Michael Gilead
Kevin N. Ochsner *Editors*

The Neural Basis of Mentalizing

Springer

The Neural Basis of Mentalizing

Michael Gilead • Kevin N. Ochsner
Editors

The Neural Basis of Mentalizing

Springer

Editors
Michael Gilead
Psychology Department
Ben-Gurion University of the Negev
Beer Sheva, Israel

Kevin N. Ochsner
Department of Psychology
Columbia University
New York, NY, USA

ISBN 978-3-030-51889-9 ISBN 978-3-030-51890-5 (eBook)
https://doi.org/10.1007/978-3-030-51890-5

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

Preface

This book, *The Neural Basis of Mentalizing*, comes out 20 years after the publication of Chris Frith and Uta Frith's (1999) groundbreaking *Science* paper "Interacting Minds—A Biological Basis," which described the first major efforts to understand the neural bases of humans' ability to *mentalize*, namely, to reflect on the mental states of the self and of others.

In the 20 years that have passed, the study of mentalizing has continued to mature and has become a central topic in the world of cognitive psychology and cognitive neuroscience. Moreover, mentalizing research plays a crucial role in such diverse fields as social and developmental psychology, emotion research, clinical psychology, linguistics, game theory, artificial intelligence, philosophy of mind, and primatology. The current book provides a comprehensive collection of perspectives on the topic and brings together researchers from these diverse fields.

We hope and believe that the publication of this volume will give rise to further progress in the study of the vital human capacity to understand themselves and others.

We wish to thank all of the numerous contributors to this book, Springer-Nature, and the production team for all of their hard work.

Beer Sheva, Israel — Michael Gilead
New York City, NY, USA — Kevin N. Ochsner

Contents

Part I
Introduction

A Guide to the Neural Bases of Mentalizing

Michael Gilead and Kevin N. Ochsner

Mentalizing is the act of thinking about mental states, i.e., reflecting on one's own mental states and estimating the mental states of others. This capacity plays a crucial role in daily life, and it is widely believed that our advanced mentalizing abilities may be one of the main elements that distinguish us from other animals. As such, much research in psychology, philosophy, and neuroscience has been devoted to studying this process.

As implied by its title, the current book reviews the extant research examining mentalizing at the *neural* level; however, as suggested by Marr (1982), a full understanding of a cognitive process also entails an understanding the *algorithms* that are instantiated by the neural tissue (i.e., the representations and cognitive processes involved in the process), which in turn entails an understanding of the *computation* subserved by these processes (i.e., the challenge that the process is meant to address). In light of this, the 34 chapters in this book provide an analysis of mentalizing across the neural, algorithmic, and computational levels (Fig. 1).

This book comes out 20 years after the publication of Frith and Frith's (1999) influential paper describing the first major efforts to understand the neural bases of the mentalizing capacity. The main findings of these initial studies were that a network of temporo-parietal, anterior temporal, posterior medial parietal, and medial prefrontal regions subserve humans' mentalizing ability. Twenty years later, these findings have been replicated numerous times. Celebrating the 20 years' anniversary of this paper, the current volume begins with a special chapter by Frith and Frith, providing a historical perspective on the emergence of mentalizing research, an overview of current research, and an outline of future directions.

M. Gilead (✉)
Psychology Department, Ben-Gurion University of the Negev, Beer Sheva, Israel
e-mail: mgilead@bgu.ac.il

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

M. Gilead, K. N. Ochsner (eds.), *The Neural Basis of Mentalizing*,
https://doi.org/10.1007/978-3-030-51890-5_1

Fig. 1 The most widely used terms throughout the 33 chapters of the book

Before reviewing the contents of the book, we wish to provide readers with a short guide to some of the theoretical terms and common anatomical labels that are widely used throughout.

Terminological Guide

There are some terms that simply overlap with the term Mentalizing. For example, we see no meaningful difference between the term Mentalizing and terms such as **Mindreading** and **Mental State Inference** or **Mental State Attribution**.

The psychological and neuroscientific literature often interchanges the term **Theory of Mind Reasoning** with the term mentalizing. As noted in the Chapter by Perner et al., a caveat of this is that it may be taken by some to suggest a commitment to a specific theory of how mentalizing takes place (namely, the *Theory Theory* of mentalizing, see chapters in Part II)—which one may or may not endorse.

