MENTALIZING AND ITS INFLUENCE ON FACE MEMORY:
A SOCIAL COGNITIVE AND NEURAL INVESTIGATION

A Dissertation in
Psychology
by
Robert G. Franklin, Jr.

© 2012 Robert G. Franklin, Jr.

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2012
The thesis of Robert G. Franklin, Jr. was reviewed and approved* by the following:

Reginald B. Adams Jr.
Assistant Professor of Psychology
Dissertation Advisor
Chair of Committee

Nancy A. Dennis
Assistant Professor of Psychology

Theresa K. Vescio
Associate Professor of Psychology

Paul Eslinger
Professor of Neurology
College of Medicine, Hershey Medical Center

Melvin Mark.
Professor of Psychology
Head of the Department of Psychology

*Signatures are on file in the Graduate School
ABSTRACT

Mentalizing, or determining what another is thinking or feeling, and face memory are two critical skills for social interaction. Despite extensive work about how we remember others and how we mentalize about them based on their face, little research has investigated if these two processes are related. I investigated if how much one mentalizes about a face affects face encoding. In Studies 1 and 2, I found that manipulating level of mentalizing led to greater memory for faces. Second, I examined the relationship between brain regions involved with mentalizing and those involved in face memory. In Study 3, faces that elicited more mentalizing led to greater activation in memory and mentalizing regions in an encoding memory task. In Study 4, experimentally manipulating mentalizing led to greater activation in mentalizing and face memory regions during a face encoding task.
# TABLE OF CONTENTS

LIST OF TABLES.................................................................................................................. vi
LIST OF FIGURES................................................................................................................ vii
ACKNOWLEDGEMENTS ....................................................................................................... viii

## CHAPTER 1. THE RELATIONSHIP BETWEEN MENTALIZING AND MEMORY ..... 1

UNDERSTANDING OTHERS ............................................................................................ 5
MECHANISMS OF MENTALIZING ................................................................................... 8
  * Simulation Theory and Mirror Neurons .................................................................... 8
  * Theory-theory ........................................................................................................ 11
  * Neural Mechanisms for Mentalizing ..................................................................... 13
SOCIAL-PERCEPTUAL MENTALIZING AND THE FACE .................................................. 15
SUMMARY ......................................................................................................................... 18
PROCESSING AND REMEMBERING FACES .................................................................. 19
COGNITIVE MODELS OF FACE MEMORY .................................................................... 20
NEURAL MODELS OF FACE MEMORY ......................................................................... 24
MENTALIZING AND MEMORY ....................................................................................... 27
MEMORY AFFECTING MENTALIZING ............................................................................ 28
  * Simulation and the Prospective Brain ..................................................................... 28
  * Person Knowledge and Stereotypes ....................................................................... 29
MENTALIZING AFFECTING MEMORY .......................................................................... 31
  * Evidence for a Relationship .................................................................................. 32
    Parallel Ingroup Biases ......................................................................................... 32
  * Facial Emotion and Memory Processing ................................................................ 34
    Direct Evidence ..................................................................................................... 35

## CHAPTER 2. BEHAVIORAL MANIPULATIONS ................................................................ 37

STUDY 1 ............................................................................................................................ 38
  * Participants ............................................................................................................. 38
  * Stimuli .................................................................................................................... 38
  * Procedure .............................................................................................................. 39
  * Results ................................................................................................................... 40
  * Discussion ............................................................................................................ 43
STUDY 2 ............................................................................................................................ 44
  * Participants ............................................................................................................. 44
  * Stimuli .................................................................................................................... 45
  * Procedure .............................................................................................................. 45
  * Results ................................................................................................................... 45
  * Discussion ............................................................................................................ 47
CHAPTER 3. NEURAL MECHANISMS OF A RELATIONSHIP ........................................ 49
STUDY 3 .................................................................................................................. 49
Participants ............................................................................................................ 50
Stimuli ..................................................................................................................... 50
Design and Procedure ............................................................................................ 50
Scanning Parameters ............................................................................................ 51
RESULTS .................................................................................................................. 52
Discussion .............................................................................................................. 56
STUDY 4 .................................................................................................................. 56
Participants ............................................................................................................ 57
Stimuli ..................................................................................................................... 58
Design and Procedure ............................................................................................ 58
Scanning Parameters and Analysis ......................................................................... 60
Results ..................................................................................................................... 62
Behavioral Results .................................................................................................. 62
fMRI Results .......................................................................................................... 64

