, but likely in combination with many other components inside and outside of that higher level. Based on a deep science approach, demonstrating a lower-level component's necessity for influencing a higher-level component need not imply sufficiency. In fact, the deep science approach offers an intermediate vision that falls between the extremes of dualism and reductionism, and remains capable of preserving distinctions between levels of analysis while simultaneously tracing causal links that connect them.

Causal Impact The linking levels framework thus implies not only the first two scientific goals of description and explanation, but also the last two scientific goals of prediction and control (Watson, 1913). The surveyed findings that link brain activity to anticipatory affect to motivated behavior over the short span of two decades indicate that researchers have moved beyond description and explanation to prediction. The ability of these findings to not only account for but also to predict choice has partially spurred the birth and growth of new hybrid fields of scientific inquiry (e.g., neuroeconomics, neurofinance, neuromarketing, decision neuroscience, consumer neuroscience, and others). Demonstrating causal links across levels of analysis also implies control (limited by inevitable noise and multicausality). Specifically, manipulating a component at one level should have the causal capacity to alter a linked component at an adjacent but higher level.

New tools developed for precise neural manipulations now make possible identification of these linked components, as well as subsequent tests of control (Namburi, Al-Hasani, Calhoon, Bruchas, & Tye, 2016). For instance, optogenetic manipulations of midbrain dopamine neural firing increase ventral striatal FMRI activity, which elicits approach toward self-administration of the optogenetic stimulus (Ferenczi et al., 2016). Identifying these causal links across levels of analysis can then lead to new predictions and tests of control. For example, recent research has indicated that reward anticipation proportionally induces low frequency electrophysiological activity in the NAcc (i.e., in the delta range), and further, that electrical interference with these signals temporarily halted an animal's approach toward appetizing stimuli (e.g., high-fat food; Wu et al., 2017). Thus, consistent with causal links across levels of analysis, manipulating brain activity necessary for anticipatory affect and associated motivated behavior can change the course of that behavior. Demonstrations of causal influence across levels of analysis could inspire more precisely targeted interventions. These interventions might include "closed loop control"—in which a device detects and then interferes with a predictive neural signature to prevent the onset of a pathological experience or behavior (Grosenick, Marshel, & Deisseroth, 2015).

Metaphorical Reframing The goal of linking levels invites reconsideration not only of lower levels of analysis (e.g., physiology) but also higher levels (e.g., purpose) (Table 7.2). Theorists have often based their metaphors for the mind on its assumed general function. Thus, behaviorists favored a reflex metaphor for the mind based on the ability of reflexes to reliably and rapidly translate input into output, whereas cognitivists favored a computer metaphor for the mind based on the capacity

of computers to faithfully process information. Here, we propose an adaptive metaphor for a mind that prioritizes survival and procreation. Such a mind would ideally need to rapidly anticipate, detect, and compare opportunities with threats in order to promote approach or avoidance. A concise phrase that captures these functions, alluded to earlier, is the "hedonic sharpener." In contrast to "computer" or "reflex" metaphors, the overarching goal of a hedonic sharpener is neither accuracy nor consistency, but rather rapid action in the service of maximizing pleasant feelings and minimizing unpleasant ones. These feelings presumably signaled potential increases or decreases in fitness and motivated appropriate behavior in the ancestral past (Panksepp, Knutson, & Burgdorf, 2002). The hedonic sharpener metaphor not only implies novel underlying components (e.g., gain anticipation, loss anticipation, value integration, motivated action), but might also better account for behavior that might appear anomalous or suboptimal in the context of alternative reflex or computer metaphors (e.g., reliance on quick heuristics, overconfidence, confirmation bias, biased assimilation of positive versus negative feedback, etc.). One counterintuitive but testable implication of this metaphorical reframing is that in the case of a reflex or computer, input should be more correlated with output than intermediate processing (since information degrades with processing). In the case of the hedonic sharpener, however, intermediate processing should be more correlated with output than input, since the goal of the system is not to faithfully represent incoming information but rather to transform it in a way that facilitates rapid adaptive action.

Conclusion Instead of a closed system, a deep science approach offers an open framework that can be extended or modified by new findings. Thus, the initial links described here raise more questions than they answer. Still, recent findings have clearly begun to link neural activity, anticipatory affect, and motivated behavior. These advances have been enabled by theoretical recognition of the influence of anticipatory affect on motivated behavior and methodological advances in measuring concepts at matching resolution. Based on the speed and promise of these advances, linking levels of analysis may provide the most direct path from the scientific goals of description and explanation to those of prediction and control. By linking previously disparate levels of analysis, the deep science approach could accelerate the development of effective interventions for enhancing human health and well-being.

Acknowledgments We thank Ingrid Haas, Yuan Chang Leong, Maital Neta, and Jeanne Tsai for feedback on earlier drafts. During manuscript preparation, the author was supported by a Wu Tsai Stanford Neurosciences Institute Grant to the NeuroChoice Initiative.

References

Adcock, R. A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., & Gabrieli, J. D. E. (2006). Reward-motivated learning: mesolimbic activation precedes memory formation. *Neuron, 50*(3), 507–517. https://doi.org/10.1016/j.neuron.2006.03.036

Ariely, D., & Berns, G. S. (2010). Neuromarketing: The hope and hype of neuroimaging in business. *Nature Reviews Neuroscience, 11*, 284–292. https://doi.org/10.1038/nrn2795

Bandettini, P. A., Wong, E. C., Hinks, R. S., Tikofsky, R. S., & Hyde, J. S. (1992). Time course EPI during task activation. *Magnetic Resonance in Medicine, 25*(2), 390–397. https://doi.org/10.1002/mrm.1910250220

Bargmann, C. I. (2012). Beyond the connectome: How neuromodulators shape neural circuits. *BioEssays, 34*(6), 458–465. https://doi.org/10.1002/bies.201100185

Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annual Review of Psychology, 58*, 373–403. https://doi.org/10.1146/annurev.psych.58.110405.085709

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage, 76*, 412–427. https://doi.org/10.1016/j.neuroimage.2013.02.063

Bechara, A., Tranel, D., Damasio, H., & Damasio, A. R. (1996). Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cerebral Cortex, 6*(2), 215–225. https://doi.org/10.1093/cercor/6.2.215

Berns, G. S., & Moore, S. E. (2012). A neural predictor of cultural popularity. *Journal of Consumer Psychology, 22*(1), 154–160. https://doi.org/10.1016/j.jcps.2011.05.001

Berridge, K. C. (2004). Motivation concepts in behavioral neuroscience. *Physiology and Behavior, 81*(2), 179–209. https://doi.org/10.1016/j.physbeh.2004.02.004

Berridge, K. C., & Robinson, T. E. (1998). What is the role of dopamine in reward: Hedonic impact, reward learning, or incentive salience? *Brain Research Reviews, 28*, 309–369. https://doi.org/10.1016/S0165-0173(98)00019-8

Cacioppo, J. T., & Berntson, G. G. (1992). Social Psychological Contributions to the Decade of the Brain: Doctrine of Multilevel Analysis. *American Psychologist, 47*(8), 1019–1028. https://doi.org/10.1037/0003-066X.47.8.1019

Canessa, N., Crespi, C., Motterlini, M., Baud-Bovy, G., Chierchia, G., Pantaleo, G., … Cappa, S. F. (2013). The functional and structural neural basis of individual differences in loss aversion. *Journal of Neuroscience, 33*(36), 14307–14317. https://doi.org/10.1523/jneurosci.0497-13.2013

Clithero, J. A., & Rangel, A. (2013). Informatic parcellation of the network involved in the computation of subjective value. *Social Cognitive and Affective Neuroscience, 9*(9), 1289–1302. https://doi.org/10.1093/scan/nst106

Cooper, J. C., & Knutson, B. (2008). Valence and salience contribute to nucleus accumbens activation. *NeuroImage, 29*, 538–547. https://doi.org/10.1016/j.neuroimage.2007.08.009

Craig, W. (1918). Appetites and aversions as constituents of instincts. *The Biological Bulletin, 34*(2), 91–107. https://doi.org/10.2307/1536346

Davidson, R. J. (2015). Comment: Affective chronometry has come of age. *Emotion Review, 7*(4), 368–370. https://doi.org/10.1177/1754073915590844

Decot, H. K., Namboodiri, V. M. K., Gao, W., McHenry, J. A., Jennings, J. H., Lee, S. H., … Stuber, G. D. (2017). Coordination of brain-wide activity dynamics by dopaminergic neurons. *Neuropsychopharmacology, 42*(3), 615–627. https://doi.org/10.1038/npp.2016.151

Delgado, M. R., Nystrom, L. E., Fissell, C., Noll, D. C., & Fiez, J. A. (2000). Tracking the hemodynamic responses to reward and punishment in the striatum. *Journal of Neurophysiology, 84*(6), 3072–3077. https://doi.org/10.1152/jn.2000.84.6.3072

Demos, K. E., Heatherton, T. F., & Kelley, W. M. (2012). Individual differences in nucleus accumbens activity to food and sexual images predict weight gain and sexual behavior. *Journal of Neuroscience, 32*(16), 5549–5552. https://doi.org/10.1523/jneurosci.5958-11.2012

Descartes, R. (1641). *Meditation IV. In the philosophical works of descartes* (pp. 1–33). Cambridge, England: Cambridge University Press.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron, 73*(3), 415–434. https://doi.org/10.1016/j.neuron.2012.01.010

Diekhof, E. K., Kaps, L., Falkai, P., & Gruber, O. (2012). The role of the human ventral striatum and the medial orbitofrontal cortex in the representation of reward magnitude: An activation

likelihood estimation meta-analysis of neuroimaging studies of passive reward expectancy and outcome processing. *Neuropsychologia, 50*(7), 1252–1266. https://doi.org/10.1016/j.neuropsychologia.2012.02.007

Elliott, R., Friston, K. J., & Dolan, R. J. (2000). Dissociable neural responses in human reward systems. *The Journal of Neuroscience, 20*(16), 6159–6165. https://doi.org/10.1523/JNEUROSCI.20-16-06159.2000

Engel, S. A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., Chichilnisky, E. J., & Shadlen, M. N. (1994). FMRI of human visual cortex. *Nature, 369*, 525. https://doi.org/10.1038/369525a0

Engelmann, J. B., Meyer, F., Fehr, E., & Ruff, C. C. (2015). Anticipatory anxiety disrupts neural valuation during risky choice. *Journal of Neuroscience, 35*(7), 3085–3099. https://doi.org/10.1523/jneurosci.2880-14.2015

Falk, E. B., Berkman, E. T., & Lieberman, M. D. (2012). From neural responses to population behavior: Neural focus group predicts population-level media effects. *Psychological Science, 23*(5), 439–445. https://doi.org/10.1177/0956797611434964

Ferenczi, E. A., Zalocusky, K. A., Liston, C., Grosenick, L., Warden, M. R., Amatya, D., … Deisseroth, K. (2016). Prefrontal cortical regulation of brainwide circuit dynamics and reward-related behavior. *Science, 351*, aac9698. https://doi.org/10.1126/science.aac9698

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making, 13*(1), 1–17. https://doi.org/10.1002/(SICI)1099-0771(200001/03)13:1<1::AID-BDM333>3.0.CO;2-S

Genevsky, A., & Knutson, B. (2015). Neural affective mechanisms predict market-level microlending. *Psychological Science, 26*(9), 1411–1422. https://doi.org/10.1177/0956797615588467

Genevsky, A., Västfjäll, D., Slovic, P., & Knutson, B. (2013). Neural underpinnings of the identifiable victim effect: Affect shifts preferences for giving. *The Journal of Neuroscience, 33*(34), 17188–17196. https://doi.org/10.1523/jneurosci.2348-13.2013

Genevsky, A., Yoon, C., & Knutson, B. (2017). When brain beats behavior: Neuroforecasting crowdfunding outcomes. *The Journal of Neuroscience, 37*(36), 8625–8634. https://doi.org/10.1523/JNEUROSCI.1633-16.2017

Grosenick, L., Marshel, J. H., & Deisseroth, K. (2015). Closed-loop and activity-guided optogenetic control. *Neuron, 86*(1), 106–139. https://doi.org/10.1016/j.neuron.2015.03.034

Haber, S. N., & Knutson, B. (2010). The reward circuit: Linking primate anatomy and human imaging. *Neuropsychopharmacology, 35*(1), 4–26. https://doi.org/10.1038/npp.2009.129

Hampton, A. N., & O'Doherty, J. P. (2007). Decoding the neural substrates of reward-related decision making with functional MRI. *Proceedings of the National Academy of Sciences, 104*(4), 1377–1382. https://doi.org/10.1073/pnas.0606297104

Harbaugh, W. T., Mayr, U., & Burghart, D. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science, 316*(5831), 1622–1626. https://doi.org/10.1126/science.1140738

Harlé, K. M., & Sanfey, A. G. (2007). Incidental Sadness Biases Social Economic Decisions in the Ultimatum Game. *Emotion, 7*(4), 876–881. https://doi.org/10.1037/1528-3542.7.4.876

Hess, W. R. (1958). *The functional organization of the diencephalon*. New York, NY: Grune & Stratton.

Huang, Y. F., Soon, C. S., Mullette-Gillman, O. A., & Hsieh, P. J. (2014). Pre-existing brain states predict risky choices. *NeuroImage, 101*, 466–472. https://doi.org/10.1016/j.neuroimage.2014.07.036

Hyman, S. E., Malenka, R. C., & Nestler, E. J. (2006). Neural mechanisms of addiction: The role of reward-related learning and memory. *Annual Review of Neuroscience, 29*(1), 565–598. https://doi.org/10.1146/annurev.neuro.29.051605.113009

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., … Wang, P. (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *The American Journal of Psychiatry, 167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Karmarkar, U., Shiv, B., & Knutson, B. (2015). Cost conscious? The neural and behavioral impact of price primacy on decision making. *Journal of Marketing Research, 52*(4), 467–481. https://doi.org/10.1509/jmr.13.0488

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science, 308*(5718), 78–83. https://doi.org/10.1126/science.1108062

Kirkland, T., & Cunningham, W. A. (2011). Neural basis of affect and emotion. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*, 656–665. https://doi.org/10.1002/wcs.145

Kirkland, T., & Cunningham, W. A. (2012). Mapping emotions through time: how affective trajectories inform the language of emotion. *Emotion, 12*(2), 268–282. https://doi.org/10.1037/a0024218

Knutson, B. (2016). Deep science. Retrieved from https://www.edge.org/response-detail/26758

Knutson, B., Adams, C. M., Fong, G. W., & Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *The Journal of Neuroscience, 21*(16), 1–5. https://doi.org/10.1523/JNEUROSCI.21-16-j0002.2001

Knutson, B., Fong, G. W., Adams, C. M., Varner, J. L., & Hommer, D. (2001). Dissociation of reward anticipation and outcome with event-related fMRI. *NeuroReport, 12*(17), 3683–3687. https://doi.org/10.1097/00001756-200112040-00016

Knutson, B., Fong, G. W., Bennett, S. M., Adams, C. M., & Hommer, D. (2003). A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: Characterization with rapid event-related fMRI. *NeuroImage, 18*(2), 263–272. https://doi.org/10.1016/S1053-8119(02)00057-5

Knutson, B., & Genevsky, A. (2018). Neuroforecasting aggregate choice. *Current Directions in Psychological Science, 27*(2), 110–115. https://doi.org/10.1177/0963721417737877

Knutson, B., & Gibbs, S. (2007). Linking nucleus accumbens dopamine and blood oxygenation. *Psychopharmacology, 191*, 813–822. https://doi.org/10.1007/s00213-006-0686-7

Knutson, B., & Greer, S. (2008). Anticipatory affect: Neural correlates and consequences for choice. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1511), 3771–3786. https://doi.org/10.1098/rstb.2008.0155

Knutson, B., & Huettel, S. (2015). The risk matrix. *Current Opinion in Behavioral Sciences, 5*, 141–146. https://doi.org/10.1016/j.cobeha.2015.10.012

Knutson, B., & Karmarkar, U. (2014). Appetite, consumption, and choice in the human brain. In *The interdisciplinary science of consumption* (pp. 163–184). Cambridge, MA: The MIT Press. https://doi.org/10.7551/mitpress/9780262027670.003.0009

Knutson, B., Katovich, K., & Suri, G. (2014). Inferring affect from fMRI data. *Trends in Cognitive Sciences, 18*(8), 422–428. https://doi.org/10.1016/j.tics.2014.04.006

Knutson, B., Rick, S., Wimmer, G. E., Prelec, D., & Loewenstein, G. (2007). Neural predictors of purchases. *Neuron, 53*(1), 147–156. https://doi.org/10.1016/j.neuron.2006.11.010

Knutson, B., Taylor, J., Kaufman, M., Peterson, R., & Glover, G. (2005). Distributed neural representation of expected value. *The Journal of Neuroscience, 25*(19), 4806–4812. https://doi.org/10.1523/JNEUROSCI.0642-05.2005

Knutson, B., Westdorp, A., Kaiser, E., & Hommer, D. (2000). FMRI visualization of brain activity during a monetary incentive delay task. *NeuroImage, 12*(1), 20–27. https://doi.org/10.1006/nimg.2000.0593

Knutson, B., & Wimmer, G. E. (2007). Reward neural circuitry for social valuation. In *Social neuroscience: integrating biological and psychological explanations of social behavior* (pp. 157–175). New York, NY: Guilford Press.