As noted, to mentalize is to reflect about one's own mental states and estimate the mental states of others. If John slaps Jack, and as a result, Jack becomes angry—it is not necessarily the case that Jack engaged in mentalizing. Only when Jack reflects upon his feelings following the slap or tries to understand why John slapped him, then it is appropriate to say that mentalizing took place. Because mentalizing involves thinking about thinking (or similarly, thinking about emotions, desires, intentions, and so on), mentalizing is often referred to as **Meta-Representational** processing (as "meta-representation" means a representation of a representation).

In order to assess whether the behavior of an individual is indeed driven by Meta-Representation, the central paradigm used in Mentalizing research is the False-Belief paradigm (discussed at length throughout the book). This paradigm assesses whether individuals utilize meta-representations by examining behavior in a situation wherein the participant's representation of the world differs from that of another person. In light of this, the capacity examined on False-Belief tasks is often referred to as **False-Belief Reasoning**. False-Belief reasoning entails mentalizing and pro-

vides evidence for Meta-Representation; however, the terms are not interchangeable (False-Belief reasoning is mentalizing but not vice versa).

As reflected in the False-Belief paradigm, it is often the case that in order for mentalizing to be adaptive, individuals need to see the world from a perspective that differs from their own. The cognitive process of trying to view and interpret the world from the perspective of another person, rather than the self, is termed **Perspective Taking**. This process is believed to require the ability to overcome "egocentric biases" or "decouple the representation of the world as seen by myself and another." The terms Mentalizing and Perspective Taking are often used interchangeably. However, again, it is worth noting that individuals may engage in mentalizing without taking the perspective of another person (despite the fact that this would be maladaptive).

Further confusion is entailed by the fact that the term Perspective Taking also refers to a specific *strategy*, widely studied in social psychology (i.e., the explicit strategy of trying to understand another person by imaging oneself in their predicament). Additionally, the capacity to see the world through the perspective of another person does not necessarily involve mentalizing; for example, it is (at least theoretically) possible that tasks of simple visual Perspective Taking rely on mechanisms that do not involve the sort of thinking about thinking which we refer to as mentalizing.

As noted, the term Perspective Taking refers to the *process* whereby individuals try to see the world from a perspective that differs from one's own. It is assumed that such attempts are often successful and generates an accurate understanding of others' perspectives. In light of this Perspective Taking is often discussed as a *skill*. Similarly, the ability to accurately figure out the contents of another's mind is often termed **Empathic Accuracy** (see discussion in the Chapter by **Hinnekens**, **Ickes**, **Berlamont**, **and Verhofstadt**).

This brings us to another loaded term that partially overlaps with the term mentalizing (and is probably the most widely used in lay discourse), namely, **Empathy**. Just like Mentalizing, Empathy entails an understanding of the mental states of others; however, unlike mentalizing, when we say that John empathized with Jack, it typically means that John shared some aspects of the emotional experience of Jack. Researchers of empathy often speak of Emotional Empathy and **Cognitive Empathy**, with the former entailing experience sharing, and the latter concept being virtually identical to mentalizing.

Finally, people sometimes use the term **Social Cognition** interchangeably with the term mentalizing. The capacity for mentalizing is indeed part of the broader class of Social-Cognitive capacities, but there are various social-cognitive capacities that should not be thought of as mentalizing (see **Malle's** chapter). The term Social Cognition is especially confusing because it also refers to a field of research (namely, social psychological research that adopts the premises and toolkit of cognitive science).

Neuroanatomical Guide

Throughout the book, there are various different anatomical regions that are mentioned with regard to the mentalizing capacity; first and foremost of these is the Mentalizing Network. The Mentalizing Network is comprised of regions in the posterior temporal lobe and parietal lobe, medial prefrontal cortex, medial parietal cortex, and the anterior temporal lobe. This network overlaps with the so-called Default Mode Network (Raichle et al., 2001). However, under some definitions of the Default Network, it includes several areas that extend beyond the classic Mentalizing Network. Below, we provide a quick neuroanatomical guide to the approximate locations of the regions discussed in the chapters.

Throughout the book, crucial parietal and posterior temporal areas for mentalizing are referred to as the pSTS (posterior superior temporal sulcus), TPJ (temporoparietal junction), Angular Gyrus, and IPL (inferior parietal lobule). In Fig. 2, we used the automated meta-analysis tool Neurosynth (Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011) to show the peak coordinates for each of these anatomical labels, as well as the peak posterior temporal and parietal locations of the mentalizing network and default-mode-network.