CHAPTER 4. SUMMARY AND DISCUSSION ....................................................... 78
CONCLUSIONS ...................................................................................................... 92

APPENDIX A: LIST OF MENTAL STATE WORDS USED FOR PRE-RATING TASK. 94
APPENDIX B. META-ANALYSIS OF BEHAVIORAL RESULTS ......................... 95

REFERENCES ........................................................................................................... 96
List of Tables

Table 1. Regions active in Study 3................................................................. 54
Table 2. Regions active for complex versus basic adjectives in Study 4............... 65
Table 3. Regions active for hits versus misses in Study 4 .................................. 69
Table 4. Regions active for interaction in Study 4............................................. 73
List of Figures

Figure 1. High confidence hits and misses for Study 1................................. 43
Figure 2. High confidence hits and misses for Study 2................................. 47
Figure 3. FMRI results for Study 3............................................................. 55
Figure 4. High confidence hits and misses for Study 4................................. 63
Figure 5. Whole-brain activation for complex minus basic adjectives in Study 4... 66
Figure 6. Selected activations for complex minus basic adjectives in Study 4........ 67
Figure 7. Whole-brain activation for hits minus misses in Study 4.................... 70
Figure 8. Selected activations for hits minus misses in Study 4....................... 71
Figure 9. Whole-brain activation for interaction in Study 4............................ 74
Figure 10. Selected activations for interaction in Study 4............................... 75
I am deeply grateful to all those who made this work possible. This work is not my own but a testament to what they have given to me; their gracious generosity and boundless kindness made this possible. First, this includes my family, my wife Rebecca and my parents Robert and Diane Franklin. They inspired me, celebrated with me through the successes, and consoled me through the favors. I am deeply grateful to them and for what they did. I am also thankful for my extended family, my grandparents, aunts and uncles, and cousins. Though I may have gone far from home into realms they may not have understood during this journey, they were with me and I with them. The work is theirs as much as it is mine. I am also grateful to my advisor Reginald Adams, Jr., who nourished and formed me academically. He is a mentor on this journey and the work and ideas contained here are as much (or more) his creative manifestation as mine. I am also deeply grateful to Amanda Gearhart, Cara Henry, Melissa Mamigonian, and Kathleen VanHorn for all their help in collecting the data in this dissertation. Without their effort, this would not have been possible. Above all, however, I am grateful, eternally and completely, to God, whose inexorable love and intractable grace manifest by Christ, makes all things possible.
Chapter 1. The Functional Relationship Between Mentalizing and Face Memory

“No man is an island, entire of itself.” - John Donne, Meditation XVII

We are born into a social world, an intricate network of connections that are essential for our survival and psychological well-being. We are born into community, a society filled with others who impact and form who we are. We are interconnected in complex ways, related to others by shared genes and experiences and the social bonds we form. And ultimately, we are connected by the simple commonality of our shared existence. The social world we live in is no accident, but an inevitable manifestation of how we are made. The human brain and mind does not merely function in a social environment but is crafted to be intrinsically social. Our brain develops in social contexts and these contexts affect our mental processes. This is true for even very basic cognitive processes such as perception, attention, cognition, and memory.

The visual system is the most important perceptual mechanism we have for learning about our social world. Based on what we see, we learn much about our social environment, perceiving who is around us, and what they are like. Because so much important social information is around us, our visual system seeks and attends to this information often without conscious awareness. Nowhere is this more apparent than how we perceive and process faces. The face is simply the most important social visual cue. It is what we attend to when we interact with another. We prefer to socialize face to face, and feel these interactions are most meaningful. A face displays in one assemblage a person’s identity, sex, race, age, attractiveness, and locus of
attention. Despite the fact we are admonished not to “judge a book by its cover,” we gather all this information about a person, even with a single glance, and use it to influence how we think about others (Zebrowitz & Montepare, 2008).