Knutson, B., Wimmer, G. E., Kuhnen, C. M., & Winkielman, P. (2008). Nucleus accumbens activation mediates the influence of reward cues on financial risk taking. *NeuroReport, 19*(5), 509–513. https://doi.org/10.1097/WNR.0b013e3282f85c01

Koenigs, M., & Tranel, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: Evidence from the ultimatum game. *Journal of Neuroscience, 27*(4), 951–956. https://doi.org/10.1523/jneurosci.4606-06.2007

Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., … Grafman, J. (2007). Neural correlates of trust. *Proceedings of the National Academy of Sciences, 104*(50), 20084–20089. https://doi.org/10.1073/pnas.0710103104

Kruschwitz, J. D., Waller, L., List, D., Wisniewski, D., Ludwig, V. U., Korb, F., … Walter, H. (2018). Anticipating the good and the bad: A study on the neural correlates of bivalent emotion anticipation and their malleability via attentional deployment. *NeuroImage, 183*, 553–564. https://doi.org/10.1016/j.neuroimage.2018.08.048

Kühn, S., & Gallinat, J. (2012). The neural correlates of subjective pleasantness. *NeuroImage, 61*, 289–294. https://doi.org/10.1016/j.neuroimage.2012.02.065

Kühn, S., Strelow, E., & Gallinat, J. (2016). Multiple "buy buttons" in the brain: Forecasting chocolate sales at point-of-sale based on functional brain activation using fMRI. *NeuroImage, 136*, 122–128. https://doi.org/10.1016/j.neuroimage.2016.05.021

Kuhnen, C. M., & Knutson, B. (2005). The neural basis of financial risk taking. *Neuron, 47*(5), 763–770. https://doi.org/10.1016/j.neuron.2005.08.008

Kuppens, P. (2015). It's about time: A special section on affect dynamics. *Emotion Review, 7*(4), 297–300. https://doi.org/10.1177/1754073915590947

Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology, 99*(6), 1042–1060. https://doi.org/10.1037/a0020962

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., … Turner, R. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences, 89*(12), 5675–5679. https://doi.org/10.1073/PNAS.89.12.5675

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., Greenwald, M. K., & Bradley, M. M. (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

Lebreton, M., Jorge, S., Michel, V., Thirion, B., & Pessiglione, M. (2009). An automatic valuation system in the human brain: Evidence from functional neuroimaging. *Neuron, 64*(3), 431–439. https://doi.org/10.1016/j.neuron.2009.09.040

LeDoux, J. (2012). Rethinking the emotional brain. *Neuron, 73*(4), 653–676. https://doi.org/10.1016/j.neuron.2012.02.004

Leong, J. K., Pestilli, F., Wu, C. C., Samanez-Larkin, G. R., & Knutson, B. (2016). White-matter tract connecting anterior insula to nucleus accumbens correlates with reduced preference for positively skewed gambles. *Neuron, 89*(1), 63–69. https://doi.org/10.1016/j.neuron.2015.12.015

Levy, D. J., & Glimcher, P. W. (2012). The root of all value: A neural common currency for choice. *Current Opinion in Neurobiology, 22*(6), 1027–1038. https://doi.org/10.1016/j.conb.2012.06.001

Levy, I., Lazzaro, S. C., Rutledge, R. B., & Glimcher, P. W. (2011). Choice from non-choice: predicting consumer preferences from blood oxygenation level-dependent signals obtained during passive viewing. *The Journal of Neuroscience, 31*(1), 118–125. https://doi.org/10.1523/JNEUROSCI.3214-10.2011

Litt, A., Plassmann, H., Shiv, B., & Rangel, A. (2011). Dissociating valuation and saliency signals during decision-making. *Cerebral Cortex, 21*(1), 95–102. https://doi.org/10.1093/cercor/bhq065

Liu, X., Hairston, J., Schrier, M., & Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: A meta-analysis of functional neuroimaging studies. *Neuroscience and Biobehavioral Reviews, 35*(5), 1219–1236. https://doi.org/10.1016/j.neubiorev.2010.12.012

Loewenstein, G. F., Hsee, C. K., Weber, E. U., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin, 127*(2), 267–286. https://doi.org/10.1037/0033-2909.127.2.267

Lohani, S., Poplawsky, A. J., Kim, S. G., & Moghaddam, B. (2017). Unexpected global impact of VTA dopamine neuron activation as measured by opto-fMRI. *Molecular Psychiatry, 22*(4), 585–594. https://doi.org/10.1038/mp.2016.102

MacNiven, K. H., Jensen, E. L. S., Borg, N., Padula, C. B., Humphreys, K., & Knutson, B. (2018). Association of neural responses to drug cues with subsequent relapse to stimulant use. *JAMA Network Open, 1*(8), 1–14. https://doi.org/10.1001/jamanetworkopen.2018.6466

Marr, D. (1982). *Vision*. New York, NY: W. H. Freeman and Company.

Nagel, T. (1998). Reductionism and antireductionism. *The Limits of Reductionism in Biology, 213*, 3–14. https://doi.org/10.1002/9780470515488.ch2

Namburi, P., Al-Hasani, R., Calhoon, G. G., Bruchas, M. R., & Tye, K. M. (2016). Architectural representation of valence in the limbic system. *Neuropsychopharmacology, 41*(7), 1697–1715. https://doi.org/10.1038/npp.2015.358

Nielsen, L., Knutson, B., & Carstensen, L. L. (2008). Affect dynamics, affective forecasting, and aging. *Emotion, 8*(3), 318–330. https://doi.org/10.1037/1528-3542.8.3.318

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain-A meta-analysis of imaging studies on the self. *NeuroImage, 31*(1), 440–457. https://doi.org/10.1016/j.neuroimage.2005.12.002

O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J., & Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience, 4*(1), 95–102. https://doi.org/10.1038/82959

O'Doherty, J. P., Deichmann, R., Critchley, H. D., & Dolan, R. J. (2002). Neural responses during anticipation of a primary taste reward. *Neuron, 33*(5), 815–826. https://doi.org/10.1016/S0896-6273(02)00603-7

Olds, J. (1955). Physiological mechanisms of reward. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1955* (pp. 73–139). Lincoln, NE: University of Nebraska Press.

Olds, J., & Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *Journal of Comparative and Physiological Psychology, 47*, 419–427.

Olds, M. E., & Fobes, J. L. (1981). Activity responses to morphine and amphetamine in rats with elevated NE levels in the pons. *Pharmacology, Biochemistry and Behavior, 15*(2), 167–171. https://doi.org/10.1016/0091-3057(81)90172-6

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Oxford, England: Illinois Press.

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. New York, NY: Oxford University Press.

Panksepp, J., Knutson, B., & Burgdorf, J. (2002). The role of brain emotional systems in addictions: A neuro-evolutionary perspective and new "self-report" animal model. *Addiction, 97*(4), 459–469. https://doi.org/10.1046/j.1360-0443.2002.00025.x

Park, B. K., Blevins, E., Knutson, B., & Tsai, J. L. (2017). Neurocultural evidence that ideal affect match promotes giving. *Social Cognitive and Affective Neuroscience, 12*(7), 1083–1096. https://doi.org/10.1093/scan/nsx047

Pessiglione, M., Petrovic, P., Daunizeau, J., Palminteri, S., Dolan, R. J., & 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

Preuschoff, K., Quartz, S. R., & Bossaerts, P. (2008). Human insula activation reflects risk prediction errors as well as risk. *The Journal of Neuroscience, 28*(11), 2745–2752. https://doi.org/10.1523/JNEUROSCI.4286-07.2008

Ramnani, N., Elliott, R., Athwal, B. S., & Passingham, R. E. (2004). Prediction error for free monetary reward in the human prefrontal cortex. *NeuroImage, 23*(3), 777–786. https://doi.org/10.1016/j.neuroimage.2004.07.028

Rao, S. M., Binder, J. R., Hammeke, T. A., Bandettini, P. A., Bobholz, J. A., Frost, J. A., … Hyde, J. S. (1995). Somatotopic mapping of the human primary motor cortex with functional magnetic resonance imaging. *Neurology, 45*(5), 919–924. https://doi.org/10.1212/WNL.45.5.919

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A neural basis for social cooperation. *Neuron, 35*(2), 395–405. https://doi.org/10.1016/S0896-6273(02)00755-9

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004). Opposing BOLD responses to reciprocated and unreciprocated altruism in putative reward pathways. *NeuroReport, 15*(16), 2539–2543. https://doi.org/10.1097/00001756-200411150-00022

Robinson, D. L., Venton, B. J., Heien, M. L. A. V., & Wightman, R. M. (2003). Detecting subsecond dopamine release with fast-scan cyclic voltammetry in vivo. *Clinical Chemistry, 49*(10), 1763–1773. https://doi.org/10.1373/49.10.1763

Ruff, C. C., & Fehr, E. (2014). The neurobiology of rewards and values in social decision making. *Nature Reviews Neuroscience, 15*, 549–562. https://doi.org/10.1038/nrn3776

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1987). *Parallel distributed processing* (Vol. 1). Cambridge, MA: MIT Press.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161–1178. https://doi.org/10.1037/h0077714

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

Salimpoor, V. N., Van Den Bosch, I., Kovacevic, N., McIntosh, A. R., Dagher, A., & Zatorre, R. J. (2013). Interactions between the nucleus accumbens and auditory cortices predict music reward value. *Science, 340*(6129), 216–219. https://doi.org/10.1126/science.1231059

Samanez-Larkin, G., Gibbs, S., Khanna, K., Nielsen, L., Carstensen, L., & Knutson, B. (2007). Anticipation of monetary gain but not loss in healthy older adults. *Nature Neuroscience, 10*(6), 787–791. https://doi.org/10.1038/nn1894

Samanez-Larkin, G. R., & Knutson, B. (2015). Decision making in the ageing brain: Changes in affective and motivational circuits. *Nature Reviews Neuroscience, 16*(5), 278–289. https://doi.org/10.1038/nrn3917

Sanfey, A. G. (2007). Social decision-making: Insights from game theory and neuroscience. *Science, 318*(5850), 598–602. https://doi.org/10.1126/science.1142996

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

Scholz, C., Baek, E. C., O'Donnell, M. B., Kim, H. S., Cappella, J. N., & Falk, E. B. (2017). A neural model of valuation and information virality. *Proceedings of the National Academy of Sciences, 114*(11), 2881–2886. https://doi.org/10.1073/pnas.1615259114

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Sejnowski, T. J., Churchland, P. S., & Movshon, J. A. (2014). Putting big data to good use in neuroscience. *Nature Neuroscience, 17*(11), 1440–1441. https://doi.org/10.1038/nn.3839

Sescousse, G., Caldú, X., Segura, B., & Dreher, J. C. (2013). Processing of primary and secondary rewards: A quantitative meta-analysis and review of human functional neuroimaging studies. *Neuroscience and Biobehavioral Reviews, 37*(4), 681–696. https://doi.org/10.1016/j.neubiorev.2013.02.002

Smith, A., Douglas Bernheim, B., Camerer, C. F., & Rangel, A. (2014). Neural activity reveals preferences without choices. *American Economic Journal: Microeconomics, 6*(2), 1–36. https://doi.org/10.1257/mic.6.2.1

Smith, D. V., Hayden, B. Y., Truong, T. K., Song, A. W., Platt, M. L., & Huettel, S. A. (2010). Distinct value signals in anterior and posterior ventromedial prefrontal cortex. *Journal of Neuroscience, 30*(7), 2490–2495. https://doi.org/10.1523/JNEUROSCI.3319-09.2010

Stuss, D. T., Gow, C. A., & Hetherington, C. R. (1992). "No longer gage": Frontal lobe dysfunction and emotional changes. *Journal of Consulting and Clinical Psychology, 60*(3), 349–359. https://doi.org/10.1037/0022-006X.60.3.349

Thayer, R. E. (1989). *The biopsychology of mood and arousal. Personality and Individual Differences*. New York, NY: Oxford University Press. https://doi.org/10.1016/0191-8869(90)90284-X

Tricomi, E. M., Delgado, M. R., & Fiez, J. A. (2004). Modulation of Caudate Activity by Action Contingency. *Neuron, 41*(2), 281–292. https://doi.org/10.1016/S0896-6273(03)00848-1

Tusche, A., Bode, S., & Haynes, J.-D. (2010). Neural Responses to Unattended Products Predict Later Consumer Choices. *Journal of Neuroscience, 30*(23), 8024–8031. https://doi.org/10.1523/jneurosci.0064-10.2010

Venkatraman, V., Dimoka, A., Pavlou, P. A., Vo, K., Hampton, W., Bollinger, B., … Winer, R. S. (2015). Predicting advertising success beyond traditional measures: New insights from neurophysiological methods and market response modeling. *Journal of Marketing Research, 52*(4), 436–452. https://doi.org/10.1509/jmr.13.0593

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98*(2), 219–235. https://doi.org/10.1037/0033-2909.98.2.219

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology, 76*(5), 820–838. https://doi.org/10.1037/0022-3514.76.5.820

Watson, J. B. (1913). Psychology as the behaviourist views it. *Psychological Review, 20*(2), 158–177. https://doi.org/10.1037/h0074428

Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. *Advances in Experimental Social Psychology, 35*, 345–411. https://doi.org/10.1016/j.pain.2011.02.015

Witten, I. B., Steinberg, E. E., Lee, S. Y., Davidson, T. J., Zalocusky, K. A., Brodsky, M., … Deisseroth, K. (2011). Recombinase-driver rat lines: Tools, techniques, and optogenetic application to dopamine-mediated reinforcement. *Neuron, 72*(5), 721–733. https://doi.org/10.1016/j.neuron.2011.10.028

Wu, C. C., Bossaerts, P., & Knutson, B. (2011). The affective impact of financial skewness on neural activity and choice. *PLoS One, 6*(2), 1–7. https://doi.org/10.1371/journal.pone.0016838

Wu, C. C., Sacchet, M. D., & Knutson, B. (2012). Toward an affective neuroscience account of financial risk taking. *Frontiers in Neuroscience, 6*(159), 1–10. https://doi.org/10.3389/fnins.2012.00159

Wu, H., Miller, K. J., Blumenfeld, Z., Williams, N. R., Ravikumar, V. K., Lee, K. E., … Halpern, C. H. (2018). Closing the loop on impulsivity via nucleus accumbens delta-band activity in mice and man. *Proceedings of the National Academy of Sciences of the United States of America, 115*(1), 192–197. https://doi.org/10.1073/pnas.1712214114

Wundt, W. (1897). *Outlines of psychology*. London, England: Williams and Norgate. https://doi.org/10.1037/12908-000

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

Zink, C. F., Pagnoni, G., Martin-Skurski, M. E., Chappelow, J. C., & Berns, G. S. (2004). Human striatal responses to monetary reward depend on saliency. *Neuron, 42*(3), 509–517. https://doi.org/10.1016/S0896-6273(04)00183-7

Chapter 8
Reproducible, Generalizable Brain Models of Affective Processes

Philip Kragel and Tor D. Wager

> Nature has placed mankind under the governance of two sovereign masters, pain and pleasure. On the one hand the standard of right and wrong, on the other the chain of causes and effects, are fastened to their throne. They govern us in all we do, in all we say, in all we think…—Jeremy Bentham, The Principles of Morals and Legislation

Pain and pleasure are primary motivating forces that underlie much of human behavior. Take pain, for example. It is defined as an aversive sensory and emotional experience. Generally, we avoid it. Many of our philosophical and religious traditions are focused around advice on how to escape, avoid, manage, or accept its reality without causing additional suffering.

But what is pain, exactly? Pain is multiple things. It is an experience, a motivating force, an elicitor of emotional responses, a driver of decisions. Sometimes it is, more or less, a "negative sensory and emotional experience" caused by activation in nociceptive pathways. But it cannot be only that, because it teaches us to fear pain in the future, and it drives our autonomic systems and the musculature that allows us to escape. It motivates goals, from the simple and immediate—take your hand off the stove!—to the elaborate and complex, even becoming a focal point around which one's life is organized. It teaches us to fear it, but sometimes also to seek it, as during the expiation of guilt or when we turn to pain to relieve emotional distress. And because it is an experience, it is consciously accessible. We will never fully understand pain until we understand consciousness.

P. Kragel
Institute of Cognitive Science, University of Colorado, Boulder, CO, USA

T. D. Wager (✉)
Institute of Cognitive Science, University of Colorado, Boulder, CO, USA

Department of Psychology and Neuroscience and the Institute of Cognitive Science, University of Colorado, Boulder, CO, USA

Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA
e-mail: tor.wager@colorado.edu

M. Neta, I. J. Haas (eds.), *Emotion in the Mind and Body*, Nebraska Symposium on Motivation 66, https://doi.org/10.1007/978-3-030-27473-3_8

Identifying the brain representations that underlie pain is a crucial step towards understanding it. There are many fundamental questions that remain unanswered. Is pain caused by a specific neural pathway or region? Is the thing that makes pain conscious the same as the thing that makes us conscious of the visual world, or of our own intentions? Is pain a single type of experience, with one neural mechanism, or a family of experiences with diverse mechanisms? What is the relationship between the processes that cause us to experience immediate pain in the moment and those that drive long-term avoidance? Is the aversiveness of physical pain neurologically similar to the aversiveness of other negative experiences—the empathic distress of seeing another person suffer, or the emotional heartbreak of being rejected by a lover or a longtime friend? Are pain and suffering the same thing, and must the one cause the other?

Without understanding the brain processes involved, our answers to these questions will be definitional. People use similar words to describe pain and emotional distress (MacDonald, 2009)—so by definition they are *conceptually* similar. But these similarities may say more about the way we organize our thoughts and the pragmatics of communication than about the deeper nature of our being.

How do we even know when someone is in pain? In one sense, it's simple: We can ask them. But, though it is tempting to take self-report as a ground-truth measure of subjective experience, experience is private. I cannot directly observe whether you are in pain, or whether the color red looks the same to you as it does to me (Chalmers, 2007). And though people's self-reports are often trustworthy, in many cases they fall short. If you burn your hand and you tell me you are experiencing 7 out of 10 pain, that pain report is very likely related to the degree of nociceptive activity traveling through your spinothalamic tract (and other tracts) up to your brain. But it is also related to your prior belief that you've been injured and that it *should* be painful. It is colored by the emotions you feel—are you cool and objective, or afraid of serious injury? Are you angry at yourself for putting your hand on the stove, or at me for leaving the stove on? And it is filtered by what you are trying to communicate. Do I look sorry for causing your injury, or empathetic, or understanding? Or do you need to make sure you are being taken seriously? Your "7" is a *communicative behavior* that results from a *judgment* that is made relative to your past experience—how bad should a "7" be?—and your appraisal of the overall context of the situation.

If we are simply to believe all pain reports in all situations, we will have to accept a number of uncomfortable things. We will have to accept that when people report pain after a fake auto crash—a controlled experiment that looked like a crash but involved no real sudden movement—the whiplash pain they report is just as real as the pain from a real accident (Castro et al., 2001). We will have to accept that males who are strongly gender identified really experience less pain than those who are not (Alabas, Tashani, Tabasam, & Johnson, 2012). We will have to accept that if a faith healer reduces a person's pain report with a sham surgery, magnets, or device that manipulates "auras," that is just as good as a drug that reduces pain by a similar amount. We will have to accept that if I give two groups scales with different anchors, and they report different pain intensities relative to those different anchors,

they really feel different levels of pain (Schwarz, 1999). And we will have to accept that people with intellectual disabilities who do not communicate pain effectively (de Knegt & Scherder, 2011) are not really experiencing it in the same way as the rest of us. This kind of blind trust in self-reports is, in part, what underlies the mistreatment of groups thought to be "incapable" of normal pain, including animals and human infants (Fitzgerald & Walker, 2009).

Brain Representations

The complexity of even "basic" affective processes like pain and pleasure and the limitations inherent in using self-reports as exclusive measures are fundamental issues that have held back the study of motivation and emotion since the inception of the scientific disciplines that study them. The hope, then, is that neuroscientific measures will enable us to identify neurophysiological processes that cause affective experiences—*brain representations* of affective states.

In some sense, studying a defined brain circuit or process that contributes to pain, or any other affective/motivational state, is much simpler than understanding pain reports (or other behaviors) as a whole. This is because behaviors emerge from the interactions among many brain processes. Identifying particular brain circuits and their relationship to overall behavior is a way of beginning to deconstruct those behaviors and thereby understand the elemental ingredients of affect and motivation.

However, the way in which we have historically approached studying the brain has, in many respects, been oversimplified in ways that do not lead to greater understanding. Some of these simplifications have been embedded in the way we analyze brain data and make inferences about the mappings between brain and mind. This chapter outlines some of these difficulties, anchoring on pain, negative affect, and empathy as exemplars. It also presents a new approach to brain-mind mapping, *predictive modeling*, that is gaining traction in the field and promises to help overcome some of the limitations of previous work. Finally, we discuss current progress using predictive modeling to understand some of the brain "ingredients" that contribute to pain, negative affect, and empathy, and how the brain processes involved relate to one another.