Crucial medial prefrontal areas for mentalizing are often referred to in the book as the vmPFC (ventromedial prefrontal cortex), dmPFC (dorsomedial prefrontal cortex), and simply mPFC (medial prefrontal cortex). In Fig. 3, we show the peak coordinates for each of these anatomical titles, as well as the peak medial prefrontal coordinates of the mentalizing network and default-mode-network.

Crucial medial parietal areas for mentalizing are referred to in the book as PCC (posterior cingulate cortex) and Precuneus. In Fig. 4, we show the peak coordinates for each of these anatomical titles, as well as the peak medial parietal coordinates of the mentalizing network and default-mode-network.

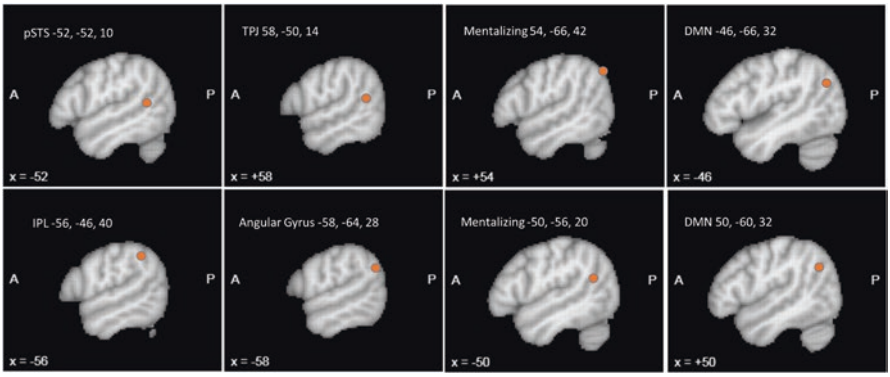

Fig. 2 Peak Neurosynth coordinates of main parietal and posterior temporal regions discussed throughout the book. Also shown are peak temporal and parietal coordinates for the terms Mentalizing and Default Mode Network

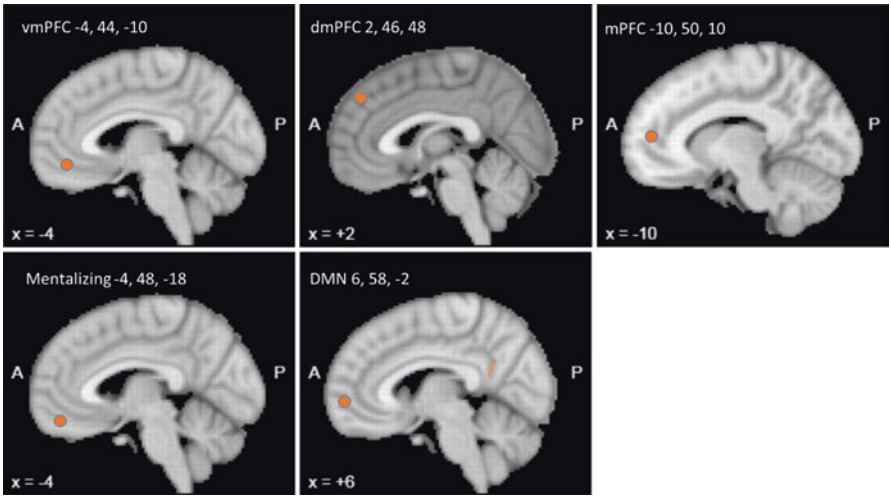

Fig. 3 Peak coordinates of medial prefrontal regions discussed throughout the book, and medial prefrontal peaks of the Mentalizing Network and Default Mode Network

Finally, a crucial anterior temporal region in the mentalizing network is often referred to in the book as the Temporal Pole. In Fig. 5, we show the peak coordinates for this anatomical title, as well as the peak anterior temporal coordinates of the mentalizing network.

Chapter Guide

As noted, the current volume begins with a special 20 years' anniversary chapter by **Frith and Frith**, providing a historical perspective on the emergence of mentalizing research, an overview of current research, and an outline of future directions. The subsequent chapters are organized into five parts: (i) The boundaries of mentalizing; (ii) Theoretical approaches to the study of mentalizing; (iii) The components of mentalizing; (iv) Mentalizing in social interactions and decision-making; (v) Mentalizing in self-referential processing and emotion.