The face is also special because it provides us a glimpse into the minds of others. We use facial cues to mentalize about others, determine what they are feeling or thinking (Amodio & Frith, 2006; Brothers, 1990; Castelli, 2005; U. Frith & C. Frith, 2001). Many perceptual and situational cues help us deduce others’ mental states, but the face is unique in how much information it conveys about what we are thinking. We are able to decode another person’s emotions and complex mental states based on facial cues quickly, accurately, and consistently (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; Ekman & Friesen, 1971). The ability to mentalize about another based on only the most subtle nuances of their facial cues makes the face an exquisite social cuing device.

In addition to displaying information about what we are thinking, the face also represents who we are. The concept of one’s “face” as representing one’s identity to the world is common to many cultural traditions, including Western and Chinese cultures. Indeed, we use the face more than any other social stimulus to identify others. We are able to sort between thousands of other people based on only minute differences in facial structure and texture and recognize others from different angles and even under conditions of high visual noise (Duchaine & Nakayama, 2006). In order to use the face as a representation of one’s identity, we must have the ability to remember thousands of others and be able to match a face we perceive from our enormous database of exemplar faces we have encoded.

Remembering others and knowing what they are thinking are only two of the mental processes we engage in when we perceive others’ faces. Instead of being overwhelmed by all this
information displayed in a face, the human visual system decodes this rich meaning effortlessly and with minimal conscious effort. In order to quickly process such information, we use shortcuts: categorizations, heuristics, and other simplified ways of processing in order to streamline the social cognitive processing demands of a complex world (Fiske & Taylor, 1984). This cognitive miser framework can help explain how we negotiate the complexity of the social world, but does not specifically mention what mechanisms by which this simplification occurs. The cognitive miser framework is consistent with the ecological model of perception, which proposes that perception is not a passive process, but an active process designed to seek out information relevant for adaptive purposes (Gibson, 1979; McArthur & Baron, 1983). Applying these models to social visual perception, both the cognitive miser framework and ecological model of perception suggest that the different sources of information from which we decode social messages may be streamlined, specifically for adaptive purposes.

Building on these approaches, the compound social cue model of face perception asserts that various facial cues (such as expression, eye gaze, gender, or race) share common underlying social messages and that our perceptual system is disposed to pick up on these messages across social modalities (Adams, Franklin, Nelson, & Stevenson, 2010). The compound social cue model asserts each of these cues may share underlying signal values especially relevant for adaptive social functioning.

Mentalizing and face memory are two critical skills for social interaction. Separate lines of research describe how we remember others and determine their mental states, describing separate cognitive and neural mechanisms for each system. However, the compound social cue model suggests that the underlying mental processes involved in decoding this information from the face are thus necessarily related. If this is the case, the separate processes of mentalizing and
face memory, with separate neural substrates and separate cognitive mechanisms, may functionally interact.

In this dissertation, I explore the relationship between mentalizing and face memory. Though research has extensively examined both of these processes separately, I present evidence here addressing how mentalizing and face memory are related on both a cognitive and neural level. Below, I first review literature examining mentalizing ability and its underlying cognitive and neural mechanisms. Then, I review research examining face memory. Finally, I review the nascent research examining the functional correspondence between the two processes.

In Chapter 2, I present two behavioral studies showing the relationship between mentalizing and face memory, indicating that greater levels of mentalizing lead to greater memory for faces. In Chapter 3, I build on these behavioral findings by examining their neural underpinnings. I present two studies using functional MRI, revealing that greater mentalizing of faces increases processing neural networks underlying both mentalizing and face memory. Further, this neural response to mentalizing predicts subsequent face memory performance. Chapter 4 reviews these findings, stressing both the theoretical implications of these findings to cognitive and neural models of face memory as well as their potential application to applied research, including the own-race bias in face memory.
Understanding Others