Betwixt Simplicity and Complexity: A Middle Road

Representations and Measures

One of the major goals in mapping brain to mind is to develop *brain measures* that capture the underlying *representation* of a mental event. "Representation" is a theoretical construct that came out of cognitive science over the past decades. A "representation" of an object, an orange for example, is an information structure that

describes the properties of the orange (orange-colored, sweet, healthy) and can be linked to actions (eat it, smell it, slice it) and similar objects (lemons, watermelon). We assume that the brain encodes such representations, so that oranges and other objects can be perceived and categorized, acted upon, remembered, and so forth. A *brain representation,* then, should be an obligatory information structure encoded into the brain; without it, one cannot recognize an orange. It should also be sufficient; if I activate a perceptual representation of "orange" in your brain, you will perceive or imagine an orange. Pain and other affective experiences are thought to be encoded similarly. Activating a representation of pain would create the experience of pain and activate at least some of its associated actions and thoughts; suppressing such a representation would block pain.

If we can identify a brain measure that is closely aligned with a representation, that provides a powerful inferential tool that enables testing of interventions. For example, if we can establish a measure of pain based on fMRI activity, that measure becomes a *target* for interventions. We can test whether various interventions modify that pattern; if they do, we might infer that they influence the brain mechanisms that generate pain (Fig. 8.1). Such tests would provide objective measures for

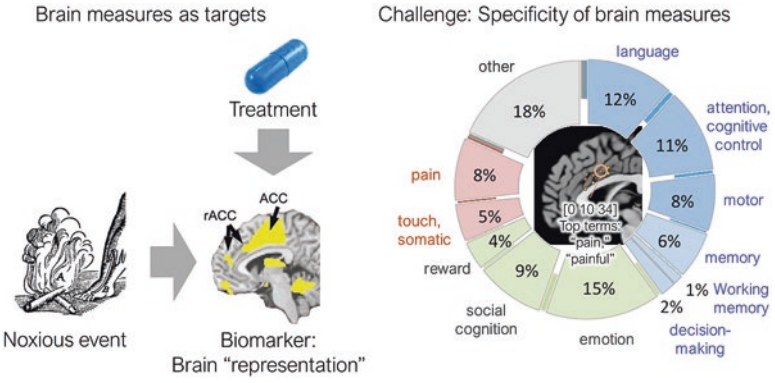

Fig. 8.1 Brain measures and brain targets: The specificity problem. Initially, neuroimaging studies defined brain "representations" of a mental construct, or category of mental events, as areas that responded to an instance of that construct. For example, brain "representations" for pain were defined as regions that responded to a painful stimulus (the image shown is a pain-related map from Wager et al., 2004). The logic was sensible: Define brain markers that correspond to a mental construct, and these would become targets for interventions. One could then compare intervention effects on those targets, characterize their changes across time, and more. However, identifying a brain representation that corresponds to a mental construct is much more difficult than we initially realized. One problem, shown in the right panel, is that individual regions or voxels are rarely highly specific for any category of mental event. The figure shows one of the most pain-selective voxels in the dorsal cingulate cortex. It is activated by approximately 200 studies in the neurosynth.org database (Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011) that correspond to many different mental categories (Wager et al., 2016). Thus, unfortunately, it is a poor measure of any type of mental event, and when testing the effects of an intervention, observing effects on the cingulate cannot tell us whether it affects pain, emotion, decision-making, motor, or language, or social cognitive processes

interventions that are less complex and variable across contexts than self-report, and put psychological, drug, and other interventions on a level playing field, enabling direct comparisons of their effect sizes across interventions.

The trouble is that a representation is an information structure, not a pattern of brain activity. Work on population coding in neuroscience suggests that in many cases, representations of objects, actions, and so forth are encoded in patterns across neurons with different tuning curves (for reviews, see Averbeck, Latham, & Pouget, 2006; Kragel, Koban, Barrett, & Wager, 2018; Sakurai, 1996). But establishing mappings between measurable brain features and "representations" of mental events is very tricky and requires an extensive set of empirical tests—and, where evidence is not available, a leap of faith. Suppose we identify a set of neurons that respond with increased firing rates to a painful event, for example a hot stimulus on the hand. We do not yet know that those neurons "represent" anything, or what they represent. One must ask several questions to get closer to understanding what the neural patterns encode. Is the increase in activity highly *sensitive* to the event, meaning that they respond the same way every time? Is it *generalizable* across different types of painful events (heat, cold, chemical, ischemia) and across body sites? Is it *specific* to painful events, or does it respond to other kinds of salient events as well (touch, food, threatening auditory cues)? Is it predictive of the onset, probability, or magnitude of pain behaviors (self-reports in humans or other behaviors in animals)? Is it *necessary*, so that if we suppress its expression in the brain we eliminate pain behavior? Is it *sufficient*, so that if we activate it exogenously, we recreate pain behavior in the absence of a stimulus?

Whether we are measuring populations of single neurons or patterns of activity in human neuroimaging studies, the same principles for understanding representations apply. We can develop measures, but to know how they relate to mental categories and behavior, we have to engage in the series of tests outlined above, and perhaps others. To the degree that a measure is validated in these ways, it can be taken as a provisional proxy for a "representation."

Viewed in this way, it is not clear whether there is one "representation" of pain (or any other state) in the brain, or many. It is also not clear whether there is a representation of pain itself that is particular to the conscious experience of pain rather than other associated processes (avoidance, withdrawal, autonomic responses, learning, memory encoding, updating of goals). In fact, "pain" is a construct, a category that we have invented. Whether or not it is coherent at the brain level is debatable. That is, does "pain" exist as a useful neurological category, or is it a category we invented for human convenience, like the categories "furniture," "heavy metal music," or "actors who played in *Hamilton*"?

But all is not lost: The set of criteria above outline an empirical framework for testing what any particular brain measure actually measures, and we can use that to make progress in understanding which kinds of mental constructs (including "pain," "empathy," "reward," or their subtypes) we can identify coherent brain measures for, and which we cannot. Humans can invent any categories they want to; some, we will find, map onto neurological systems in coherent ways because they are innate, developed early in life, or developed through human experience shaped

by neuroplasticity. Such categories are likely to have predictive and explanatory power in terms of our innate tendencies and predispositions. Biology can also form a basis for agreement among scholars on what the essential constructs are and where the boundaries between them lie. Others do not; these may still be useful for human communication, but they likely will have little predictive power in terms of our innate tendencies and predispositions. And, if we invent psychological categories in a way unmoored to biology, we run the risk of simply inventing our own "convenient truths," with little agreement on constructs, definitions, and boundaries. Our contention here is that we should anchor psychological constructs to biology, and that the set of criteria outlined above provides a way to validate psychological constructs at the brain level.

Over-Simplified Measures

In the grand timeline of the study of mind and brain, neuroimaging arrived very late on the scene, a few decades ago. It grew up in the intellectual soil of neuropsychological research conducted over the previous two centuries. One of the legacies of neuropsychology and early philosophy of mind was that brain processes are implemented in single, discrete chunks of brain (Lindquist & Barrett, 2012). Some dramatic successes identified patients with focal lesions and distinct, circumscribed cognitive deficits—in, e.g., speech production, language comprehension, perception and action (reviewed in Banich, 2004; Brett, Johnsrude, & Owen, 2002). These successes defined the field, and the way in which neuroimaging data were analyzed—one region at a time, with hopes of finding a single, key region that was crucially involved in the behavior.

A "one-region-one-function" ideology licensed several problematic assumptions. First is the assumption that scientists can understand the brain by examining one region at a time. This is the way the vast majority of neuroimaging analyses are still conducted, with "mass-univariate" outcomes that treat each unit of brain (or "voxel") as an outcome in a separate analysis. Maps are collections of effect estimates across voxels. Second is the assumption that if one identifies an area that responds to a particular stimulus or correlates with a particular behavior, it is sufficient to take activity in that region alone as a measure. For example, it has been typical to assume that one can identify regions whose activity represents pain by identifying one or more regions that respond to painful stimuli.

The same assumption plays out across different areas in affective science. Because the amygdala responds to negative affective stimuli, it is widely assumed that the amygdala *represents* negative affect. Amygdala activity has thus been taken as a measure of negative affect and used as a *target* for manipulations of the social context, mental health interventions, and more. The same is true for the nucleus accumbens/ventral striatum and reward; the dorsal cingulate and pain or "conflict"; the ventromedial prefrontal cortex and value, reward, or emotion more generally; and more.

One way to describe the problems with this assumption is to consider the formal inference being made. If we want to understand which brain areas encode the intensity of a painful stimulus, we need to infer on the probability (or effect size) of activity in a local region given an increase in stimulus intensity. This is called a *forward inference*, and it is what statistics in standard brain mapping studies are designed to address. The ability to make forward inferences depends on the measure's *sensitivity* to the event in question. However, if I am interested in whether activating a brain measure implies that a particular type of mental event (e.g., pain) has occurred, this is a different type of inference. It is known as *reverse inference* in the brain mapping literature (Poldrack, 2006) because it involves inference on the stimuli or causal events rather than their effects on the brain. The ability to make reverse inferences depends on more than sensitivity; it depends on the *positive predictive value* of the measure, which depends jointly on *sensitivity*, *specificity*, and the *base rate* of the mental event in question. If a single brain region or network is active during many different behaviors or tasks, inferring mental states based on its activation becomes impossible, because one of many different constructs could be driving changes in activity.

This is precisely the case with brain regions typically used as measures of pain and other affective processes. For example, Fig. 8.1 shows a breakdown of the various types of tasks that activate one of the most "pain-selective" regions of the dorsal cingulate cortex, part of the anterior midcingulate (aMCC). Though the aMCC does contain single neurons that encode the incidence and intensity of noxious events (Hutchison, Davis, Lozano, Tasker, & Dostrovsky, 1999; Koyama, Tanaka, & Mikami, 1998; Sikes & Vogt, 1992), fMRI activity is observed in a wide variety of non-pain-related tasks as well, including cognitive, motor, and language functions. Thus, little can be inferred about the mental processes engaged based on finding activity increases at this location—and if a pain-related intervention influences activity in aMCC, little can be said about which of the processes potentially involved (pain perception, emotion, decision making, etc.) are being represented.

But again, all is not lost. Whether the aMCC is activated is only a small fraction of the information available in a brain image. In addition, we have information about the precise locations within aMCC, the magnitude of activation in each location, and the relative activation across locations. This pattern information can be substantially more sensitive and specific to pain and other mental categories, as we shall see below.

In sum, identifying brain measures for mental constructs is a worthy goal. These measures can tell us a great deal about the physiological architecture that supports the mind, and they can form a useful set of physiological targets for interventions. But problems arise with over-simplified definitions of brain measures and hasty, superficial validation. Showing that the dorsal cingulate responds to painful events, or the amygdala responds to negative images, is only the first in a long series of steps outlined above for understanding what constructs the brain measures represent. The strategy is not wrong, but the development and validation of the brain measures we use as proxies for affective representations is incomplete. In addition, the measures that come from standard hypothesis tests (is region x active in task y) are too coarse to have high positive predictive value for mental constructs.

Hyper-Complexity

The combination of sophisticated machine learning approaches and non-invasive whole-brain imaging has produced brain measures that are increasing in complexity. Rather than basing predictions on the activity of a single brain region, it is possible to develop more complex measures for mental constructs. Currently, many brain measures include on the order of a hundred of thousand parameters to make predictions. And a new class of models based on brain connectivity expands the space of parameters even more dramatically. For example, many studies now develop predictions based on connectivity across pairs of voxels (Dosenbach et al., 2010; Drysdale et al., 2016; Rosenberg et al., 2016; Shen et al., 2017; Turk-Browne, 2013). A standard fMRI scan at spatial resolutions now easily accessible has 62 billion pairs of connections, leading to models with up to 62 billion parameters. And although this may lead to better predictive performance, it comes with its own problems including interpretability and transparency.

The more complex a brain measure, the more difficult it is to interpret and explain how it makes predictions. A simple brain measure based on the activity of single region is easy to interpret: if the brain region is active above a set threshold, then we can make a probabilistic claim about the likelihood of a mental state. There is a single parameter to interpret and model predictions can be explained in a straightforward manner. This model is transparent (Lipton, 2016) because it is easy to contemplate in its entirety, it utilizes a simple (linear) learning algorithm, and each component can be intuitively explained. On the other hand, a complex brain measure that is based on the joint activation of every region of the brain is not so straightforward to understand. Because model parameters are jointly estimated, no single brain area is guaranteed to predict a mental phenomenon on its own. Dependencies across brain regions, estimated by their functional covariation, are also learned by the model in many cases. Often, modeling covariation across regions can help to more accurately predict outcomes of interest (Woo, Schmidt, et al., 2017). This means that the role of individual regions in these complex models can be difficult to infer (Haufe et al., 2014), and inferences are most accurate when the model is examined in its entirety.

The tendency for complex models to outperform simpler ones in many cases has given rise to the popular notion of a tradeoff between prediction and explanation. If prediction is the primary goal, we may not worry much about a model's complexity. Conversely, if explanation is primary, we may care less about predictive accuracy. However, characterizing prediction and explanation as a simple tradeoff is a very limited way of thinking. Typically, we want both. A model that "explains" a phenomenon without being able to predict new instances of that phenomenon is a myth. For example, "the volcano erupted because the volcano god is angry" is one of many explanations, but if it has no predictive validity, there is no substantive reason to prefer it over any other explanation. Conversely, a model that predicts instances without explaining—the proverbial "black box"—is problematic because without knowing why a model makes one prediction vs. another, one never knows

when the model will fail. It might make accurate predictions in one context, but be wildly inaccurate in others. For example, a brain model that predicts ADHD status might be based on head movement (Couvy-Duchesne et al., 2016; Eloyan et al., 2012), which obviously has little explanatory power for the brain basis of ADHD and might have predictive validity only until better motion-correction algorithms are available. Prediction and explanation might be better thought of as the two parents of understanding—sometimes at odds, but more often working together toward a common goal.

Beyond complicating our understanding of how they work, complex models can often be less generalizable. Models that are overly complex tend to characterize idiosyncrasies of the data used to build models to make predictions, rather than core features related to a mental construct. For example, a hypothetical brain measure of a broad construct like negative emotion may rely not on responses related to unpleasantness or negative motivation per se, it could also be driven by responses that are related to affective arousal or the attentional demands of negative stimuli. This does not suggest that such a model is inherently flawed, simply that it uses many correlated features to make predictions about negative emotion. However, such a model is not likely to be broadly generalizable; it may not fare well in less frequent cases of negative emotion that are low in arousal, do not demand attention, or are superficially dissimilar for other reasons.

Thus, models should ideally be as simple and transparent as possible, but still retain good measurement properties. Determining the balance between simplicity and complexity is a well-known problem in machine learning, known as the bias-variance tradeoff: a model can either be more complex to precisely characterize the training data used to build it (a case of low bias or overfitting) or it can be less complex to minimize the variance in model predictions (a case of low variance or underfitting). Increasing complexity decreases bias but increases variance. The goal in model development is to find the proper balance of bias and variance to optimize model complexity. Many helpful approaches from machine learning have been developed to meet this goal (e.g., cross-validation and regularization, among others), and ideally produce more transparent, interpretable models. However, using tools from machine learning alone does not guarantee psychologically interpretable models. Insights from psychology and psychometrics are also crucial when it comes to developing brain predictors of mental constructs.

From Maps to Models of Brain Function

The Difference Between Maps and Models

The aim of brain mapping is to identify which brain regions are consistently activated by different experimental manipulations of mental state. The classical outcome of a brain mapping study is a parametric map that shows how different

experimental conditions are associated with fMRI activity spanning the entire brain: perceiving faces (relative to other objects) produces a map with peaks in the lateral occipital cortex, fusiform gyrus, and amygdala; receiving rewards reveals a map with high activation the ventromedial prefrontal cortex, ventral striatum, and amygdala. Brain maps identify the neural correlates of different manipulations, substantiating forward inferences about the brain regions that will be active during a particular mental state.

Superficially, a brain model may not appear very different from a brain map. This is because they both comprise a set parameter estimates from a regression model (or related statistics) across local areas of the brain. Both reveal patterns of brain activity that are related to some mental phenomenon. However, the purpose of brain maps and models is quite different. Whereas brain maps characterize which brain regions respond to different stimuli or mental events, multivariate brain models are designed to make reverse inferences about (or predict) an individual person's mental state or behavior based on their brain activity (Kragel, Koban, et al., 2018; Woo, Chang, Lindquist, & Wager, 2017).

The utility of brain models lies in their ability to quantify reverse inferences by making objective, testable predictions about mental states based on brain activity. This allows researchers to focus on models that have desirable measurement properties, to falsify models by making strong predictions, to establish the reproducibility of models, and to identify the mental constructs with which a model is most consistent. Brain maps contribute useful knowledge about the function of brain regions, particularly when accumulated across studies and summarized in mega- and meta-analyses (Wager, Lindquist, & Kaplan, 2007; Yarkoni et al., 2011). But brain models are more useful for understanding how the brain constructs mental states and organizes the space of mental phenomena, through their capacity to predict outcomes of interest and characterize brain representations of mental constructs.

Predictive brain models come in many forms (for review, see Kragel, Koban, et al., 2018), but perhaps the most common form is a map of linear associations between voxels in multiple brain regions and an outcome of interest. An outcome of interest can be estimated by computing the dot product of this map, often called a weight map or predictive map (Fig. 8.2), with brain activity measured during scanning. This estimation procedure involves computing the product of the predictive map and measured brain activity at every voxel, and then summing across all voxels. This way, brain activity measured at every voxel in the map contributes to the prediction. Depending on the form of the model, outcomes of interest could be a continuous measure of behavior, a subjective measure of self-report, a probability of a mental state, or a diagnostic outcome such as clinical status.