Part I: The Boundaries of Mentalizing

One of the ways by which scientists can study a phenomenon is by carefully mapping its boundaries. In the case of mentalizing such research has attempted to examine the extent to which humans are able to accurately infer the mental states of others and outline the degree to which different populations (e.g., infants,

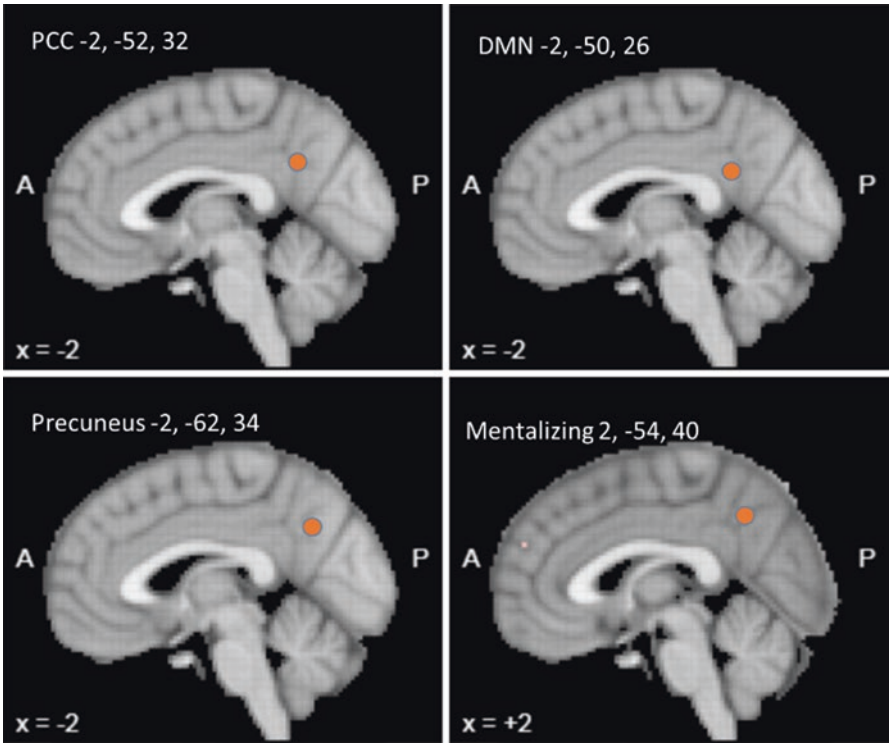

Fig. 4 Peak coordinates of posterior medial parietal regions discussed throughout the book, and posterior medial parietal peaks of the Mentalizing network and Default Mode Network

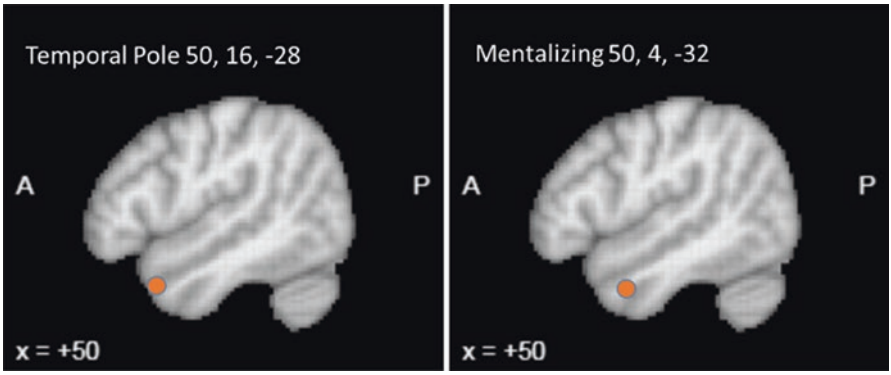

Fig. 5 Peak coordinates of anterior temporal regions discussed throughout the book, and anterior temporal peaks of the Mentalizing network

neuroatypical individuals) and different species (e.g., nonhuman primates) exhibit mindreading capacities. Thus, the first part of the book describes research that maps the mentalizing ability by establishing its boundaries.

As reviewed in the chapter by Frith and Frith, the extensive research program on humans' capacity for mental state processing can be largely traced back to a highly influential paper, by Premack and Woodruff (1978), that asked the question of whether nonhuman primates have an explicit "theory of mind." As a response to this paper, several philosophers suggested a litmus test to determine whether an organism indeed represents the mental states of others—a test subsequently termed the false-belief task. Several years later Wimmer and Perner (1983) devised a false-belief task and ran it on young children and discovered that it is only by the age of 4 that children are able to perform this task successfully.