“Though our brother is upon the rack, as long as we ourselves are at our ease, our senses will never inform us of what he suffers. They never did, and never can, carry us beyond our own person, and it is by the imagination only that we can form any conception of what are his sensations. Neither can that faculty help us to this any other way, than by representing to us what would be our own, if we were in his case. It is the impressions of our own senses only, not those of his, which our imaginations copy.” - Adam Smith, *The Theory of Moral Sentiments*

Within each of us exists a rich mental experience, made up of an awareness of our physiological state and environment, as well as the thoughts and feelings that make up our conscious existence. We believe these invisible processes guide us and make up who we are, and it is by those experiences that we can know that we exist (Descartes, 1641/1901). Because of our own mental experiences, we assume others have the same experiences, even though we have no direct evidence of others' mental states and experiences (Dennett, 1987). We assume others are like us and view the world as we do, thinking and feeling in the same way (e.g., Birch & Bloom, 2004) and so we strive to understand what those invisible thoughts and feelings that guide others are.

We assume that we are able to understand what others are feeling and thinking and that those feelings or thoughts lead to their behaviors. Because of this, we place a high social value on inferring the mental states of others. We want to please others and avoid angering them. We also judge people’s actions based on their mental states. We praise others’ skill in success when people intentionally succeed but attribute their accidental successes as “luck”. We also judge others’ morality based on these mental states. Western legal systems punish criminals based on
culpability, or the concept that crime is “constituted only from concurrence of an evil-meaning mind,” thus affirming the necessity of a *mens rea*, or a guilty mind, in criminal proceedings (Morissette v. United States, 1952). This is not just the case in Western law traditions, as ancient penal codes dating back to the ancient Code of Hammurabi assert the importance of a criminal's mental states in administering justice to lawbreakers (Roth, 2005). Implicit in all of these examples is that we are able to understand others’ minds.

As we place a high importance on knowing what their thoughts, feelings, and intentions are, and as such, considerable psychological study has examined this ability. The ability to understand another's mental state is called by a variety of names within psychological literature, such as mentalizing (Amodio & C. D. Frith, 2006; U. Frith & C. D. Frith, 2001; C. D. Frith & U. Frith, 2006), theory of mind (Allison, Puce, & McCarthy, 2000), mental-state reasoning (Birch & Bloom, 2004), and mind-reading (Baron-Cohen, Jolliffe, Mortimore, & Robertson 1997). Though these terms may refer to different perspectives about decoding others' mental states, the basic premise of these terms is the same. We use information in our environment, such as others’ behavior, expressions, and situational cues to understand their mental experiences.

Typically developed humans tend to understand that other humans are capable of independent feelings, desires, or experiences than their own. The ability to understand others’ mental states such as these develops in infancy, with evidence as early as between 9 and 15 months after birth (Carpenter, Nagell, & Tomasello, 1998; U. Frith & C. D. Frith, 2003). Throughout infancy, humans gradually develop the ability to know that others have different perspectives, different sets of knowledge, and can feel different emotions than one feels one’s self.
Mentalizing develops into a powerful psychological process and even goes beyond processing other humans, as we tend to anthropomorphize nonhuman beings to assume that they also have mental states. We read mental states into the faces and actions of animate beings, such as animals (Gosling, Kwan, & John, 2003; Kwan, Gosling, & John, 2008; Zebrowitz et al., 2011), and even into the actions of inanimate objects. In a film displaying two triangles and a circle moving around a box, Heider and Simmel (1944) showed that participants created elaborate stories to explain the shapes' movement in terms of human attributes of behavior. Participants described the large triangle as “chasing” the small triangle and the circle, implying the movement of the arbitrary shapes had intent, even though no social cues were present.

We can read the mental states of others from several sources of nonverbal information. We can read emotional states from others’ posture and movements (de Gelder, 2006). We can also understand where their attention is based on movements such as looking and pointing (Langton, Watt, & Bruce, 2000). Nonverbal vocal cues also can convey what another's mental state is, as we can glean information about mental processes from pitch and loudness of speech (Hornak, Rolls, & Wade, 1996; Scherer & Wallbott, 1994).