To build models that achieve the goals described above, models need to be trained with generalizability and falsifiability in mind. Increasing the amount of training data should improve performance and decrease the likelihood of overfitting. Verifying that the model performs well on data that is independent from that used to train it is essential to be certain it does not only predict idiosyncrasies of the training data (cross-validation is one way to estimate the generalizability of a model, see Fig. 8.2). Training on data from different manipulations that are all conceptually

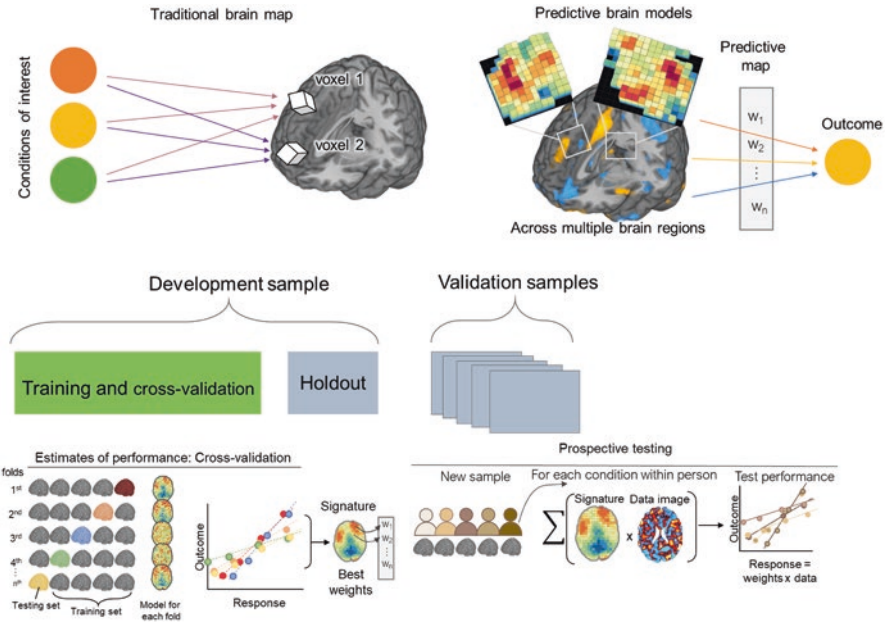

Fig. 8.2 Brain mapping versus brain modeling, and valid tests of model accuracy. The standard brain mapping procedure (top left) identifies brain areas that respond to known (or assumed) psychological events. It provides a map of local effects, but not a model of how these various brain effects work separately or together to construct a psychological state. The predictive modeling approach (top right) reverses the equation, treating psychological states as outcomes. These outcomes are jointly predicted by combinations of brain regions (including connectivity and integrative network properties if desired). The bottom panels show some desirable schemes for testing the predictive accuracy of a model on new individual participants. One of the most efficient ways is *cross-validation* (bottom left), in which a subset of participants is held out as a test set. The model is trained (i.e., its parameters estimated) on the remaining participants (the *training set*), and the trained model's performance is evaluated on the test set. The procedure is repeated with a different test set until every participant has been tested at least once. If done properly, this provides a virtually unbiased estimate of how well a model trained on one group of individuals will generalize to others. However, there are many ways in which the assumptions underlying cross-validation can be violated, resulting in over-optimistic estimates. For this reason, it is desirable to also test the model prospectively on an additional hold-out sample that is tested *only once*. Repeated testing with different models will produce an over-optimistic bias and invalidate the test. Models that hold up to validation in this way can be validated on other samples as well

related to the outcome of interest but that differ from one another in superficial ways (e.g., modeling brain responses to aversive images and sounds) is one approach to increase generalization across contexts. Training models on data from a sample (or multiple samples) of independent subjects has multiple advantages. It greatly increases the amount of data that can be used to train models, as data can be pooled across large and diverse samples. It also allows models to be tested prospectively in new studies, enabling new hypotheses about the specificity and generalizability to be tested.

Three Examples: Models of Pain, Affect, and Emotion

To date, we have developed over 18 predictive models focused on basic affective and emotional processes, all of which were designed to generalize to new individuals (for a partial review, see Kragel, Koban, et al., 2018; Woo, Chang, et al., 2017). These have been tested across independent samples to varying degrees, fostering the process of prospective testing and validation. Although these models are at different stages of development, and some affective processes are easier to manipulate and measure with fMRI, here we focus on several examples that have been successfully evaluated in multiple prospective tests, leading to a better understanding of which cognitive and affective processes they respond to, and which they do not. These models aim to characterize the neural substrates that best describe affective processing related to physical pain, negative affect, and discrete emotional experiences. They show that it is possible to develop brain models that robustly and reproducibly predict individual people's affective experiences.

The Neurologic Pain Signature: A Neural Marker at the Core of Nociception

Perhaps the most extensively validated brain model of a basic affective process is the Neurologic Pain Signature (NPS, Wager et al., 2013). The NPS was designed to predict differences in subjective pain reports based on fMRI activity in areas commonly associated with painful stimulation (Yarkoni et al., 2011), including dorsal cingulate cortex, insula, somatosensory cortex, thalamus, midbrain, and a number of other regions (Fig. 8.3). The model was developed using brain responses to thermal stimulation and was found to discriminate between painful heat and nonpainful warmth, the anticipation of pain, and pain recall with over 94% sensitivity and specificity. In a prospective test, the NPS was found to be sensitive to the subjective intensity of thermal stimulation in an independent study and specific to physical pain, as the NPS responded strongly to painful thermal stimulation, but showed little response to nonpainful stimulation and no response to "social pain" evoked by viewing an image of an ex-romantic partner and recalling an experience that evoked rejection-related distress (Wager et al., 2013). This was surprising in light of other work highlighting the similarity of somatic pain and rejection (Eisenberger, 2015; Keysers, Kaas, & Gazzola, 2010; Kross, Berman, Mischel, Smith, & Wager, 2011). Because rejection has been considered one of the experiences most similar to somatic pain in its brain representation, finding no NPS response to rejection was a particularly important demonstration of specificity. In addition, many researchers have focused on the similarities between pain-related activation and activation related to other salient events (Legrain, Iannetti, Plaghki, & Mouraux, 2011). But rejection-related stimuli and other stimuli that fail to activate the NPS—e.g., observing others' pain (Krishnan et al., 2016) and highly aversive emotional pictures (Chang, Gianaros, Manuck, Krishnan, & Wager, 2015) are highly salient, suggesting that the NPS is not tracking general salience, attentional demand, or arousal.

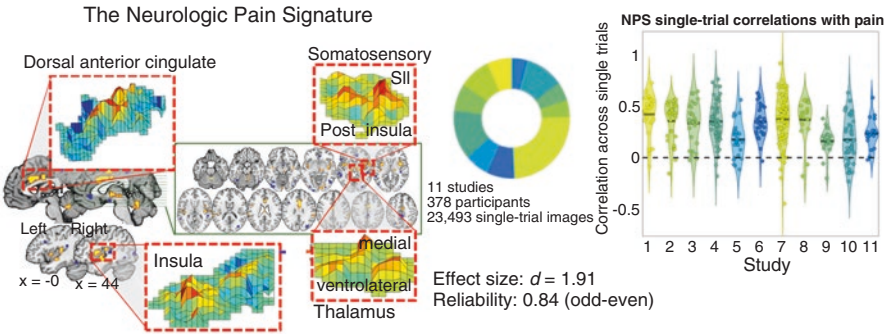

Fig. 8.3 The Neurologic Pain Signature (NPS). The left panel shows the local activity patterns that together comprise the NPS, which was trained on voxels covering about 10% of the brain (Wager et al., 2013). Many regions involved in nociception, like the dorsal cingulate and posterior insula, are included in the NPS. The important thing, however, is that the model comprises specific local patterns in these regions that are difficult to name, often do not respect anatomical boundaries precisely, and specify the relative levels of activity in neighboring voxels. This is very different from a "model" that stipulates that one should find some activity somewhere in the cingulate: It is much more precise. The right panel shows the performance of the NPS in tracking trial-to-trial variations in subjective pain reports, in nearly 400 participants across 11 studies collected in our laboratory. Each data point shows the correlation between pain and brain responses across single trials. Though this is expected to be quite noisy, as single-trial responses are highly variable, the vast majority of participants (over 95%) show a positive correlation between NPS responses and pain reports. The "violins" in the plot show the distribution in these correlation values (*y*-axis) across individual participants for each study. This shows that the NPS's correlation with pain is highly reproducible across individuals and samples, without any model re-fitting or any special techniques. The NPS was also reliable, with an odd-even trial reliability of 0.84 on average across the studies. Though accuracy could increase if acquisition and preprocessing procedures were standardized across studies, good performance across studies shows robustness to some variations in paradigm, acquisition (including field strength), and analysis

Since its development, the NPS has been validated cross-culturally using brain responses to pain in over 34 independent cohorts at the time of writing (e.g., see Zunhammer, Bingel, Wager, & Placebo Imaging Consortium, 2018). Figure 8.3 depicts a recent analysis examining the relationship between trial-by-trial variation in the NPS response and self-reports of pain. Although individual studies vary in terms of stimulation parameters, imaging protocols, sample demographics, and concurrent cognitive and affective demands, the NPS is reliably associated with pain reports in each study, with notably large effect sizes.

The generalizability and specificity of the NPS has been evaluated in an ongoing series of studies (summarized in Kragel, Koban, et al., 2018; Woo, Chang, et al., 2017). The specificity of the NPS has been tested against brain activation during a range of non-painful events, such as viewing aversive images, observing others in pain, performing challenging cognitive tasks, and viewing images of ex-romantic partners. In addition, a number of interventions that influence reported pain have no apparent effects on NPS responses. These include cognitive self-regulation (Woo, Roy, Buhle, & Wager, 2015), most placebo effects (Zunhammer et al., 2018),

reward preceding pain (Becker, Gandhi, Pomares, Wager, & Schweinhardt, 2017), and manipulations of expectations (Geuter, Boll, Eippert, & Büchel, 2017; Woo, Schmidt, et al., 2017) and perceived control (Bräscher, Becker, Hoeppli, & Schweinhardt, 2016; Woo, Schmidt, et al., 2017). These null findings suggest that the NPS is not influenced by "top down" effects in most cases, with the exception of some forms of conditioned (learned) influences on pain (Jepma, Koban, van Doorn, Jones, & Wager, 2018) and some manipulations of social context (López-Solà, Koban, & Wager, 2018; Sola, Koban, Geuter, Coan, & Wager, 2019). This suggests it mediates core nociceptive and affective aspects of pain, rather than evaluative aspects that are known to influence pain reports. Motivated in part by debates about the degree to which the NPS is a marker of pain *specifically* or is also responsive to stimulus salience (Hu & Iannetti, 2016), current efforts focus on evaluating the sensitivity of the NPS to a broader array of painful events, including visceral and mechanical stimulation and its specificity against other potentially iso-salient, aversive stimuli such as unpleasant and even "painful" sounds, and breathlessness, among others.

These findings also illustrate two additional principles related to model validation across studies. First, the NPS tracks pain in some, but not all, contexts. For example, it responds to pain increases caused by turning up the heat, but not by *imagining* more intense heat (Woo et al., 2015) or expecting more intense pain (Zunhammer et al., 2018). Does this falsify the NPS as a pain-related measure? No, we do not think so! No brain measure can ever perfectly measure "pain," or any other subjective experience. Brain measures measure *brain systems*, which are linked to pain and may play a role in creating it. But these brain systems can be ignored; my "pain systems" may be firing like crazy but I may be ignoring them, unconscious, or just stoic and unwilling to report my experience as "pain." Since the publication of the NPS, it has become increasingly clear that it reflects one system (with subsystems) that contributes to pain, but other brain systems are important for capturing other aspects of pain, including the change in negative evaluation that occurs when one imagines that a stimulus is damaging or harmless (Woo et al., 2015).

Second, testing a validated pain-related measure can provide a new window into which interventions are effective in shaping the construction of pain. Most psychological and behavioral interventions (e.g., placebo and cognitive regulation) do not affect the NPS, implying that they affect a *later* stage in pain construction or evaluation, or at least different brain processes that contribute to pain reports. Some interventions influence the NPS (e.g., generating expectations of higher levels of pain, see Jepma et al., 2018), which—because the NPS is very sensitive to painful peripheral input—implies a deeper level of influence on earlier aspects of pain sensation and perception. The magnitude of an intervention's effect on the NPS may turn out to be relatively unrelated to the effects on pain reports, leading to interesting new questions: Is minimizing NPS responses helpful in terms of long-term pain, avoidance, and physiological harms? Or will minimizing self-reported pain always be the sine qua non of pain treatment?

The Picture Induced Negative Emotion Signature: Multiple Brain Systems Engaged in Processing Unpleasant Images

One of the most prominent organizing features of affective states is their valence: whether they are pleasant and positively reinforce behavior or whether they are aversive and act as negative reinforcers. The Picture Induced Negative Emotion Signature (PINES) was developed to characterize the brain systems involved in predicting negative affect generated during picture viewing. The PINES was trained in a sample of 121 individuals, using whole-brain patterns of fMRI responses to scenes and objects to predict subjective ratings of negative affective experience. Cross-validation in this training sample revealed exceptional prediction of negative affect in independent subjects: the root mean squared error was only 1.23 points on a 5-point rating scale, and correlations between observed and predicted ratings were high (r = 0.85, Cohen's d = 3.23). Holdout testing in 61 participants not used for training the model showed similar performance in a completely independent sample, discriminating between negative and neutral images (Fig. 8.4).

Because high levels of negative affect are often associated with salience or arousal, a follow-up test was conducted to examine the specificity of the PINES to another unpleasant experience that is highly arousing: painful thermal stimulation. This test revealed that while the PINES is sensitive to the intensity of negative affect

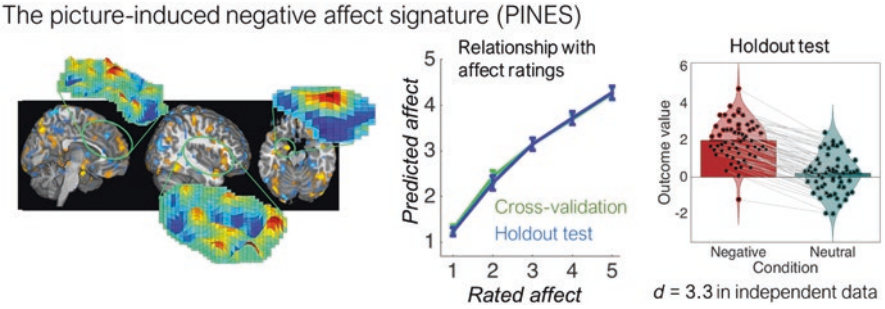

Fig. 8.4 The picture-induced negative affect signature (PINES). This brain model was trained to predict the intensity of reported negative affect on a 1–5 scale, across approximately 180 participants (Chang et al., 2015). The pattern is shown in the left panel; it includes local patterns in the amygdala, insula, dorsomedial prefrontal cortex, hypothalamus, periaqueductal gray, and other regions related to affect and social cognition. The middle panel shows that it tracks ratings across all levels of negative affect in an approximately linear fashion, with nearly identical performance for the cross-validation set and prospective holdout set, which was tested only once. This similarity indicates a lack of bias in the cross-validation test. The right panel shows the performance on the holdout set (tested once) on responses to negative versus neutral images. Each pair of dots connected by a line represents data from one participant. The model correctly predicted which image was the negative one in 100% of the participants, with a massive effect size of d = 3.3. This is an unbiased estimate because the model was tested only once on these data, without re-fitting parameters. Measures with large effects like this one offer substantially more power than single-region effects as targets for interventions (e.g., Gilead et al., 2016; Koban, Kross, Woo, Ruzic, & Wager, 2017; Reddan, Wager, & Schiller, 2018)

evoked by images, it is not sensitive to thermal stimulation, ruling out the possibility that a simple common factor such as arousal adequately describes the mental processes characterized by the PINES. Since its development, efforts are underway to determine if the PINES is better thought of as a general marker of negative emotion or is fine-tuned to a particular set of appraisals, such as evaluations of threat or prospection about negative outcomes. In particular, work has focused on whether the PINES is sensitive to affect evoked by aversive sounds and to positive images to see whether it captures appraisals common to highly salient stimuli.

Relatedly, the PINES has been validated in a prospective test examining whether perspective taking can modulate affective responding to negative images (Gilead et al., 2016). In this study, participants were presented negative and neutral images, and were instructed to either take the perspective of a tough individual who feels little emotion or a more emotionally sensitive and squeamish person who is more prone to responding emotionally. The PINES robustly generalized to this independent sample; brain responses to negative images evoked greater PINES responses compared to neutral images (Cohen's $d = 2.3$). Moreover, PINES responses were diminished by perspective taking. Responses were lower when participants took the perspective of a tough individual compared to the perspective of someone with high levels of emotional sensitivity, although with a smaller effect size (Cohen's $d = 0.37$). This result demonstrates that, unlike the NPS, the PINES is sensitive to cognitive self-regulation—making it a potentially useful target for clinical interventions.

Brain-Based Markers for Emotion Categories: Distributed Representations Identify Qualitatively Distinct Kinds of Emotional Experience

In addition to models that characterize continuous affective dimensions, such as the intensity of pain and negative affect, predictive models have also been developed to identify brain states that distinguish emotional experiences that are rated as being categorically distinct (Kragel & LaBar, 2015). These brain-based models of discrete emotions (Fig. 8.5) were identified by modeling whole-brain patterns of fMRI response to cinematic films and instrumental music that participants rated as evoking distinct feelings of either contentment, amusement, surprise, fear, anger, sadness, or the absence of emotion which was rated as neutral. The decision to include both films and music in the training dataset for these models illustrates a powerful principle: To train models to be maximally generalizable, it is a good idea to include examples of the constructs (here, emotion categories) that are as distinct as possible from one another on superficial features, such as the sensory modality used to elicit emotion. This reduces the chances that the model picks up on confounding characteristics (e.g., different visual properties of movies that evoke feelings of anger or happiness) and increases the chances that its predictions will generalize to new stimulus sets and tasks.

Cross-validation across independent subsamples of subjects revealed that brain responses to single movie and music clips could be classified into one of these seven categories of emotional experience with a medium effect size (Cohen's $d = 0.55$)

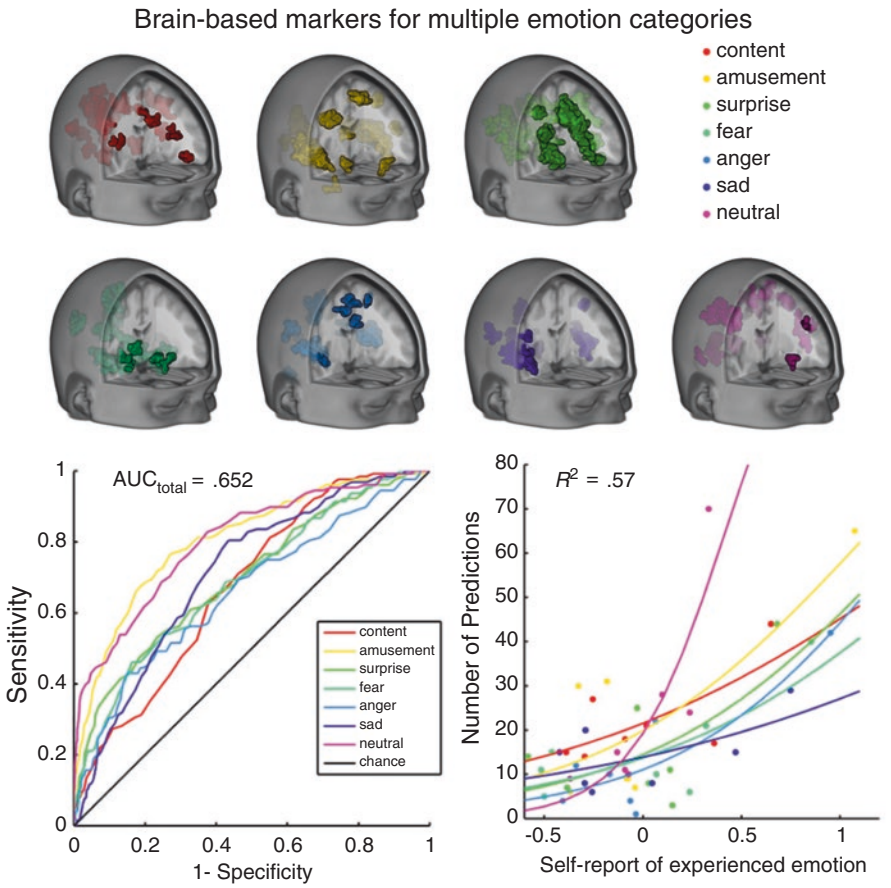

Fig. 8.5 Brain-based markers for multiple emotion categories. These distributed brain models were trained to predict the emotion category labels assigned to both emotional videos and music clips. As with other models, the peak regions are shown, but the models included a broader pattern across the brain. The models were able to classify single-trial responses into one of seven emotion categories with an average accuracy of just over 37%—nearly triple the chance level of performance. As shown in the lower left panel, pairwise classification of single trials revealed effects robustly above chance levels, with an average area under the ROC curve = 0.652. In terms of self-reported experience, the frequency of model classifications explained 57% of the variance in self-report across the seven emotion categories (based on a binomial regression model). Thus, even though the models were trained using a categorical framework, and did not include information about self-report, they are sensitive to differences in emotional experience across a priori categories

and that the models could predict 57% of the variance in self-reported emotional experience.