Following the finding that children fail to pass false-belief tasks before the age of 4, an additional crucial milestone in mentalizing research came from the discovery that children with Autism Spectrum Disorder likewise often fail false-belief tasks (Baron-Cohen et al., 1985). Such findings have led to an extensive research program on impairments in mentalizing capacities in neurotypical and neuroatypical populations. In their chapter, **Simantov**, **Lombardo**, **Baron-Cohen**, **and Uzefovsky** discuss one specific process that may be impaired in neuroatypical individuals and may play a crucial role in subserving the capacity to mentalize—namely, the capacity for *self-other distinction*. An alternative view of mentalizing impairments in Autism is discussed by **Herrington**, **Parish-Morris**, and **Schultz**, implicating decreased social motivation in neuroatypical individuals.

Whereas the research program initiated by Wimmer and Perner's (1983) false-belief paradigm research suggested that a full-blown theory of mind does not develop until the age of 4, further research has suggested a dissociation between infants' explicit responses on false-belief tasks and their behavior. This research (e.g., Onishi & Baillargeon, 2005; Southgate, Senju, & Csibra, 2007) showed that eye-movement patterns of infants as young as 2 years old might reflect advanced implicit mentalizing capacities. In recent years, these results and their interpretation have been hotly debated; in their chapter, **Grosse Wiesmann and Southgate** provide an attempt towards an integration of supposedly incongruent past findings—that arrives at the interesting conclusion that while infants may not have fully fledged mentalizing capacities, they may represent the world in an altercentric manner; namely, the default condition for humans may be to see the world as others see it.

In a nice example of a long scientific journey, the eye-tracking measures developed for the study of mentalizing ability in young children provided a breakthrough in research into Premack and Woodruff (1978) question concerning the existence of theory-of-mind in nonhuman primates. After almost 40 years of research, Krupenye, Kano, Hirata, Call, and Tomasello (2016) utilized eye-tracking to provide compelling evidence that, indeed, chimpanzees seem to pass the false-belief tasks, and as such may have meta-representational capacities. In the current volume, **Krupenye** discusses this research and what it means for questions regarding the evolution of Mentalizing. In contrast, **Arre and Santos** voice a more pessimistic view of the

mentalizing capacities of nonhuman primates and suggest that further research is warranted in order to rule out the possibility that mentalizing is a uniquely human capacity.

Valuable insights concerning mentalizing and its limitations come from behavioral research on neuro*typical* adults. For example, much research has utilized paradigms gauging participants' empathic accuracy (Ickes, 1993) to quantify the extent to which individuals are able to correctly read the minds of others. The chapter by **Hinnekens**, **Ickes**, **Berlamont**, **and Verhofstadt** reviews the rich literature and the consequence of empathic accuracy to social interaction and well-being. **Jhurry and Harris** review the contextual factors that determine when and why individuals fail to mentalize and the different implications of reduced mentalizing. Finally, the chapter by **Atias and Aviezer** evaluates the extent to which people are able to infer mental states from facial expressions and vocalizations. They highlight how, contrary to popular belief, it is often nearly impossible to gauge people's emotions based purely on such nonverbal information.

Part II: Theoretical Approaches to the Study of Mentalizing

Throughout the years several prominent theoretical perspectives have attempted to explain the mechanisms by which mentalizing takes place. The chapters in Part II review these different scientific approaches.

Houlihan, **Tenenbaum**, **and Saxe** present an overview of their extant work on computational modeling of the mentalizing capacity with the "Bayesian Theory of Mind" approach, as well as their research on the type of information represented within the mentalizing network. They discuss their view according to which the computations performed by the mentalizing network can be best described as operations that occur over structured representations of mental content, wherein the different elements stand in specified relations to each other—just like in a scientific *theory* (Carey, 2009; Gopnik & Wellman, 1994).

As Houlihan et al. note, this approach, which describes mental state inference as in terms of structured relations between representations of mental states and behaviors, stands in contrast to the recent computational model proposed by Tamir and Thornton. This approach, described in the chapter by **Thornton and Tamir**, suggests that we infer people's mental states by forming a low-dimensional representation of mental states, traits, and actions—and the transition probabilities between different points along this representation. The findings described in Thornton and Tamir's chapter provide evidence that the mentalizing network may indeed represent the mental world by relying on nonstructured associations between elements in a low-dimensional space, thus challenging the structured representation approach. In contrast, findings reported in Houlihan et al. challenge the idea that a low-dimensional representation suffices to explain the variance in mental state attributions.