When determining what someone else is thinking or feeling, however, none of these cues is as important or as informative as the face (C. D. Frith & U. Frith, 2006). The face has complex and coordinated musculature that is argued to exist primarily for the purpose of communicating with others (e.g, Huber, 1931). Indeed, we can reliably and consistently read other people’s basic emotional (Ekman & Friesen, 1971) and complex mental states (Baron-Cohen, Wheelwright, Raste, et al., 2001) based on subtle cues varying within these muscles. Eye gaze also gives clues about what another is thinking. We can detect where people are looking from the eyes (and the orientation of the head) and are sensitive to even millimeter shifts in pupil location (Langton,
Watt, & Bruce, 2000). Further, changes in eye gaze give us cues as to another’s emotional state (e.g., Calder et al., 2002) and influences how we detect emotions from others (Adams & Kleck, 2003, 2005; Adams et al., 2011). Though the face and eyes make up a small part of the human body, their subtle behaviors allow others an unparalleled peek into the processes of our minds.

**Mechanisms of Mentalizing**

The ability to mentalize is extremely complex, requiring a variety of different cognitive resources. Different theories explain our ability to mentalize based on different cognitive and neural mechanisms. However, what is common about each of these explanations is that each asserts just how important the face is in mentalizing.

**Simulation Theory and Mirror Neurons.**

A Native American proverb asserts that humans cannot understand others until they “walk a mile in their moccasins.” Teachings of the same vein exist across cultures and throughout much of history. For centuries, philosophers have asserted that our ability to simulate the experiences of others is an essential part of understanding what they think or feel. For instance, David Hume (1739/1888) asserted that “the minds of men are mirrors to one another.” (p. 365). Among others, Adam Smith and Immanuel Kant expressed similar sentiments in the 18th century (Goldman, 2006). Despite the long history of simulation in philosophy, it is only recently that simulation theory has existed within psychology as an explanation for our ability to mentalize about others.

Simulation theories assert that we mentalize about others by having a mental representation of what another person is experiencing and then we use that representation as a basis to derive what they are thinking (Gallese & Goldman, 1998). Thus, simulation requires the ability to perceive another's situation and then mentally represent that situation based on those
perceptions. Then we use those representations as the basis for understanding what another is thinking. Thus, we are able to then use ourselves and our understandings of how we feel in similar situations as a base, and as Smith asserts in the above quote, “imagine” what another must be feeling in that situation. Importantly, simulation theories assume that we can do this separately from our own mental states, so that a person can simulate another's mental state without actually feeling that state (Goldman, 2006).

The ability to simulate others requires us to be able to mentally represent another’s situation. Indeed, as Hume asserts, our minds are mirrors to one another. Since Lipps (1906), theorists have suggested that we naturally and subconsciously mimic others’ behaviors and that this is important in recognizing what they are thinking. Babies naturally mimic adults’ facial expressions (Termine & Izard, 1998). We mimic others’ behaviors, postures, and even yawns (e.g., Bavelas, Black, Chovil, Lemery, & Mullett, 1988; Bernieri, 1988; Chartrand & Bargh, 1999; Provine, 1986). This mimicry is important for understanding what others are feeling, as mimicry speeds recognition of emotions (Stel & van Knippenberg, 2008). This ability to mimic others suggests that our perceptions may be unconsciously coupled to behavioral representations of our actions (Chartrand, Maddux, & Lakin, 2005).

The putative human mirror neuron system provides a neural mechanism for mimicry and simulation (Goldman, 2006). Mirror neurons are populations of neurons that respond similarly when one is performing an action and when one is perceiving another performing that same action. Single-cell recordings in the premotor cortex and inferior parietal lobule of monkeys (area F5) respond both when performing certain motions as well as perceiving others' performing the same motions (Rizzolatti, Fogassi, & Gallese, 2001). Humans also show evidence of a mirror neuron system (MNS) in the premotor cortex and posterior parietal regions, which are activated
by perceiving the motions of others and when planning to perform the same motions oneself (Hari et al., 1998; Iacoboni et al., 1999). In humans, mirror neurons are thought to be part of an elaborate system of imitation that involves the posterior inferior frontal gyrus (IFG), premotor cortex, inferior parietal lobule, and posterior STS (Iacoboni & Dapretto, 2006).