Given the initial success classifying brain responses to films and music in independent subjects, the generalizability of these models was prospectively tested in the absence of stimulation to see if the brain states identified aspects of emotional experience that were stimulus-related, or if they captured more general aspects of

emotion that are shared across internally and externally generated feelings. This test was conducted by classifying brain activity during resting-state scanning (in a sample of 499 participants in the Duke Neurogenetics Study—Kragel, Knodt, Hariri, & LaBar, 2016) and evaluating the relationship between individual differences in emotional states (anxiety and depression) and traits (anxiety, angry hostility, and depression). Associations were found between participants' state anxiety and model-predicted fear, depressive symptoms and model-predicted sadness, trait anxiety and model-predicted fear, trait angry hostility and model-predicted anger, and trait depression and model-predicted sadness. Although effect sizes were modest (on the order of Cohen's $d = 0.1$), which is common when examining individual differences, these findings validated the emotion markers by showing selective correlations with conceptually related state and trait measures of emotion.

Learning About the Brain, Learning About the Mind

Key questions in affective science have traditionally focused on *either* the mind *or* the brain. Ongoing debates in psychology concern whether affective states and constructs should be characterized as points in a multi-dimensional space, represented as different kinds or categories, or as some combination of the two. These types of debates span multiple areas of affective science, including pain (Davis, Kucyi, & Moayedi, 2015) and emotion (e.g., Barrett, Khan, Dy, & Brooks, 2018; Cowen & Keltner, 2017, 2018). Until recently, evidence from neuroimaging has played a relatively minor role in constraining theories of pain and emotion. Analogously, debates in affective neuroscience tend to focus on mapping different psychological processes onto different neural substrates: does the amygdala selectively process information related to valence, threat, fear, or salience? Is the dorsal cingulate a pain selective region (Lieberman, Burns, Torre, & Eisenberger, 2016; Lieberman & Eisenberger, 2015), or does it integrate multiple different computations involved in valuation and action (Apps, Rushworth, & Chang, 2016; Brown & Alexander, 2017; Kolling et al., 2016; Kvitsiani et al., 2013; Shenhav, Straccia, Cohen, & Botvinick, 2014; Wager et al., 2016)? For the most part, hypotheses about the brain and mind have been separate, making it difficult to use our knowledge of the brain to advance our understanding of the mind, and vice versa. The predictive modeling framework aims to overcome this issue by making links between psychological theory and brain models explicit, with the goal of simultaneously uncovering knowledge about both the brain and mind.

Learning About the Brain: Using Models to Understand Brain Representation

Predictive brain models can be used to answer many different questions about brain representation. One line of questions explores the relationship between brain structure and mental events: Which neural structures are important for a mental

construct? Are certain networks or groups of brain regions more important than others? Is the brain representation of a construct distributed, or is it engendered by local codes? These questions reveal insight into the nature of brain representations, providing a rich way of comparing predictive models of mental phenomena.

These questions can be answered by crafting models using different approaches and comparing the results. To understand which brain regions are important for a mental construct, multiple models can be built using brain activity from different sets of areas. If a single region, or as is more often the case, if a set of brain regions is *sufficient* for predicting an outcome of interest (e.g., the intensity of negative affective experience), then increasing the complexity of the model by including signals from additional brain regions should not improve the accuracy or performance of the model. Conversely, if a brain region is *necessary* for predicting an outcome of interest, then any predictive model that does *not* include it should perform worse than if it had been included in the model.

As an example, consider two predictive models that are both good predictors of an outcome of interest that were trained using two non-overlapping brain regions. Model performance is the same regardless of which brain region is used to build a predictive model. Thus, *either* brain region is sufficient for prediction, but *neither* brain region is necessary. In this scenario, the outcome of interest may be coded similarly in each of these brain regions, because no information is gained by adding signals from both regions to a common model. In this case, the regions could be considered redundant from an information theoretic perspective.

Related to the problem of identifying which brain regions are necessary and sufficient for prediction, brain representations can be characterized either as local or distributed codes. Local representations are spatially restricted to a single brain region (or circumscribed neural circuit). Distributed representations are spatially extended, and contain multiple codes that on their own do not directly reflect the outcome of interest but only do so when considered together. The distinction between local and distributed representations could apply to coding in single neurons vs. populations of neurons in a brain region (Averbeck et al., 2006), or to single brain regions vs. large-scale distributed networks or combinations of networks (Kragel, Koban, et al., 2018).

The representation of objects in inferotemporal cortex is one particularly well-studied example of distributed representation. This brain region contains neurons which code for different high-level visual features, such as color and form (Tanaka, 1996). Individually, these neurons cannot effectively represent an object. However, when considered jointly, populations of these feature-selective neurons can be used to code for many different types of objects. Consider, for instance, populations of neurons that selectively respond to objects that are orange in color, or objects that have curved edges, or are somewhat glossy, or that have a dimpled texture, and so on. Any single one of these features is not sufficient to represent the fruit "orange," but when enough of these features are combined, they can form a distributed representation for "oranges." In addition, a number of studies have shown that the individual neurons that respond most strongly to a given object type (e.g., oranges) are not sufficient to decode object categories—i.e., to discriminate oranges from others

(Kiani, Esteky, Mirpour, & Tanaka, 2007). Distributed population codes also appear to be crucial in other areas as well, from motor control to emotion (reviewed briefly in Kragel, Koban, et al., 2018).

With fMRI, this logic can be extended to the analysis of distributed codes that span the entire brain. A goal in model development is to identify the full set of brain regions that are internally consistent (i.e., that reliably code for a single feature) and which improve performance when added to a predictive model (i.e., is not redundant with other brain features). In this case, each individual feature is necessary for the distributed representation, but no single feature is a sufficient prediction. Characterizing the different aspects of distributed representations can help characterize the nature of complex mental constructs. As an example, there are many different features related to negative affect: valuation of poor outcomes, unpleasant feelings, high levels of arousal, increased attention, motor activation, and so forth. This kind of distributed representation of negative affect would not likely be coded in a single brain region, but would be processed in parallel by multiple systems specialized for different processes. Predictive modeling of negative affect using fMRI provides evidence for such a representation: the PINES, which predicts the intensity of negative affective experience, is composed of multiple subnetworks (including visual, somatosensory, limbic, subcortical, among other brain regions). Although each of these subnetworks independently contributes to predictions of negative affect, no single region is *necessary* or *sufficient* for prediction—providing evidence that brain representations of negative affect are distributed in nature.

Our model predicting negative emotion from brain activity patterns (Chang et al., 2015) exhibited characteristics of a broadly distributed process: No single resting-state network was either necessary or sufficient to predict the intensity of reported negative affect. In addition, a model that combined voxels across multiple large-scale networks was vastly superior to models restricted to any single region (Fig. 8.6). Likewise, models of somatic and vicarious pain constructed from territories spanning multiple brain networks outperform those constrained to single brain regions.

Validating predictive models can additionally be used to show which contexts and variables a brain representation generalizes to or is specific against. For example, the amygdala is an important structure for acquiring conditioned skin conductance responses to tones paired with aversive outcomes. However, is the same amygdala representation also critical for fear-potentiated startle? Is the same representation involved in learning the negative value of certain tones also utilized in learning which tastes should be avoided? Often, we do not know which features are important for affective behavior; we make assumptions, but the boundaries of generalization are usually untested.

Prospectively testing predictive brain models moves beyond assuming brain representations are shared across these factors by making tests explicit. Showing that the NPS responds robustly with the intensity of thermal and mechanical stimulation, but not other emotionally salient events like "feeling" another's pain or viewing aversive images makes it clear that the NPS is not *just* a model of exteroceptive salience, but that it is uniquely predictive of intense sensory events that lead to physical pain.

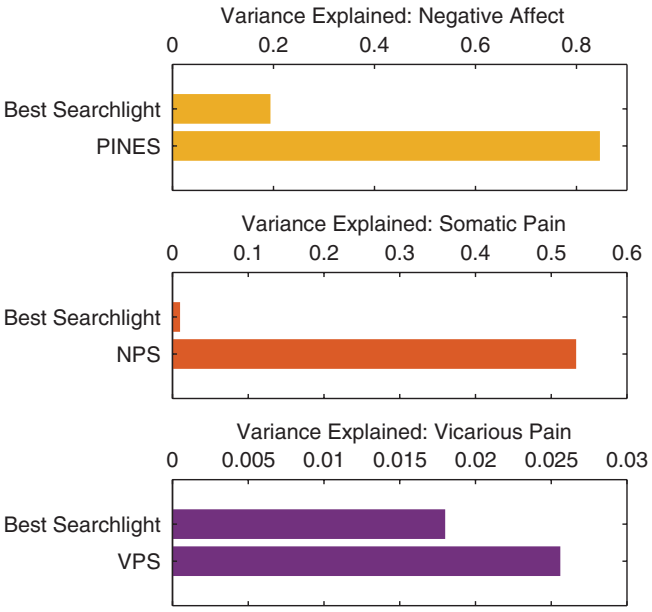

Fig. 8.6 Testing the necessary and sufficient basis for prediction. Comparing predictive models and testing their relative accuracy can inform us about the features of a model that are necessary and sufficient for prediction. One important aspect is the question of how much brain "real-estate" is needed to accurately predict an outcome? Perhaps a single region, like the amygdala, is enough. Perhaps the critical voxels are all contained within one coherent network, like the "default mode" network. Or maybe the situation is more complex and multiple networks are required. Though distributed models appear to produce more accurate results with larger effect sizes in many studies—with outcomes ranging from memory to sustained attention to pain and emotion—the benefits of distributed models are rarely tested systematically. Here, we show two examples of such comparisons, for negative emotion (top), somatic pain (middle), and vicarious pain (bottom). In each case, a model including the whole brain substantially outperformed even the best single regions identified in searchlight analyses across the brain (e.g., the amygdala for emotion, or posterior insula for pain). These analyses show that for both outcomes, negative affect is truly encoded in a distributed network, and no single region is adequate

Just as these tests can tighten the boundaries of a predictive model, by showing specificity, so too can they broaden the limits of presumed generalization. For instance, predictive brain models for distinct kinds of emotional experience (e.g., fear and sadness) respond not only to rich stimuli such as narrative film, but also to individual differences in self-generated feelings in the absence of stimulation.

These examples show how systematically evaluating predictive models—whether testing in an independent subject or a different population, and testing the response of a model to related psychological manipulations—provides insight regarding whether a mental construct has a reliable brain basis, and what the nature of brain representation might look like.

Learning About the Mind: Distinct Systems for Different Types of Affect

Comparing brain representations to one another sheds light on which constructs are more similar to one another, and may be conceptually linked. Often, we make psychological distinctions based on behavior, language, subjective experience, or more generally based on long-held assumptions about how the mind works. Comparing and contrasting models based on the brain can shed new insight into the structure of the mind.

As an example, consider relationships among emotional experiences. Most of the time, correlations among self-reports and judgments of conceptual similarity show that anger and sadness are more similar to each other than to happiness. This is assumed to be the case because anger and sadness are both associated with negative affect whereas happiness is a positive emotion. But our assumptions are often not holding up when validated against human brain activity. Both meta-analyses (Murphy, Nimmo-Smith, & Lawrence, 2003; Wager et al., 2015) and individual studies (Kragel & LaBar, 2015) have shown that sadness and happiness are, relatively speaking, more similar to one another, and that anger is farther away in brain space. This challenges the notion that the emotions are organized primarily based on valence, and that other dimensions of appraisal and affective experience, such as self-relevance and internal orientation (Wager et al., 2015), may be equally if not more important in organizing emotions. Thus, understanding the brain can be used to update current theories about how the mind works by identifying commonalities and differences between mental constructs.

When we begin to compare models that predict various kinds of affective outcomes, a very interesting pattern emerges. The models are largely distinct, suggesting that different affective outcomes are related to different patterns across brain systems. The similarity in the spatial patterns for 18 predictive models developed in our lab is shown in Fig. 8.7. The matrix of intercorrelations shows that the maximal correlation among any pair of models is around $r = 0.2$, suggesting that each model is distinct. There are caveats; comparing cross-prediction of outcomes is a stronger

Fig. 8.7 (continued) $r = 0.2$. This suggests that each brain model is distinct. We must be cautious here, as two brain models can be spatially dissimilar but make the *exact same* predictions (it's true!). Thus, spatial similarity is only part of the story, and cross-prediction and tests of separate modifiability provide a stronger way to evaluate whether two models predict different things. Nonetheless, the picture that emerges from both comparing spatial similarity across models and patterns of separate modifiability within individual studies is that different affective outcomes are predicted by different brain patterns. The dendrogram shows the group of the models; those closest together are most similar. The model groups models for similar outcomes together: neighbors include models of pain (NPS and SIIPS), empathy (care and distress), autonomic responses to stress (heart rate [HR] and galvanic skin response [GSR]), similar emotions (fear and surprise), and fibromyalgia (pain and multisensory responses). Some are less similar than expected: the Vicarious Pain Signature (VPS), trained on pictures and validated on non-visual stimuli, is different from a model of empathic distress trained on auditory narratives and linked with charitable donation

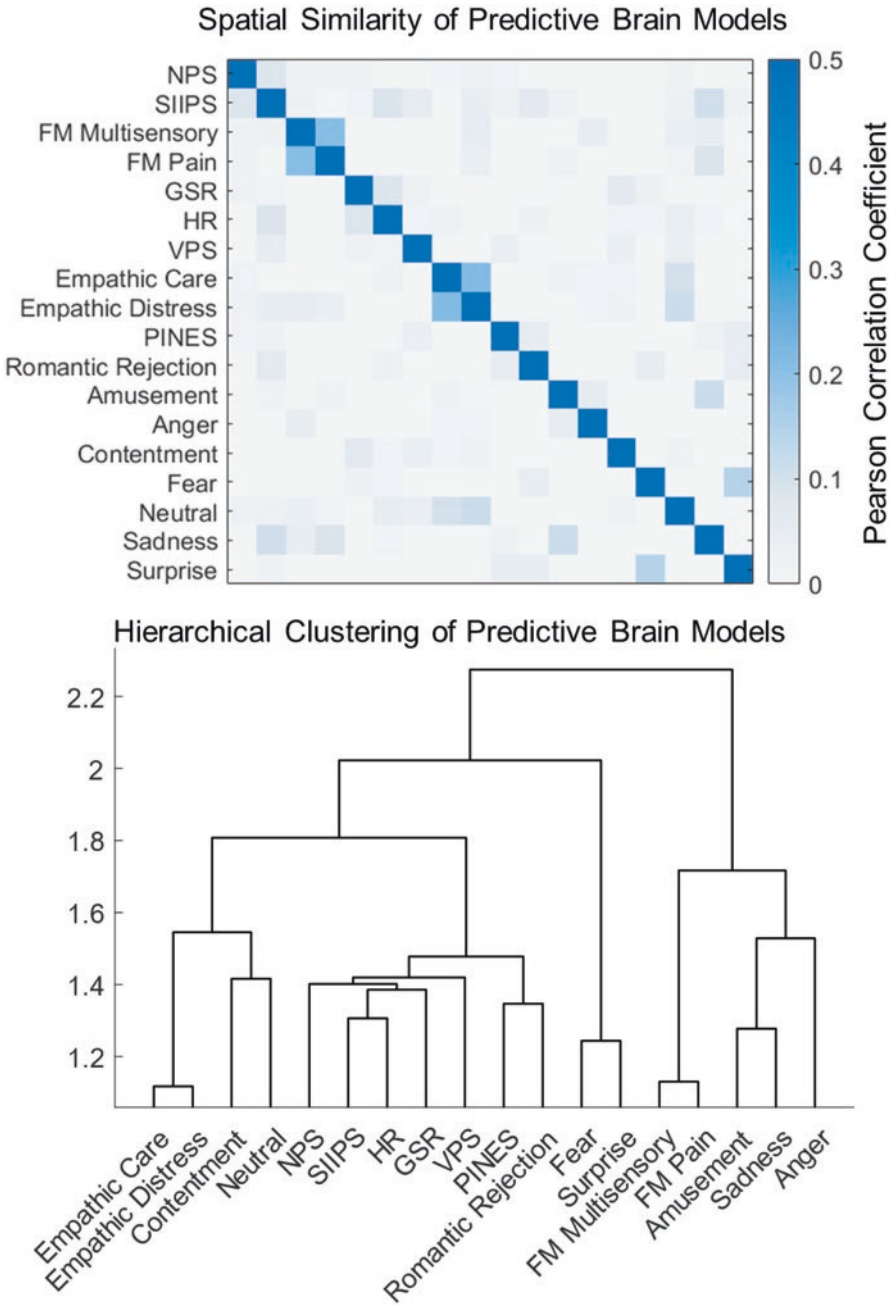

Fig. 8.7 Similarity across brain models of affective processes. The matrix image shows the correlations in the spatial patterns (weights predicting affective outcomes) across 18 models, each designed to predict a specific affective outcome across participants. These models are remarkably dissimilar from one another: The maximum correlation between any pair of models is around

criterion for assessing similarity across models (see Fig. 8.7 legend for discussion), and the low correlations could be due in part to noise. Nonetheless, when paired with the dissociations in outcome prediction we have observed, the results suggest that there is much more differentiation among affective brain processes than we, at least, had previously imagined.

Empathy: A Case Study

The study of empathy provides a useful example of how the modeling approach we describe above can reveal a new picture of how perceiving pain in oneself and others intersect, and why it matters. Empathy can mean different things to different people and has multiple aspects (Klimecki, Leiberg, Ricard, & Singer, 2014; Preston & de Waal, 2002; Zaki & Ochsner, 2012). We view those as existing along a continuum from (a) perception of another's suffering to (b) "feeling with" them, including feelings of personal distress and empathic care (warmth and tenderness), to (c) deciding to take action and give care or support. Here, we'll focus examples on two aspects of the empathy continuum, recognizing another's distress and feelings of care and personal distress.