Such conflicting models may provide a roadmap for empirical research into the computational and neural bases of mentalizing. Furthermore, as reviewed in **Gonzalez and Chang's** guide to computational models of mentalizing, in recent years, several additional computational models provide researchers with viable alternatives, and that will likely guide future empirical investigation in the field.

The approach described by Houlihan et al. wherein mentalizing is achieved via statistical inference processes upon structured meta-representations is a reincarnation of the classic philosophical view termed the theory-theory of mental state inference. An article published in 1986 by Robert Gordon provided compelling challenges to the theory-theory approach and instead argued for what is termed the *Simulation* theory. Simulation theory suggests that in order to understand an organism like myself, I do not have to generate a theory of myself—but rather simply examine my own responses in similar situations. In his chapter, **Gordon** discusses the simulation view as a natural consequence of recent computational models of cognitive processing, namely, the *predictive processing* approach. Furthermore, Gordon argues that the computational toolkit utilized by recent incarnations of the theory-theory (i.e., the *probabilistic generative modeling* approach) naturally invites the view that mental state inference does not involve operations upon a model of mental states, but rather an inversion of one's own mechanisms for action selection.

One of the classic arguments against the simulation approach is that the self is oftentimes an inappropriate model for thinking about the mental states of others. Like Gordon, **Perner**, **Aichhorn**, **Tholen**, **and Schurz** suggest that people nonetheless start off with an assumption of identity between self and other. The *Teleological* theory, described by Perner et al., presents a stark alternative to specialized accounts of mental state inference and argues that mentalizing happens via reliance on general-purpose cognitive processes that we use in order to think about nonmental content. Specifically, the ability to think about reasons for action and the ability to find the identity between different senses of the same referent object.

While the different perspectives presented in this chapter significantly differ from each other, they are all comfortably situated within the realms of the *information processing approach*; namely, the idea that "the mind is a system of organs of computation, designed by natural selection to solve the kinds of problems our ancestors faced in their foraging way of life" (Pinker, 1997, p. 21, cited in Carpendale et al.). **Carpendale**, **Müller**, **Lewis**, **and Wallbridge** suggest that explaining mentalizing (and cognition more broadly) by using an ontology of such "organs of computation" is the wrong idea. Instead, they present a perspective on mentalizing they term the "process-relational" account, which focuses on understanding how children construct meaning through an interaction with their physical and social world.

Part III: The Components of Mentalizing

In contrast to the dissenting view championed by Carpendale et al., the way cognitive scientists have typically attempted to explain the process of mentalizing is by breaking this process down to its constituent computational components, and trying

to understand how these components interact in the mind and brain. In Part III we discuss the research into these purported cognitive components and how they are implemented in the brain.

In order to provide a basis for such a componential view, **Malle** contextualizes the process of mentalizing within the broader realm of social-cognitive capacities and proposes a hierarchically organized set of different computations that allow humans to understand the minds of others. This model of a "tree of social cognition" provides an ontology of the components of social cognition and their relation to each other. **Apperly** describes a process model that explicates the temporal progression of the cognitive components involved in mentalizing. Specifically, Apperly proposes a dual-system approach to mindreading that constitutes a slow-but-flexible processing stream that heavily relies on retrieval from long-term memory and on cognitive control and a fast-but-inflexible stream that entails lower cognitive demands.

The chapter by **Van Overwalle and Heleven** provides an in-depth review of the neuroscientific research into many of the component processes discussed in Apperly's process model. Specifically, the chapter reviews extant research on the neural basis of the "slow" and "fast" social cognition and on the cognitive control of memory during mentalizing. Furthermore, they describe a theory of how different regions of the Mentalizing Network play a role in these different component processes.

While *social*-cognitive neuroscientists often study the Mentalizing Network in the context of social cognition tasks, this network has also been the focus of much research by *cognitive* neuroscientists that have attempted to characterize the broader (and potentially non-social) underlying computations of this network of regions, also termed the Default Mode Network.

For example, research into semantic cognition has revealed that the neural bases of semantic processing overlap with the Default Network (Binder, Desai, Graves, & Conant, 2009). In light of this, in his chapter, **Binder** proposes that researchers should think of mentalizing-related activity as a manifestation of conceptual processing; namely, that mentalizing is simply an application of humans' more general capacity for conceptual thought in the social domain. Relatedly, in her chapter, **de Villiers** surveys the research that examined the role of language processing in mentalizing and considers the evidence for several views regarding the relation mentalizing conceptual and linguistic processing.