Though mirror neurons are present in several brain regions, the most often researched is within the premotor cortex, the posterior IFG. This area, including Brodmann area 44, is thought to be a homolog of monkey area F5, or the mirror system isolated in monkeys (Geyer, Matelli, Luppino, & Zilles, 2000; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). The human IFG is also involved in mirroring others’ responses (e.g. Buccino et al., 2001; Grezes, Costes, & Decety, 1998; Iacoboni et al., 1999). Indeed, a suppression paradigm, which uses fMRI to examine the reduction of a neural response in a specific population of neurons when those neurons are repeatedly activated by a specific task, showed overlap in the suppression of neural responses in the IFG for perceiving an action and performing that same action, supporting the contention that the same neurons were active in both tasks (Kilner, Neal, Weiskoph, Friston, & Frith, 2009). This again supports the contention of mirror neurons in the human IFG.

Only indirect evidence implicates the MNS in reading others' mental states. However, several studies indicate a role of the IFG in mentalizing and mimicry. Reading others' mental states from facial cues leads to activation in the IFG (Adams et al., 2010; Schulte-Rüther, Markowitsch, Fink, & Piefke, 2007). Imitating and perceiving emotional faces elicits activity in the IFG (Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003; van der Gaag, Minderaa, & Keysers, 2007). Further, those with higher trait levels of empathy show more left IFG activity when perceiving dynamic displays of facial emotions (Chakrabarti, Bullmore, & Baron-Cohen, 2006) as well as when perceiving others’ motor movements (Kaplan & Iacoboni, 2006). Despite
this evidence for the existence of the MNS and its role in understanding others, the role of the MNS in mentalizing is unclear. Studies that have implicated the MNS in empathy have not localized MNS-specific neurons in the IFG, as the IFG contains many neurons that are not thought to be part of the MNS (Turella, Pierno, Tubaldi, & Castiello, 2009). Therefore, since the IFG is involved in other tasks that may be related to mentalizing but are not related to the MNS, such as task switching and cognitive control, it is still uncertain if the MNS is involved in mentalizing.

**Theory-theory**

The types of errors people make in mentalizing do not support the notion that simulation itself is sufficient to explain what others are thinking (Saxe, 2005). For example, as children develop in mentalizing ability, they make predictable errors in determining others' mental states (such as assuming that ignorance, or “not knowing” is equivalent to “being wrong”). These errors are consistent biases toward specific mentalizing mistakes, which suggests that instead of being random errors, that these errors are based on incorrect presumptions or theories as to how others act. Thus, this suggests that some of mentalizing is based on theories of how others act. Further, the perceptual representations that are used as a basis for simulating others are unable themselves to represent the stable attitudes about others that we use to guide mentalizing (Mitchell, Banaji, & Macrae, 2006). Supporting this, a meta-analysis of fMRI studies on understanding the intentions of others found no involvement of the mirror system in understanding others in the absence of perceptions of other’s behaviors, but did find evidence of the MNS in understanding others based on perceptual cues (Overwalle & Baetens, 2009). Further, activation in MNS regions is not sufficient to account for mentalizing about perceptions
of others’ actions, as understanding these actions utilizes a large network of other brain regions not implicated in the MNS (Spunt, Satpute & Lieberman, 2011). This evidence indicates a more cognitive-based mechanism exists for understanding the mental states of others that may operate alongside simulation in how we mentalize.

When it comes to understanding the world, we create folk theories about life based on our own experiences. Theory-theory, or folk psychology, extends this idea to how we understand others, asserting that we use our own theories to understand what is in others’ minds. At a very basic level, we create our own folk theories to try to explain others’ behavior based on sets of rules we learn (Carruthers & Smith, 1996). In this perspective, we assume unseen mental states or goals drive others’ behavior, and we use our knowledge about others in general and our observations about the behavior of specific others in order to determine what their mental states or goals are.