Recognizing others' distress can involve at least two kinds of processes. One is fast and relatively reflexive and automatic, involving little conscious thought. It tends to produce "experience sharing" or "state matching" (e.g., I am distressed at your distress), which is believed to underlie emotional contagion and some specific forms of helping (and aggressive or fearful) behavior in animals (Preston & de Waal, 2002). At the brain level, representing others' actions, and perhaps emotional states, is associated with "mirror neurons" in the premotor and inferior frontal cortex. Additionally, perceiving others' distress is associated with activation of the anterior insula and cingulate cortex (Lamm, Decety, & Singer, 2011). The second process is slower, more deliberate, and is thought to involve higher-level cognition and, in particular, mentalizing about others' distress. Mentalizing requires the ability to recognize that another's mind is distinct from one's own and conceive a theory about their mental state based on (potentially) multiple context clues. At a brain level, it is thought to rely on cortical networks associated with social cognition, including dorsomedial prefrontal cortex and superior temporal sulcus. It remains largely an open question how important each of these facets of empathy are when it comes to generating feelings for others and helping behavior.

For our purposes here, we'll focus in on one aspect of this broad picture: How similar are brain responses to one's own pain vs. vicarious "pain" from observing others? The idea that they activate the same brain systems, particularly the anterior insula and cingulate (Singer et al., 2004), has been used to argue that state-matching mechanisms are important for empathic care and helping responses in humans. The implications of the answers to this question go beyond understanding how empathy works; as David Brooks wrote in his article "The Archipelago of Pain" (Brooks, 2014), if emotional pain is the same as somatic pain at a brain level, why treat them

differently in our legal system and policymaking endeavors? Brooks argued that if social isolation is like being physically tortured, maybe it should be just as illegal. And if observing someone else in pain activates our pain circuits, does that qualify it as a type of harm as well?

Studies comparing self-pain with other-pain (or vicarious pain) reveal very different patterns of brain overlap depending on whether one is looking at overlap in univariate brain responses to stimuli or comparing patterns that decode experiences of distress. Most early studies have found overlap in the anterior insula and cingulate, but there are a couple of problems with concluding that this reflects a "shared mechanism." First, these studies generally do not establish a strong link between activity and vicarious pain, and they do not establish that the responses are specific to "painful" events rather than other classes of emotional, cognitive, language, decision, and motor processes. Thus, if there is overlapping activation when experiencing heat pain and looking at pictures of others in pain, what does this overlap mean in terms of which processes are shared? Any set of mental processes common to self- and other-pain and reduced in the control tasks (generally non-painful warmth and neutral pictures) could be driving shared activation. This includes greater attention, greater salience or relevance, stronger autonomic responses, and greater demand on action planning mechanisms. To infer that what is shared is specifically related to empathy involves ruling out these and other alternatives.

A second problem is that these studies typically focused on the overlapping areas and assume all the differences are essentially due to noise. For example, if self-pain and other-pain overlap in 5% of the voxels tested, does this mean that a "shared mechanism" has been found? It might be prudent to consider the differences as well—perhaps it means that self- and other-pain are only 5% similar. But this conclusion would be premature as well, for at least three reasons. First, overlapping voxels are a poor way to assess similarity in mental processes because the number of activated voxels is not a *measure* of any particular process, as described above, and may not be related to empathy at all. Secondly, it ignores the magnitude of the responses. What if the voxels in common are activated twice as strongly by one condition than another? Should this still be counted as an identical response for purposes of assessing overlap? Third, the similarity metric will be strongly influenced by measurement noise as well as the similarity in the underlying processes.

The first of these is the most conceptually profound, and points to some fundamental uncertainties in how we should use brain similarity to infer similarity in mental processes. Even if we quantify the degree of overlapping vs. non-overlapping voxels, we need to understand the relationship between activation patterns and the behavior we are interested in (e.g., vicarious pain). This is a conceptual problem, not a problem with measurement noise. An illustrative example comes from a recent study by Carrillo et al. (2019). They quantified the proportions of cells in rodent dorsal anterior cingulate cortex (dACC) that respond to (1) painful shocks, (2) observation of another rodent receiving painful shocks, and (3) a threatening sound conditioned to painful shock. Some neurons responded to each of the three motivationally relevant conditions, and subsets responded selectively to pairs of conditions or to all three. They interpreted those neurons that responded to self-pain and other-

pain as "empathy selective." This is appropriate—but does it mean that the dACC contains a "mechanism" for empathy? Maybe, but we have to make an additional assumption: We must assume that shared neural activation implies a shared process. However, this need not be the case. A wealth of population-coding evidence in neuroscience indicates that in many domains, the *pattern* across neural population carries information about stimulus types, mental categories, and motor responses—not the individual neurons. If we accept the "empathy mechanism" account of neuronal overlap, we must also posit that there is a "pain and emotion but not empathy" process also represented in the cingulate, and a "empathy and emotion but not pain." A more parsimonious alternative is that the cingulate contains *distributed representations* related to each of the three types of events. These may share some neurons in common, but involve distinct or even completely independent neural patterns. In this case, a natural similarity metric would be the spatial similarity across the populations of neurons—do they share 5%, 50%, or 90% of their neurons in common? This can be captured by calculating the spatial correlation across neural patterns, perhaps considering the continuous intensity of the neural responses to each type as well as whether or not they pass a statistical threshold and are thus considered "responsive."

We must also recognize that linear spatial similarity may be insufficient and other metrics apply—or even that the overlap in individual units has little bearing on the population-level representation. As an analogy, consider three words: GRASS, FLOWERS, and BAGS. Imagine that each word represents a mental process, or construct, and each letter a neuron that fires in response to that process. Some neurons ("S") respond to all three, and some ("A," "R," "G") to only two. If we summarize these overlaps, we will find that there are some "grass-flower" units ("R"), which we might capture a mental process common to the two conditions (R encodes "living plants"). If we calculate the similarity across units, we might infer that GRASS and FLOWER share slight overlap, but GRASS and BAGS are very similar. In fact, they are! They are *orthographically* similar, but they are not semantically similar. The relationships between letters and conceptual meaning are not linear, and one cannot construct a function of the similarity in letters and come up with an answer for the similarity in meaning. This example shows us that counting overlapping neural populations, or even assessing spatial neural similarity, may not always give us the right answers when it comes to inferring similarity in mental processes.

We are not particularly nihilistic about the situation, and there are solutions. Multivariate pattern analysis provides a complementary way of looking at neural populations—whether fMRI voxels or individual neurons—that partially solves the problems raised above. First, multivariate decoding provides a set of predictive models that can quantify how much variance in an outcome (e.g., reported vicarious pain experience) is captured by the model. If the predictive validity is high, then we can be more certain that we are studying brain measures *related to empathy*. In addition, if the predictive models can be tested across studies, some alternative processes can be ruled out—e.g., if other tasks that enhance attention do not increase responses in the model, enhanced attention can be ruled out as an explanation for

what the brain model is measuring. Second, we can provide an unbiased estimate of brain similarity, in two ways. We can assess the spatial similarity across two multivariate models, as described above. Or we can assess cross-prediction: How much variance in one outcome (e.g., somatic pain experience) is predicted by a model of a comparator outcome (e.g., vicarious pain experience), and vice versa. Third, cross-prediction provides an unbiased estimate of similarity, controlling for effects of measurement noise. For example, if Brain Model 1 explains 25% of the variance in somatic pain ratings and 25% of the variance in vicarious pain ratings, we might infer that it reflects both somatic and vicarious pain. If it explains 25% of somatic pain ratings but only 5% of vicarious pain ratings, we might infer that the effects are five times stronger for somatic pain. And if there is more noise or the outcomes are unreliable, we can still assess the relative predictive power—e.g., 5% for somatic but only 1% for somatic pain. Finally, assessing cross-prediction avoids some of the problems with the mapping between neural units and conceptual categories discussed above, because it assesses the pattern as a whole, and whether that pattern is specific for one condition or general across both. We need not assume that the units of the model (voxels or neurons) are individually interpretable in relation to the mental category that the pattern reflects.

In a series of studies comparing pain and empathy, we tested whether self- and other-pain activate similar brain representations. We identified whether multivariate brain patterns that predict pain and vicarious pain are similar or different, and analyzed both spatial pattern similarity and cross-prediction to compare the models that predicted each. The design of the first study (Krishnan et al., 2016) involved presenting a randomized series of trials of self- and other-pain. In one fMRI session, participants experienced three levels of somatic pain (heat at 44, 45, and 46 °C), selected to reliably sample the range from nonpainful/barely painful to moderately painful (Green, 2004), on two body sites: the upper forearm and foot. In a second session, participants viewed pictures of painful events on others' hands or feet (e.g., a toe being caught in a door), which were selected in pilot testing to span three levels of vicarious pain approximately matched to the heat in subjective intensity. We selected these images because they have been used extensively in past work (Jackson, Brunet, Meltzoff, & Decety, 2006), and have been shown to activate areas in the dorsal cingulate and anterior insula that putatively encode the affective dimension of pain. Participants were also instructed to take the perspective of the experiencer and imagine the painful stimulation was happening to them; this has been found to increase dorsal cingulate and anterior insula activation as well (Jackson, Meltzoff, & Decety, 2006). Before each stimulus, participants saw a cue instructing them to get ready for the upcoming trial, which allowed us to identify anticipatory activity and compare it to activity during stimulus viewing. After each trial, following a time-varying delay that allowed us to separate stimulus-related from rating-related brain activity, participants rated the subjective intensity of the experience. Though we selected stimuli at three levels of intensity for each of the arm and foot stimuli in each of the self- and other-pain modalities, our analyses focused on predicting variation in trial-by-trial intensity ratings in each modality.

Four other design features are particularly important for our ability to compare self-pain and other-pain brain representations. *First*, we developed models predicting within-person variation in reported experience. This focuses on brain patterns that are *related to experience*, and furthermore is much less noisy and subject to fewer confounds than predicting individual differences in reported experience. *Second*, we developed models that can make predictions about previously unobserved individual participants, or "population-level" models. This allows the model to be tested for specificity and generalization across different task variants by testing the model on new samples. This validation tells us much more about what each pattern measures than any single study is likely to be able to do. We compared the population-level models to idiographic models, in which the brain patterns predicting self-pain and other-pain are customized for each individual; the performance of these models was only marginally better than the population-level model, indicating that the brain patterns that predict experience are stable across individuals. *Third*, before the main fMRI study, we conducted pilot studies to select self-pain and other-pain stimuli that are matched in subjective intensity, eliminating a potential confound. In addition, subjective intensity is controlled for in the analysis, in the sense that we are looking for similarities and differences in brain measures that predict subjective intensity. And *fourth*, we built in two design features that allow us to test the representation of subjective intensity: (1) selecting stimuli that spanned the range of low, medium, and high subjective intensity in each of the self-pain and other-pain modalities; and (2) testing two body sites (upper and lower limb) in each of the self-pain and other-pain modalities, allowing us to analyze somatotopy and compare it to the established somatotopic organization of pain-related areas.

So what did we learn about shared brain representations for self- and other-pain? When we simply analyzed high-intensity stimuli vs. rest, we observed strong overlapping activation in the dorsal cingulate and anterior insula, among other regions, replicating the pattern found in previous studies. But when we compared the models trained to predict experience, the brain representations for self-pain and other-pain were distinct, involving many different areas across the brain and different local patterns within the dACC, insula, and other regions. We can see the same qualitative pattern across a series of analyses, each providing a slightly different window into shared representation. We'll walk through the main analyses here.

A first way to look at shared representation is to compare whole-brain, population-level models. We have found that because such models capture patterns of activity within local regions and across large-scale systems, they often more accurately predict affect ratings than any single local region or individual "network" (see above for an analysis and examples). Testing the NPS, which was previously validated to track pain across multiple studies, we found that NPS responses strongly tracked stimulus categories and predicted pain ratings across both upper- and lower-limb body sites. This makes sense, because while some areas are somatotopically organized (particularly somatosensory S1, S2, and dorsal posterior insula), painful stimuli activate broad, bilateral patterns that overlap across body sites, and many individual nociceptive neurons have broad receptive fields that span body sites. However, though the NPS tracked somatic pain intensity strongly, it showed no

response to images of others in pain (Fig. 8.8). In fact, NPS responses were significantly below zero. Further analysis revealed that this is because the NPS includes negative weights in some regions activated by pictures, e.g., visual cortex, and that focusing only on areas with positive pain-predicting weights (e.g., dACC, insula, S2, posterior insula, which are also nociceptive targets) showed no significant activation or deactivation for observed pain. Conversely, a population-level model trained to predict vicarious pain intensity (Fig. 8.8, right) showed very strong out-of-sample prediction of vicarious pain in new individuals, but showed no response to painful somatic events.

This pattern of cross-prediction results, in which one brain pattern tracks one and only one effect (self- or other-pain), is called "separate modifiability" (*34*). This pattern provides strong evidence for the separability of the brain processes underlying each type of "pain." In particular, it helps to rule out the presence of potential shared, confounding processes like enhanced salience or attention. Imagine that each pattern was driven by a common process, which we'll call "salience." Salience

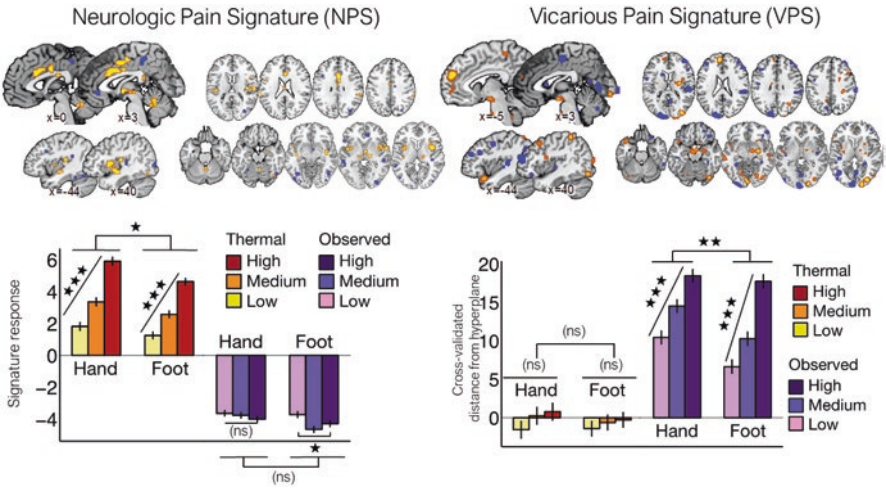

Fig. 8.8 Vicarious vs. experienced pain: Separate modifiability. This figure shows data from Krishnan et al. (2016), who tested three levels of each of somatic (heat) pain and vicarious pain (images of others in pain) in each of two body sites (upper and lower limb). The left panel shows the NPS (top), and brain responses in the NPS to somatic and vicarious pain stimuli, in warm and cool colors, respectively. The NPS responded only to somatic pain stimuli, across both body sites, with a magnitude proportional to stimulus intensity (and reported pain intensity). The right panel shows the "Vicarious Pain Signature" (VPS), a model trained to predict the intensity of vicarious pain ratings across both body sites. It responded in a graded manner to vicarious pain, but showed no response to somatic pain. This pattern of results, termed *separate modifiability*, indicates that neither pattern is strongly driven by shared psychological processes common to both conditions, including salience, arousal, and demand on attention. The NPS shows separate modifiability with other patterns for romantic rejection (Wager et al., 2013) and negative emotion (Chang et al., 2015) as well, suggesting that these various types of salient, arousing events have distinct brain bases. $***P < 0.001$, $**P < 0.01$, $*P < 0.05$, error bars reflect within-participant standard error of the mean

is enhanced for high-pain vs. low-pain stimuli in each modality; is this what our models are capturing? If so, we should see at least some cross-prediction from self- to other-pain and vice versa. That is, a model trained on self-pain that actually captures salience should respond to the higher-salience vicarious pain stimuli more strongly than the lower-salience ones. But this is not what we observed. Therefore, these data suggest that the two brain patterns are not representing any common process that is shared by painful heat and observation of others' pain.

The two patterns also involve different regions, and different patterns within regions. By taking bootstrap samples and re-running the predictive model many (e.g., 5000) times, we can obtain *P*-values for how reproducible each voxel's contribution to the overall prediction is across participants. This allows us to interpret the statistically significant areas for both somatic and vicarious pain, which are shown in Fig. 8.8 at $q < 0.05$ false discovery rate corrected. The somatic pain signature (NPS) most strongly involved many areas that receive nociceptive information from the body. The vicarious pain signature (VPS) most strongly involved some areas related to mentalizing and social cognition, including the dorsomedial prefrontal cortex, and other areas less important for pain prediction here, including the amygdala. In addition, the global patterns and local patterns within dACC and other regions were uncorrelated. We can perform a more systematic analysis of the large-scale differences in the networks involved by comparing each predictive pattern (across all voxels, not only the significant ones) to identified resting-state cortical networks. This is a useful technique for helping to interpret the brain patterns. As Fig. 8.9 shows, the somatic pain-related NPS showed a concentration of positive weights in the "ventral attention" and "somatomotor" networks, and negative weights in the "dorsal attention," "visual," and "default mode" networks as defined by Yeo et al. (2011). The vicarious pain-related VPS displayed a very different pattern, including positive weights in the "default mode" and (to a lesser degree) "ventral attention" networks and negative weights elsewhere, including in the "somatomotor" network associated with somatic pain.

Of course, pain and images involve different sensory modalities and cognitive processes, and so it should not be surprising that their brain patterns are differentiable. But the claim these data support is stronger than that: They suggest that the patterns that *predict experience* are not shared, even within the regions that are thought to encode common representations related to pain affect. A series of additional analyses provided converging evidence for this basic conclusion (Krishnan et al., 2016):

1. A "searchlight" analysis of local regions revealed that though some local regions predicted ratings in each modality, *no brain regions* showed substantial evidence for cross-prediction, and cross-prediction results were much weaker than training within-modality. These tests are fair and unbiased because whether a model is trained on the same modality as the test data (e.g., somatic pain model predicting somatic pain test data) or a different modality, the test data are taken from new individuals not used in model training.
2. Quantifying the effect sizes shows successful prediction within-modality, but weak cross-prediction. Training on vicarious pain explains *9 times less* variance

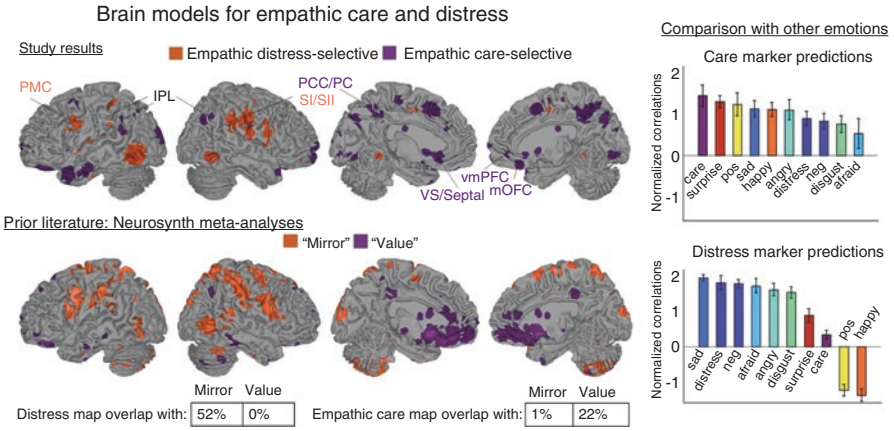

Fig. 8.9 Different types of affect, different brain networks. Interpreting multivariate models in terms of identified systems is a challenge. One lens through which to view them is their loading on (spatial similarity to) large-scale resting-state networks. Here, we show the similarity of the NPS (left) and VPS (right) to each of seven resting-state networks identified by Yeo et al. (2011). The inner dark circle on the polar plots marks the zero-correlation point, so that points outside it show positive correlations between the brain model and network, and points inside it show negative correlations. These plots reveal quite different patterns across large-scale networks for the NPS and VPS, with similar positive loadings on the "ventral attention" network and negative loadings on the "visual" network for both, but very different (usually opposite) correlations with each of the other networks. These networks do not fully capture the models, and similar correlations with the "ventral attention" network does not imply that the two models activate the same locations or patterns within this network—but the pattern of differences across networks illustrates that the two models are different in their macroscopic as well as their mesoscopic (local pattern) organization

in somatic pain and training on somatic pain explains about *500 times less* variance in vicarious pain.