Whereas much research suggests the Default Network is associated with *semantic* processing and semantic memory, this network is also widely implicated in supporting *episodic* memory (e.g., Schacter, Norman, & Koutstaal, 1998). In their chapter, **Van Genugten and Schacter** review the evidence that parts of the DMN subserve the retrieval of specific event details and explain how such retrieval processes can be used to simulate the content of another person's mind and their future behavior.

Finally, given the involvement of the DMN in both semantic and episodic memory, **Baror**, **Aminoff**, **and Bar** present their view of this network as subserving a potentially more fundamental process of context-based associative processing.

They explain how such contextual associations may play a role in semantic cognition, episodic memory, and mentalizing.

Part IV: Mentalizing in Social Interaction and Decision-Making

A comprehensive understanding of the process of mentalizing entails understanding its component processes, neural basis, but also, importantly, the role of this process in adaptive human behavior. In recent years, much research by social-cognitive neuroscientists has elucidated how mentalizing subserves social interaction and decision-making. In Part IV, the contributors review these research endeavors.

It is widely argued that the evolutionary success of *Homo sapiens* relied on our ability to cooperate with each other and accumulate cultural knowledge (Boyd, Richerson, & Henrich, 2011). This unique human capacity is predicated on our ability to effectively communicate with each other and on our motivation to help each other out (Tomasello, Carpenter, Call, Behne, & Moll, 2005). As discussed in the chapter by **Parelman**, **Doré**, **and Falk**, neuroscientific research highlights the importance of mentalizing processes and of the mentalizing network in communication behaviors. Additionally, **Franklin-Gillette and Shamay-Tsoory** discuss the neuroscientific literature on empathy and its role in prosocial behavior such as providing emotional support for others. These two chapters explain how mentalizing lies at the heart of our species' highly interdependent, cooperative and helpful nature.

To survive in the evolutionary arms race, organisms developed mechanisms that help them attain survival-related goals. Within the world of decision science, the attainment of survival-related goals is described as providing organisms with some *utility* or *value*; decision scientists try to understand the mechanisms that subserve utility or value-seeking behavior. As highlighted by **Charpentier and O'Doherty**, because humans are fundamentally social creatures, attainment of value often occurs with the help of others (e.g., observing their behaviors to understand which actions yield positive consequences) or by competing with others (e.g., in economic negotiations)—and as such, relies on mental state inferences. In their chapter, they describe neuroscientific research that utilized computational modeling of decision-making, explicating the role of mentalizing in such decision process. Whereas Charpentier and O'doherty primarily focus on reward seeking, **Espinosa, Golkar and Olsson** mainly discuss research on learning from the behaviors of others with regard to punishments and highlight how such learning plays a crucial role in shaping our affect experience and decisions—and the role of mentalizing in such processes.

As highlighted by **Civai and Sanfey**, because our lives are embedded within social contexts, many decisions we make do not only affect material rewards and punishments but also influence our standing in the social world. In their chapter, they review neuroeconomic research showing how social considerations such as prosocial motivation, fairness, and reputation affect decision-making processes;

moreover, they discuss how our decision-making processes are shaped by our expectations of others' social behavior.

A fundamental process in forming such expectation involves generating an impression of people's stable dispositions; as argued by **Ray**, **Mende-Siedlecki**, **Gantman**, **and Van-Bavel**, the main factor in dispositional inferences is an evaluation of a person's moral character. The findings discussed in these two chapters highlight the involvement of mentalizing network in shaping these important social decisions and inferences.

One of the assumptions of decision science is that human beings are rational creatures (i.e., the so-called "homo-economicus" perspective). A significant challenge to this economic perspective to human behavior is human cognition is often biased and appears to be suboptimal. For example, in social decision-making people often perceive reality as they want it to be (e.g., everybody loves me), rather than how it truly is. In the final chapter of this part, **Park**, **Kim**, **and Young** discuss the notion of rational choice and suggest that activity in the mentalizing network may be a marker of procedurally rational behavioral (i.e., decision that is consistent with Bayesian Decision Theory). Together, the findings reviewed in Part IV highlight how understanding the process of mentalizing is crucial for understanding other vital aspects of human behavior.