Though debate exists between simulation theory and theory-theory as to which one explains how we mentalize about others, it is likely that mentalizing relies on both of these processes or a hybrid of both of these processes (Apperly, 2008; Carruthers & Smith, 1996). Understanding others likely combines our own representation of others along with the person knowledge that we store about others. Keysers & Gazzola (2007) proposed a model integrating simulation and theory accounts of mentalizing by dividing between the representations of others based on the knowledge of others and the reflections of others based on how we would simulate their mental states. Both of these types of knowledge are combined to then understand another’s mental state.
Neural Mechanisms for Mentalizing.

Several brain regions are implicated in mentalizing. In addition to the MNS in the inferior frontal gyrus discussed above, the medial prefrontal cortex (mPFC), precuneus and posterior cingulate (PCC), superior temporal sulcus (STS) and temporoparietal junction (TPJ) are active in mentalizing tasks.

Regions of the medial prefrontal cortex (mPFC) are consistently implicated in mentalizing tasks. These regions are thought to be the basis of more theory-based cognitive approaches of processing others’ mental states and are also involved in social cognition tasks more broadly (Uddin, Iacoboni, Lange, & Keenan, 2007). Different regions of the mPFC are involved in different aspects of mentalizing, as shown by dissociations among various mentalizing tasks. For instance, ventral regions of the mPFC (vmPFC) are involved in understanding similar others (Mitchell et al., 2006) as well as emotional perspective taking (Hynes, Baird, & Grafton, 2006). The dorsal mPFC (dmPFC), however, is thought to involve using stereotypical knowledge about others to understand what they are thinking. Supporting this, the dmPFC is involved in computationally determining others’ beliefs during games that involve knowing another’s strategy (Hampton, Bossaerts, & O’Doherty, 2008; Yoshida, Seymour, Friston, & Dolan, 2010).

This evidence suggests that the dmPFC might be a neural basis of the person knowledge required for theory-theory accounts (Keysers & Gazzola, 2007). Such a dissociation has led to the proposal that the vmPFC is involved with decoding the emotions of others while more dmPFC is involved with understanding meaning through the actions of others (Amodio & C. D. Frith, 2006). Thus it is possible the dmPFC contributes more to theory-based accounts of mentalizing while the vmPFC may be linked to activation in the MNS for simulation-based
mentalizing, as evidence suggests greater simulation in mentalizing about those who are most similar to us (Beeney, Franklin, Levy, & Adams, 2011; Preston & de Waal, 2002).

The precuneus and posterior cingulate (PCC) is also implicated in understanding others. Several studies involving thinking from others’ perspectives elicit precuneus and PCC activation (e.g. Pfeifer, Lieberman, & Dapretto, 2007; Ruby & Decety, 2001). Specifically judging what another person thinks about a third party (such as the Queen of England) versus judging what one thinks themselves about a third party elicited greater PCC activation as well, indicating the role of the PCC in mentalizing specifically (Lombardo et al., 2009). The role of the PCC in mentalizing is not well-understood but may involve the importance of this region in memory retrieval processes that are important in mentalizing (Cavanna & Trimble, 2006). This suggests a role of the PCC in retrieving information that are important for understanding others from a theory-theory perspective.

The temporoparietal junction (TPJ) is also involved in inferring others’ mental states. The right TPJ and to a lesser extent the left TPJ is thought to be involved in representing the thoughts or beliefs of others (Saxe & Kanwisher, 2003). For instance, when reading stories containing information about a person’s beliefs, the right TPJ was active more than when reading stories containing other information about a person that was not related to belief (Saxe & Powell, 2006). Indeed, temporary lesions of the right TPJ induced by transcranial magnetic stimulation (TMS) reduce the degree to which participants take into account a person’s belief when making a moral judgment about that person (L. Young, Camprodon, Hauser, Pascual-Leone, & Saxe, 2010). Activation in the TPJ overlaps activation in various attentional cuing tasks, which suggests specifically that the TPJ is involved most critically in computing the orientation and direction of behavior (Overwalle & Baetens, 2009).
The superior temporal sulcus (STS) is also involved in understanding the mental states of others. Like the TPJ, several studies implicate the STS in processing the goals of others. Activations in the posterior STS are also commonly found in tasks using mentalizing and indeed overlap regions in the TPJ discussed above (e.g. Saxe & Kanwischer, 2003). Specifically, the STS is more active in viewing goal-directed actions in others, such as reaching for an object, than non-goal directed actions (Pelphrey, Morris, & McCarthy, 2004). Likewise, attending to the goal-directed actions of one shape “chasing” another shape increased activations in the STS than just merely watching the shapes (Schultz, Imamizu, Kawato, & Frith, 2004). Switching from one goal-related activity to another also elicits activation in the STS (Zacks et al., 2001).