3. Training a vicarious pain model without the visual cortex resulted in prediction that was just as accurate as the whole-brain model, and is not driven simply by visual activation or attention.
4. Re-training both pain-predictive and vicarious pain-predictive models within this study resulted in the same pattern of separate modifiability.
5. Analyzing the time-courses of NPS and VPS responses before, during, and after stimulation revealed that the responses were specific to the stimulus period in both models, and did not respond to either anticipation or post-trial response selection and reporting periods.
6. Training models with patterns customized for each person revealed the same pattern of separate modifiability. Such models are more susceptible to confounds, as they are much more flexible (different patterns for different individuals) and can pick up on different types of artifacts and confounds for different individuals (Todd, Nystrom, & Cohen, 2013). Here, customizing the models yielded little benefit in predictive accuracy.
7. Somatotopy models trained to predict whether stimulation was on the upper or lower limb in each modality showed strong somatotopy for somatic pain in

sensory cortex and posterior insula, with hand and foot regions corresponding to those found in previous studies. These patterns could predict whether stimulation occurred on the upper or lower limb for an individual (averaging over same-site trials) with 90% accuracy. But no such somatotopy was observed in the vicarious pain condition, indicating that self-pain nociceptive pathways are not activated by observing others' pain. Surprisingly, we also observed differentiable brain patterns for observed pain on hands and feet, which could be predicted with close to 90% accuracy from brain data—but these patterns did not transfer to somatic pain. Thus, somatotopic representations for self- and other-pain are also qualitatively distinct.

As may be evident by now, meaningful tests of shared representation across two types of mental events is possible, but it requires a number of analyses from different angles. Most importantly, the tests are only as good as the models: To compare brain representations of somatic and vicarious pain, one must develop models that are validated to track each type of pain, and ideally rule out other kinds of processes that the models might capture. In any domain, this will likely require prospective tests of pre-identified, population-level models like the NPS and VPS across multiple studies. Fortunately, this also seems possible.

Part of our process of testing the NPS and VPS was to test their performance, and in particular their separate modifiability, across studies. For example, is the NPS really an adequate model of pain? It could be that vicarious pain is like a different kind of pain, not heat, with brain patterns similar to that kind of pain. In Studies 2 and 3 of Krishnan et al. (2016), we tested the NPS on mechanical and electrical pain, respectively, and found robust responses. Subsequently, it has been generalized to other types of pain as well, as described above. Study 3 included both painful shocks and pictures of others in pain, allowing us to conduct a prospective test of whether the NPS and the VPS trained in Study 1 show separate modifiability in a new sample. They did.

Another type of test involves testing whether the VPS really captures "vicarious pain" in general, or whether it is capturing something related to the particular images we used or emotionally intense images in general. We tested this in a subsequent study, which also replicated the separate modifiability pattern for somatic and vicarious pain (López-Solà, Koban, Krishnan, & Wager, 2017). In this study, heterosexual women arrived for the fMRI session with their male romantic partners. The partners sat in the scanner room and a thermode was attached to their arm; the women viewed the male partner with pain-induction device attached through the scanner mirror. In the somatic condition, women experienced painful heat during fMRI. In the vicarious pain condition, a small change in the fixation cross indicated that the partner was receiving pain, and there were no other sensory cues. Even in this conceptually driven "cued empathy" situation, the VPS responded strongly and specifically to vicarious pain, and the NPS responded strongly and specifically to self-pain. The separate modifiability criterion held, supporting the independence of the brain processes involved and the generalizability of the models to new samples and task variants.

The space of "prosocial emotions" like vicarious pain is still relatively unknown territory, especially when it comes to their brain bases. Recognizing others' pain could just as easily lead to schadenfreude (joy in others' suffering) and motivation to harm others as to empathic distress and helping. In some cases, this decision may be instinctual, but often—and particularly in humans—the decision to help others requires a deliberate choice (Schumann, Zaki, & Dweck, 2014). In addition to mentalizing about others' suffering, there must be an act of affiliation, a recognition of the suffering other as worthy of help and comfort. Empathic distress is not the only, or perhaps even the primary, emotion that motivates helping behavior. It has an ambiguous and context-dependent relationship with helping motivation, as distress can lead to disengagement and burnout. In some caregiving professions and spiritual traditions, practitioners are taught to avoid getting "lost" in personal distress. Feelings of warmth and tenderness—or empathic care—may be more consistently related to helping. They may also be more sustainable, as they may be rewarding for those who experience them. Empathic care is intertwined with affiliation, a sense that another is close to or aligned with oneself.

In our work, we have developed models predicting how much people will donate their experimental earnings to charity. In one study, we created biographies of potential aid recipients—pictures and stories—that varied along a number of dimensions. The pictures varied on whether the recipient was old or young, black or white, male or female. The stories varied on whether the recipient was prosocial (e.g., helping others, volunteering), whether they were more or less responsible for their hardship (e.g., contracting AIDS because of a childhood blood transfusion or by injecting illegal drugs), whether monetary aid would have an instrumental value (be likely to help improve their condition), along with clues about the political and social identity of the recipient. By creating a "grammar" of statements that can be recombined in many ways, we created hundreds of unique stories, allowing us to investigate which variables predicted higher donation amounts to a target. One recipient was selected at random and the decision enacted; the money participants gave was, in fact, given to charity.

The results of this study indicated that giving was predicted by a combination of emotions and social cognitive judgments and attributions. Both personal distress and empathic care led to more giving, as did judgments that the person was not responsible for their hardship and that giving would have instrumental value. Perceived similarity—whether external (race, gender, age, socioeconomic status) or internal (values and attitudes)—did not predict donation amounts. A quantitative model combining feelings and judgments correlated over $r = 0.6$ with within-person variations in donations.

In a subsequent fMRI study, we asked whether empathic care and distress could be predicted by distinct multivariate brain models. Participants listened to 30-s audio stories of real individuals taken from charity websites, then subsequently received a reminder about each story and were asked to donate up to $100 (100%) of their experimental earnings to a charity that would help individuals in similar situations. After the fMRI session, participants made second-by-second ratings of empathic care or personal distress, which were averaged across participants to

create normative time-courses for each biography. We trained multivariate models to predict the time-courses of each prosocial emotion, testing the models for sensitivity and specificity to each emotion in held-out participants (i.e., using cross-validation). The results identified a pattern that robustly predicted ratings of both empathic care and distress.

The care model, which was selective for care (not distress), involved increased activity in parts of ventromedial prefrontal cortex (vmPFC) and ventral striatum (VS), posterior cingulate, temporal-parietal junction, and anterior temporal cortex (Fig. 8.10). Among these, the vmPFC and VS are particularly associated with appetitive value and reward, including vicarious reward to others (Zaki, Schirmer, & Mitchell, 2011) and prediction of purchasing decisions (Grosenick, Greer, & Knutson, 2008; Knutson & Genevsky, 2018) and donations to charity (Genevsky & Knutson, 2015; Genevsky, Yoon, & Knutson, 2017; Hare, Camerer, Knoepfle, & Rangel, 2010). The vmPFC is particularly associated with self-referential thoughts (Denny, Kober, Wager, & Ochsner, 2012), perceived closeness (Krienen, Tu, & Buckner, 2010; Tamir & Mitchell, 2010), and compassion (Klimecki et al., 2014), and increases after compassion meditation training (Engen & Singer, 2015). And the portion of the TPJ and posterior cingulate activated are particularly associated with social cognition and recognizing others' actions and intentions (Carter & Huettel, 2013; Miele, Wager, Mitchell, & Metcalfe, 2011). Thus, the brain model that predicts empathic care brings together elements of self- and other-oriented cognition and positive valuation.

The personal distress model strongly tracked distress but was not highly selective, as it also tracked empathic care. This model involved premotor regions associated with "mirror neurons" and recognition and imitation of others' actions (Keysers, 2009; Losin, Iacoboni, Martin, Cross, & Dapretto, 2012; Losin et al., 2015). To test whether these two models were reliably different and help interpret them, we compared each model to two meta-analytic patterns derived from neurosynth.org, an online meta-analysis tool that links reported coordinates from thousands of neuroimaging studies to terms and topics used in the papers (Yarkoni et al., 2011). The empathic care pattern was very similar to the meta-analytic map for "value," as indicated by a Dice coefficient—a measure of overlap across binary maps—of 22%, but not to the meta-analytic map for "mirror" (1%). The empathic distress pattern showed the opposite, overlapping substantially with the meta-analytic map for "mirror" (52%) but not with "value" (0%).

Another open question concerns which emotions these models track most strongly. Does the empathic care model track something different than "happiness" or "positive affect" in general? An advantage of training normative, population-level models is that they can be annotated with additional data. In this case, we collected second-by-second ratings of each biography from 200 additional participants in an online study. The time courses of each brain model were then correlated against ratings of each of ten different emotions (Fig. 8.10). The care model correlated with empathic care and a blend of other emotions including surprise and positivity, suggesting that it is a relatively unique experience that is not reducible to generic positive affect or emotional salience. The distress model correlated with a

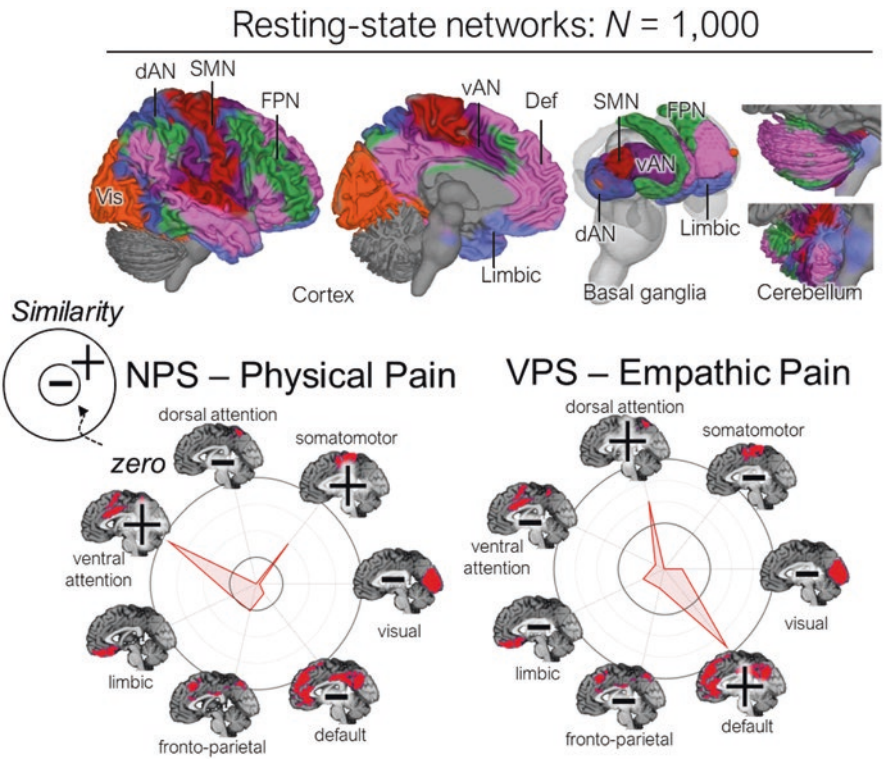

Fig. 8.10 Brain models for empathic care and distress. In this study, participants listened to audio narratives of other hardships. Two models were trained to track normative (group) moment-by-moment ratings of empathic care (warmth and tenderness) and distress. The empathic care-selective model (purple, top) included high weights in regions associated with reward, value, and self-relatedness. We compared this with a neurosynth-generated meta-analytic map for "value" (purple, bottom), and identified strong overlap in the voxels included. The empathic distress-selective model (orange, top) included premotor and inferior frontal regions associated with self-other action mirroring and negative affect in other models (e.g., for romantic rejection). We compared this with neurosynth's "mirror" map (orange, bottom) and identified substantial overlap. These findings suggest that these two empathy-related emotions have distinct brain bases. The right panel shows how these models can be further explored and validated in new studies. Because they were trained on group rating data, we were able to conduct an online study in which participants rated the same narratives on other emotions, shown at right. These time courses were regressed on the fMRI time courses for the two models. The empathic care model correlated most strongly with reported "care" in the online sample, and less strongly with other emotions. The empathic distress model correlated positively with a range of negative emotions, and negatively with ratings of "positive" and "happy." Thus, the brain systems underlying empathic care in particular may be relatively unique to care as opposed to other positive emotions

range of negative emotion ratings approximately equally strongly, but less strongly with positive emotions, suggesting that it tracks a relatively undifferentiated form of negative affect. This illustrates how population-level brain models can be retroactively tested by correlating them with measures collected on the same stimuli in subsequent experiments.

This work leaves a number of questions to be addressed in future studies: What is the relationship between vicarious pain and empathic distress across stimulus types? The VPS and empathic care patterns appear to be qualitatively different; is this a function of the types of stimuli used, the social judgments made, or participants' intentions in the context of the different types of studies? And what is the relationship with helping behavior? Stronger activity in both models predicted donations, but only weakly, suggesting a variable relationship between the brain systems underlying these feelings and action. Our view is that helping decisions are complex, and depend only partly on empathic feelings and attributions—they are also influenced by cognitive policies about how much to give, relative comparisons about how much one has given previously and whether one has "done enough," personal thoughts about the value of keeping the money for oneself, and likely other factors.

Conclusions and Implications

Given the complexities and variability involved in human affect and decision-making, it is remarkable that it is possible to identify consistent brain correlates of multiple types of affect across individuals. This appears to be true for complex social emotions like empathic care as well as basic, evolutionarily conserved processes such as pain. This did not have to be the case; the construction of pain, feelings of rejection, vicarious pain, anger, sadness, and other emotions could have been very different across different individuals, making it impossible to identify stable brain predictors. In a famous example from philosophy, one can never be sure whether my experience of the color "red" and yours are similar or completely different; we have learned to label them the same way regardless of our inner experience (Chalmers, 2007). However, in this case, the brain systems most closely linked to feelings—and which presumably play a central role in their construction—appear to be relatively conserved across individuals. This has been much more extensively tested for some forms of affect (experimentally evoked pain) than for others (empathy), and there is much more work to do, but the way forward seems promising.

These brain models reveal another interesting conclusion about how affective processes are organized in the brain. We are used to thinking of rejection, vicarious pain, somatic pain, hunger, thirst, and disgust as birds of a feather in some respects: They are all negative experiences, shaped over the course of evolution to elicit escape and avoidance. Descriptive models that group human judgments have consistently found that negative emotions are grouped together in our conception and set in contradistinction to positive emotions. But this conceptual similarity need not reflect similarity in the underlying brain systems, which may more precisely determine how they jointly or separately arise and how they might interact with one another. Put simply, there may be no common representation of "negative affect" in the brain. Different systems might, having evolved to respond to particular environmental demands, "feel like something" (or, in more technical terms, be associated

with conscious "qualia"). Viewed in this light, it is easy to imagine one system that evolved to escape from the imminent damage caused by coming too near a fire, and a different one evolved to avoid dying of thirst or being attacked by an angry group member. We may have many systems that represent many types of negative and positive affect, organized in part by the eliciting stimuli, the canonical organism-environment relationships or "situations" involved, the types of actions afforded and the time scale involved, and more. The critical variables that carve the affective brain "at its joints," determining when one affective brain system versus another is engaged, remain to be determined.

This "multiple affect systems" view stands in stark contrast to recent conceptions of the organization of the affective brain. Many studies have highlighted the broad convergence of multiple types of affect on the anterior insula and cingulate (Klimecki et al., 2014; Lamm et al., 2011). Lieberman et al., for example, proposed that the dACC reflects a unitary system that responds to events "relevant for survival" (Lieberman & Eisenberger, 2015). Eisenberger and Cole relate physiological responses in the endocrine and immune systems to a general "alarm system" in the brain (Eisenberger & Cole, 2012). And our own work has highlighted commonalities in the brain systems involved in constructing multiple types of emotional experiences (Ashar, Andrews-Hanna, Dimidjian, & Wager, 2016; Kober et al., 2008; Roy, Shohamy, & Wager, 2012). There may indeed be commonalities at the level of broad systems and concepts like constructing representations of schemas and assigning personal meaning to actions and events. But at the circuit level, the brain processes that mediate particular forms of negative affect—and our behavioral and physiological responses to them—need to be defined with increasing specificity. In animals, there is now overwhelming evidence that specific neural populations mediate specific affective behaviors in response to specific contexts (Lammel, Tye, & Warden, 2014). For example, different populations of neurons in the dACC mediate different types of foraging behavior (Kvitsiani et al., 2013) and specific aspects of pain (Dale et al., 2018; Tan et al., 2017; Zhang et al., 2018). In humans, different local patterns within the dACC track evoked pain and specific types of negative affect and cognitive control (Kragel, Kano, et al., 2018), though there appears to be a pattern that generalizes across multiple types of pain. Even within autonomic responses to stress, largely different systems mediate increases in skin conductance and heart rate (Eisenbarth, Chang, & Wager, 2016), though there appears to be a common core related to the vmPFC.

The implications of this "multiple affect systems" view are not merely academic abstractions. They bear concretely on how we move forward in identifying the systems that confer resilience or risk to disease, and target those systems with biological and psychosocial interventions. If there is a unitary system for "negative affect," one can measure its function with multiple behavioral readouts in humans and animals, and probe it with a variety of interchangeable stimuli (faces, sounds, memory cues). We could identify the molecular substrates of this system and develop pills to cure depression, anxiety, and pain. We could characterize genetic and environmental precursors that lead to its dysfunction, and thus direct prevention and treatment resources to those at greatest risk. But if biology has taught us one overarching

lesson, it is that this way of thinking will not work. Biology is complex, and its pathways and interactions are myriad. Molecular mechanisms that operate in one strain of mouse may not be operative in another, much less in other species. As a result, advancing treatments for mental health disorders has proven extremely difficult. By some accounts, there are no new classes of drugs for depression, anxiety, or pain, in spite of over a trillion dollars spent on drug development over the past decade. Identifying specific pathways and targets, and relating these to specific treatments, is a promising way forward. The future of neuroscience lies in embracing complexity—but in a limited way, simplifying and generalizing where we can, and throughout hewing to the lines nature draws for us that carve the affective brain at its joints.

Acknowledgments We are grateful for support from the National Institutes of Health, including grants R01DA035484 (Tor Wager), R01MH076136 (Tor Wager), R01MH116026 (Luke Chang), T32DA017637 (Philip Kragel), and U01 500470-78051 (Lisa Feldman Barrett). MATLAB code for analyses used in papers and figures is available at: https://github.com/canlab.