Part V: Mentalizing in Self-Referential Processing and Emotion

In the final part of the book, the authors discuss how mentalizing shapes our sense of self and our emotional experience—and thereby how it plays a role in psychopathology.

According to the symbolic interactionist perspective (Mead, 1934), our mental representation of the self is construed from representations of other people and of the way they see us—and as such, the sense of self depends on the process of mentalizing. Indeed, as reviewed in the chapter by **Chavez**, the neuroscientific research on representation of self and others shows that the neural basis of self- and other-related processing is highly interdependent.

It has been suggested (e.g., Leary & Baumeister, 2000) that the very reason that we need a sense of self is because it allows us to keep track of how others view us (e.g., whether they think we are valued members of society), and thereby assess our social standing, and the risks of being ostracized by our peers. Because of this, perceptions of the self are intimately related to our affective experience and well-being.

Much research, reviewed by **Sahi and Eisenberger**, suggests that the feeling of being negatively evaluated by others—which relies on process of mentalizing—is literally a painful experience. Moreover, in light of the potentially aversive nature of self-evaluation, it should be unsurprising that when self-evaluative processes become dysregulated, psychopathology ensues. Based on such reasoning, **Maresh and Andrews-Hanna** discuss how aberrations in self- and other-related processing can give rise to social anxiety disorder.

The idea that mentalizing processes play a central role in psychopathology has become central in the world of psychotherapy research and practice, wherein it is referred to as the *mentalization* approach (e.g., Fonagy & Bateman, 2008). This approach has given rise to one of the leading psychotherapeutic methods in modern clinical psychological science, namely, *mentalization-based treatment*. The premises of the rich theory and their relation to the neuroscientific findings on mentalizing are reviewed in a chapter by **Luyten**, **De Meulemeester**, **and Fonagy**.

Summary

More than 40 years after Premack and Woodruff (1978) and 20 years after Frith and Frith (1999), the study of mentalizing continues to blossom, with new theories and discoveries published every week. While there are many unanswered questions and controversies, it seems to us that the steady advances in research into this topic highlight that the study of mentalizing is one of the most successful enterprises in psychological science. The 33 chapters in the current volume provide what is likely the most comprehensive collection of perspectives published to date on the topic. Hopefully, the publication of this volume will prompt further advances in the study of the fundamental and crucial human capacity to understand themselves and others.

References

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex, 19*(12), 2767–2796.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*(1), 37–46.

Boyd, R., Richerson, P. J., & Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences, 108*(Suppl 2), 10918–10925.

Carey, S. (2009). *The Origin of Concepts*. New York: Oxford University Press.

Fonagy, P., & Bateman, A. (2008). The development of borderline personality disorder - A mentalizing model. *Journal of Personality Disorders, 22*(1), 4–21. https://doi.org/10.1521/pedi.2008.22.1.4

Frith, C. D., & Frith, U. (1999). Cognitive psychology - Interacting minds - A biological basis. *Science, 286*(5445), 1692–1695. https://doi.org/10.1126/science.286.5445.1692

Gopnik, A., & Wellman, H. M. (1994). The theory. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (p. 257). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511752902.011

Ickes, W. (1993). Empathic accuracy. *Journal of Personality, 61*(4), 587–610. https://doi.org/10.1111/j.1467-6494.1993.tb00783.x

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science, 354*(6308), 110–114. https://doi.org/10.1126/science.aaf8110

Leary, M. R., & Baumeister, R. F. (2000). The nature and function of self-esteem: Sociometer theory. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 32, pp. 1–62). San Diego, CA: Elsevier Academic Press Inc..

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information* (2nd ed.). New York, NY: Henry Holt and Co. Inc.

Mead, G. H. (1934). *Works of George Herbert Mead: Vol. 1. Mind, self, and society: From the standpoint of a social behaviorist*. Chicago, IL: University of Chicago Press.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*(5719), 255–258. https://doi.org/10.1126/science.1107621

Pinker, S. (1997). *How the mind works*. New York, NY: W. W. Norton & Company

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(04), 515–526.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences, 98*(2), 676–682.

Schacter, D. L., Norman, K. A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology, 49*, 289–318. https://doi.org/10.1146/annurev.psych.49.1.289

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science, 18*(7), 587–592. https://doi.org/10.1111/j.1467-9280.2007.01944.x

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28*(5), 675. https://doi.org/10.1017/s0140525x05000129

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs - Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

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–U695. https://doi.org/10.1038/nmeth.1635