References

Alabas, O. A., Tashani, O. A., Tabasam, G., & Johnson, M. I. (2012). Gender role affects experimental pain responses: A systematic review with meta-analysis. *European Journal of Pain, 16*(9), 1211–1223.

Apps, M. A. J., Rushworth, M. F. S., & Chang, S. W. C. (2016). The anterior cingulate gyrus and social cognition: Tracking the motivation of others. *Neuron, 90*(4), 692–707.

Ashar, Y. K., Andrews-Hanna, J. R., Dimidjian, S., & Wager, T. D. (2016). Towards a neuroscience of compassion: A brain systems-based model and research agenda. In J. D. Greene (Ed.), *Positive neuroscience* (pp. 1–27). New York, NY: Oxford University Press.

Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nature Reviews Neuroscience, 7*(5), 358–366.

Banich, M. T. (2004). *Cognitive neuroscience and neuropsychology*. Retrieved from https://scholar.google.ca/scholar?cluster=12756220486095662084&hl=en&as_sdt=0,5&sciodt=0,5

Barrett, L. F., Khan, Z., Dy, J., & Brooks, D. (2018). Nature of emotion categories: Comment on Cowen and Keltner. *Trends in Cognitive Sciences, 22*(2), 97–99.

Becker, S., Gandhi, W., Pomares, F., Wager, T. D., & Schweinhardt, P. (2017). Orbitofrontal cortex mediates pain inhibition by monetary reward. *Social Cognitive and Affective Neuroscience, 12*(4), 651–661.

Bräscher, A.-K., Becker, S., Hoeppli, M.-E., & Schweinhardt, P. (2016). Different brain circuitries mediating controllable and uncontrollable pain. *The Journal of Neuroscience, 36*(18), 5013–5025.

Brett, M., Johnsrude, I. S., & Owen, A. M. (2002). The problem of functional localization in the human brain. *Nature Reviews Neuroscience, 3*(3), 243–249.

Brooks, D. (2014). The archipelago of pain. *The New York Times*, 7.

Brown, J. W., & Alexander, W. H. (2017). Foraging value, risk avoidance, and multiple control signals: How the anterior cingulate cortex controls value-based decision-making. *Journal of Cognitive Neuroscience, 29*(10), 1656–1673.

Carrillo, M., Han, Y., Migliorati, F., Liu, M., Gazzola, V., & Keysers, C. (2019). Emotional mirror neurons in the rat's anterior cingulate cortex. *Current Biology, 29*(8), 1301–1312.e6.

Carter, R. M., & Huettel, S. A. (2013). A nexus model of the temporal–parietal junction. *Trends in Cognitive Sciences, 17*(7), 328–336.

Castro, W. H. M., Meyer, S. J., Becke, M. E. R., Nentwig, C. G., Hein, M. F., Ercan, B. I., … Du Chesne, A. E. (2001). No stress—No whiplash? *International Journal of Legal Medicine, 114*(6), 316–322.

Chalmers, D. (2007). The hard problem of consciousness. In M. Velmans & S. Schneider (Eds.), *The Blackwell companion to consciousness* (pp. 225–235). Oxford, England: Blackwell.

Chang, L. J., Gianaros, P. J., Manuck, S. B., Krishnan, A., & Wager, T. D. (2015). A sensitive and specific neural signature for picture-induced negative affect. *PLoS Biology, 13*(6), e1002180.

Couvy-Duchesne, B., Ebejer, J. L., Gillespie, N. A., Duffy, D. L., Hickie, I. B., Thompson, P. M., … Wright, M. J. (2016). Head motion and inattention/hyperactivity share common genetic influences: Implications for fMRI studies of ADHD. *PLoS One, 11*(1), e0146271.

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences of the United States of America, 114*(38), E7900–E7909.

Cowen, A. S., & Keltner, D. (2018). Clarifying the conceptualization, dimensionality, and structure of emotion: Response to Barrett and colleagues. *Trends in Cognitive Sciences, 22*(4), 274–276.

Dale, J., Zhou, H., Zhang, Q., Martinez, E., Hu, S., Liu, K., … Wang, J. (2018). Scaling up cortical control inhibits pain. *Cell Reports, 23*(5), 1301–1313.

Davis, K. D., Kucyi, A., & Moayedi, M. (2015). The pain switch: An "ouch" detector. *Pain, 156*(11), 2164.

de Knegt, N., & Scherder, E. (2011). Pain in adults with intellectual disabilities. *Pain, 152*(5), 971–974.

Denny, B. T., Kober, H., Wager, T. D., & Ochsner, K. N. (2012). A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *Journal of Cognitive Neuroscience, 24*(8), 1742–1752.

Dosenbach, N. U. F., Nardos, B., Cohen, A. L., Fair, D. A., Power, J. D., Church, J. A., … Schlaggar, B. L. (2010). Prediction of individual brain maturity using fMRI. *Science, 329*(5997), 1358–1361.

Drysdale, A. T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., … Liston, C. (2016). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine, 23*(1), 28–38.

Eisenbarth, H., Chang, L. J., & Wager, T. D. (2016). Multivariate brain prediction of heart rate and skin conductance responses to social threat. *The Journal of Neuroscience, 36*(47), 11987–11998.

Eisenberger, N. I. (2015). Social pain and the brain: Controversies, questions, and where to go from here. *Annual Review of Psychology, 66*, 601–629.

Eisenberger, N. I., & Cole, S. W. (2012). Social neuroscience and health: Neurophysiological mechanisms linking social ties with physical health. *Nature Neuroscience, 15*(5), 669–674.

Eloyan, A., Muschelli, J., Nebel, M. B., Liu, H., Han, F., Zhao, T., … Caffo, B. (2012). Automated diagnoses of attention deficit hyperactive disorder using magnetic resonance imaging. *Frontiers in Systems Neuroscience, 6*, 61.

Engen, H. G., & Singer, T. (2015). Compassion-based emotion regulation up-regulates experienced positive affect and associated neural networks. *Social Cognitive and Affective Neuroscience, 10*(9), 1291–1301.

Fitzgerald, M., & Walker, S. M. (2009). Infant pain management: A developmental neurobiological approach. *Nature Clinical Practice Neurology, 5*(1), 35–50.

Genevsky, A., & Knutson, B. (2015). Neural affective mechanisms predict market-level microlending. *Psychological Science, 26*(9), 1411–1422.

Genevsky, A., Yoon, C., & Knutson, B. (2017). When brain beats behavior: neuroforecasting crowdfunding outcomes. *The Journal of Neuroscience, 37*(36), 8625–8634.

Geuter, S., Boll, S., Eippert, F., & Büchel, C. (2017). Functional dissociation of stimulus intensity encoding and predictive coding of pain in the insula. *eLife, 6*. https://doi.org/10.7554/eLife.24770

Gilead, M., Boccagno, C., Silverman, M., Hassin, R. R., Weber, J., & Ochsner, K. N. (2016). Self-regulation via neural simulation. *Proceedings of the National Academy of Sciences of the United States of America, 113*(36), 10037–10042.

Green, B. G. (2004). Temperature perception and nociception. *Journal of Neurobiology, 61*(1), 13–29.

Grosenick, L., Greer, S., & Knutson, B. (2008). Interpretable classifiers for FMRI improve prediction of purchases. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 16*(6), 539–548.

Hare, T. A., Camerer, C. F., Knoepfle, D. T., & Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 30*(2), 583–590.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage, 87*, 96–110.

Hu, L., & Iannetti, G. D. (2016). Painful issues in pain prediction. *Trends in Neurosciences, 39*(4), 212–220.

Hutchison, W. D., Davis, K. D., Lozano, A. M., Tasker, R. R., & Dostrovsky, J. O. (1999). Pain-related neurons in the human cingulate cortex. *Nature Neuroscience, 2*(5), 403–405.

Jackson, P. L., Brunet, E., Meltzoff, A. N., & Decety, J. (2006). Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. *Neuropsychologia, 44*(5), 752–761.

Jackson, P. L., Meltzoff, A. N., & Decety, J. (2006). Neural circuits involved in imitation and perspective-taking. *NeuroImage, 31*(1), 429–439.

Jepma, M., Koban, L., van Doorn, J., Jones, M., & Wager, T. D. (2018). Behavioural and neural evidence for self-reinforcing expectancy effects on pain. *Nature Human Behaviour, 2*(11), 838–855.

Keysers, C. (2009). Mirror neurons. *Current Biology, 19*(21), R971–R973.

Keysers, C., Kaas, J. H., & Gazzola, V. (2010). Somatosensation in social perception. *Nature Reviews Neuroscience, 11*(6), 417–428.

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology, 97*(6), 4296–4309.

Klimecki, O. M., Leiberg, S., Ricard, M., & Singer, T. (2014). Differential pattern of functional brain plasticity after compassion and empathy training. *Social Cognitive and Affective Neuroscience, 9*(6), 873–879.

Knutson, B., & Genevsky, A. (2018). Neuroforecasting aggregate choice. *Current Directions in Psychological Science, 27*(2), 110–115.

Koban, L., Kross, E., Woo, C.-W., Ruzic, L., & Wager, T. D. (2017). Frontal-Brainstem pathways mediating placebo effects on social rejection. *The Journal of Neuroscience, 37*(13), 3621–3631.

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.

Kolling, N., Wittmann, M. K., Behrens, T. E. J., Boorman, E. D., Mars, R. B., & Rushworth, M. F. S. (2016). Value, search, persistence and model updating in anterior cingulate cortex. *Nature Neuroscience, 19*(10), 1280–1285.

Koyama, T., Tanaka, Y. Z., & Mikami, A. (1998). Nociceptive neurons in the macaque anterior cingulate activate during anticipation of pain. *Neuroreport, 9*(11), 2663–2667.

Kragel, P. A., Kano, M., Van Oudenhove, L., Ly, H. G., Dupont, P., Rubio, A., … Wager, T. D. (2018). Generalizable representations of pain, cognitive control, and negative emotion in medial frontal cortex. *Nature Neuroscience, 21*(2), 283–289.

Kragel, P. A., Knodt, A. R., Hariri, A. R., & LaBar, K. S. (2016). Decoding spontaneous emotional states in the human brain. *PLoS Biology, 14*(9), e2000106.

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.

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.

Krienen, F. M., Tu, P.-C., & Buckner, R. L. (2010). Clan mentality: Evidence that the medial prefrontal cortex responds to close others. *The Journal of Neuroscience, 30*(41), 13906–13915.

Krishnan, A., Woo, C.-W., Chang, L. J., Ruzic, L., Gu, X., López-Solà, M., … Wager, T. D. (2016). Somatic and vicarious pain are represented by dissociable multivariate brain patterns. *eLife, 5*. https://doi.org/10.7554/eLife.15166

Kross, E., Berman, M. G., Mischel, W., Smith, E. E., & Wager, T. D. (2011). Social rejection shares somatosensory representations with physical pain. *Proceedings of the National Academy of Sciences of the United States of America, 108*(15), 6270–6275.

Kvitsiani, D., Ranade, S., Hangya, B., Taniguchi, H., Huang, J. Z., & Kepecs, A. (2013). Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature, 498*(7454), 363–366.

Lamm, C., Decety, J., & Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *NeuroImage, 54*(3), 2492–2502.

Lammel, S., Tye, K. M., & Warden, M. R. (2014). Progress in understanding mood disorders: Optogenetic dissection of neural circuits. *Genes, Brain, and Behavior, 13*(1), 38–51.

Legrain, V., Iannetti, G. D., Plaghki, L., & Mouraux, A. (2011). The pain matrix reloaded: A salience detection system for the body. *Progress in Neurobiology, 93*(1), 111–124.

Lieberman, M. D., Burns, S. M., Torre, J. B., & Eisenberger, N. I. (2016). Reply to Wager et al.: Pain and the dACC: The importance of hit rate-adjusted effects and posterior probabilities with fair priors. *Proceedings of the National Academy of Sciences, 113*(18), E2476–E2479.

Lieberman, M. D., & Eisenberger, N. I. (2015). The dorsal anterior cingulate cortex is selective for pain: Results from large-scale reverse inference. *Proceedings of the National Academy of Sciences of the United States of America, 112*(49), 15250–15255.

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.

Lipton, Z. C. (2016). The mythos of model interpretability. *arXiv [cs.LG]*. Retrieved from http://arxiv.org/abs/1606.03490

López-Solà, M., Koban, L., Krishnan, A., & Wager, T. D. (2017). When pain really matters: A vicarious-pain brain marker tracks empathy for pain in the romantic partner. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2017.07.012

López-Solà, M., Koban, L., & Wager, T. D. (2018). Transforming pain with prosocial meaning: A functional magnetic resonance imaging study. *Psychosomatic Medicine, 80*(9), 814–825.

Losin, E. A. R., Iacoboni, M., Martin, A., Cross, K. A., & Dapretto, M. (2012). Race modulates neural activity during imitation. *NeuroImage, 59*(4), 3594–3603.

Losin, E. A. R., Woo, C.-W., Krishnan, A., Wager, T. D., Iacoboni, M., & Dapretto, M. (2015). Brain and psychological mediators of imitation: Sociocultural versus physical traits. *Culture and Brain, 3*(2), 93–111.

MacDonald, G. (2009). Social pain and hurt feelings. In P. J. Corr & G. Matthews (Eds.), *Cambridge handbook of personality psychology* (pp. 541–555). Cambridge, England: Cambridge University Press.

Miele, D. B., Wager, T. D., Mitchell, J. P., & Metcalfe, J. (2011). Dissociating neural correlates of action monitoring and metacognition of agency. *Journal of Cognitive Neuroscience, 23*(11), 3620–3636.

Murphy, F. C., Nimmo-Smith, I., & Lawrence, A. D. (2003). Functional neuroanatomy of emotions: A meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience, 3*(3), 207–233.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences, 10*(2), 59–63.

Preston, S. D., & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *The Behavioral and Brain Sciences, 25*(1), 1–20. Discussion 20–71.

Reddan, M. C., Wager, T. D., & Schiller, D. (2018). Attenuating neural threat expression with imagination. *Neuron, 100*(4), 994–1005.e4.

Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., & Chun, M. M. (2016). A neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience, 19*(1), 165–171.

Roy, M., Shohamy, D., & Wager, T. D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends in Cognitive Sciences, 16*(3), 147–156.

Sakurai, Y. (1996). Population coding by cell assemblies—What it really is in the brain. *Neuroscience Research, 26*(1), 1–16.

Schumann, K., Zaki, J., & Dweck, C. S. (2014). Addressing the empathy deficit: Beliefs about the malleability of empathy predict effortful responses when empathy is challenging. *Journal of Personality and Social Psychology, 107*(3), 475–493.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *The American Psychologist, 54*(2), 93.

Shenhav, A., Straccia, M. A., Cohen, J. D., & Botvinick, M. M. (2014). Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value. *Nature Neuroscience, 17*(9), 1249–1254.

Shen, X., Finn, E. S., Scheinost, D., Rosenberg, M. D., Chun, M. M., Papademetris, X., & Constable, R. T. (2017). Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nature Protocols, 12*(3), 506–518.

Sikes, R. W., & Vogt, B. A. (1992). Nociceptive neurons in area 24 of rabbit cingulate cortex. *Journal of Neurophysiology, 68*(5), 1720–1732.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science, 303*(5661), 1157–1162.

Sola, M. L., Koban, L., Geuter, S., Coan, J., & Wager, T. (2019). (304) Brain mediators of hand-holding analgesia. *The Journal of Pain, 20*(4, Supplement), S50.

Tamir, D. I., & Mitchell, J. P. (2010). Neural correlates of anchoring-and-adjustment during mentalizing. *Proceedings of the National Academy of Sciences of the United States of America, 107*(24), 10827–10832.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience, 19*, 109–139.

Tan, L. L., Pelzer, P., Heinl, C., Tang, W., Gangadharan, V., Flor, H., … Kuner, R. (2017). A pathway from midcingulate cortex to posterior insula gates nociceptive hypersensitivity. *Nature Neuroscience, 20*, 1591. https://doi.org/10.1038/nn.4645

Todd, M. T., Nystrom, L. E., & Cohen, J. D. (2013). Confounds in multivariate pattern analysis: Theory and rule representation case study. *NeuroImage, 77*, 157–165.

Turk-Browne, N. B. (2013). Functional interactions as big data in the human brain. *Science, 342*(6158), 580–584.

Wager, T. D., Atlas, L. Y., Botvinick, M. M., Chang, L. J., Coghill, R. C., Davis, K. D., … Yarkoni, T. (2016). Pain in the ACC? *Proceedings of the National Academy of Sciences, 113*(18), E2474–E2475.

Wager, T. D., Atlas, L. Y., Lindquist, M. A., Roy, M., Woo, C.-W., & Kross, E. (2013). An fMRI-based neurologic signature of physical pain. *The New England Journal of Medicine, 368*(15), 1388–1397.

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.

Wager, T. D., Lindquist, M., & Kaplan, L. (2007). Meta-analysis of functional neuroimaging data: Current and future directions. *Social Cognitive and Affective Neuroscience, 2*(2), 150–158.

Wager, T. D., Rilling, J. K., Smith, E. E., Sokolik, A., Casey, K. L., Davidson, R. J., … Cohen, J. D. (2004). Placebo-induced changes in FMRI in the anticipation and experience of pain. *Science, 303*(5661), 1162–1167.

Woo, C.-W., Chang, L. J., Lindquist, M. A., & Wager, T. D. (2017). Building better biomarkers: Brain models in translational neuroimaging. *Nature Neuroscience, 20*(3), 365–377.

Woo, C.-W., Roy, M., Buhle, J. T., & Wager, T. D. (2015). Distinct brain systems mediate the effects of nociceptive input and self-regulation on pain. *PLoS Biology, 13*(1), e1002036.

Woo, C.-W., Schmidt, L., Krishnan, A., Jepma, M., Roy, M., Lindquist, M. A., … Wager. (2017). Quantifying cerebral contributions to pain beyond nociception. *Nature Communications, 8*, 14211.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods, 8*(8), 665–670.

Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., … Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology, 106*(3), 1125–1165.

Zaki, J., & Ochsner, K. N. (2012). The neuroscience of empathy: Progress, pitfalls and promise. *Nature Neuroscience, 15*(5), 675–680.

Zaki, J., Schirmer, J., & Mitchell, J. P. (2011). Social influence modulates the neural computation of value. *Psychological Science, 22*(7), 894–900.

Zhang, Q., Xiao, Z., Huang, C., Hu, S., Kulkarni, P., Martinez, E., … Wang, J. (2018). Local field potential decoding of the onset and intensity of acute pain in rats. *Scientific Reports, 8*(1), 8299.

Zunhammer, M., Bingel, U., Wager, T. D., & Placebo Imaging Consortium. (2018). Placebo effects on the neurologic pain signature: a meta-analysis of individual participant functional magnetic resonance imaging data. *JAMA Neurology.* https://doi.org/10.1001/jamaneurol.2018.2017

Index

M. Neta, I. J. Haas (eds.), *Emotion in the Mind and Body*, Nebraska Symposium on Motivation 66, https://doi.org/10.1007/978-3-030-27473-3