**Social-perceptual Mentalizing and the Face**

A variety of different ways exist for us to gather the information necessary about another in order to make a judgment about what they are thinking. We can gather information about others’ mental states based on stories or other descriptions of their behavior and actions. We can also gain insight into others mental states by reading emotions or complex mental states from their nonverbal expressions. Among these perceptual cues, the face provides several sources of information about another that is relevant to mentalizing about them. It provides information that is related to person knowledge, such as another’s age, race, and sex, as well as knowledge about one’s emotional state. All of this is used in mentalizing about another.

Tager-Flusberg and Sullivan (2000) proposed that mentalizing ability could be distinguished between social-cognitive mentalizing, which involves mentalizing about others based on cognitive information about others’ actions and behaviors, and social-perceptual mentalizing, which involves mentalizing about others based on perceptual information. Thus,
social-perceptual mentalizing is related to decoding others’ affective states based on facial or body cues. This dissociation was based on work understanding mentalizing in those with Williams Syndrome, a rare inherited disease that causes mild to moderate mental retardation, yet leaving some social skills intact. Adults with Williams Syndrome are impaired when trying to mentalize based on cognitive information in traditional Theory of Mind tasks, yet show no deficits when performing tasks that involve mentalizing based on perceptual cues, such as the eye region of faces (Tager-Flusberg, Boshart, & Baron-Cohen, 1998).

The behavioral dissociation between social-cognitive and social-perceptual mentalizing suggests that mentalizing about faces may have some different underlying neural networks than mentalizing using cognitive information. Indeed, mentalizing using facial information utilizes additional regions that are involved with face processing in addition to those implicated in mentalizing.

In addition to decoding the goals of others, the STS is consistently involved with decoding others’ emotions from facial cues. Neural models of face perception speculate the STS is involved with decoding changeable aspects of the face, especially emotion (Haxby, Hoffman, & Gobbini, 2000). Studies using local field potentials in monkeys (e.g., Hasselmo, Rolls, & Baylis, 1989) and human studies (e.g. Narumoto, Okada, Sadato, Fukui, & Yonekura, 2001) indicate the STS is preferentially active for facial emotion displays. The STS is also active in processing the direction of another’s eye gaze, which is also important for understanding a person’s mental state (Haxby, Hoffman, & Gobbini 2002; Puce, Allison, Bentin, Gore, & McCarthy, 1998). Activation of the posterior STS to eye gaze is thought to be related to understanding another’s intention (e.g., Nummenmaa & Calder, 2009), which is similar to the
research reviewed above that implicated the posterior STS and adjacent TPJ in understanding intention.

The amygdala is another region that is involved in understanding others’ mental states based on facial cues. Tager-Flusberg and Sullivan (2000) assert that the amygdala is of critical importance in social-perceptual mentalizing. Supporting this assertion is extensive evidence showing the amygdala is important in understanding social cues. Lesions to the amygdala impair the ability to read social situations. Additionally, those with amygdala lesions fail to anthropomorphize the shapes in the Heider and Simmel task in the same way as those with intact amygdalae (Heberlein & Adolphs, 2004). Most of the research as to the purpose of the amygdala in face processing implicates it in regard to processing fear. The amygdala is differentially engaged by threatening stimuli and is critical for evaluating and understanding threat in the environment (Phelps, 2006). The amygdala is important and active in processing a variety of different facial emotions (e.g. Phan, Wager, Taylor, & Liberonz, 2002); however it is most active in processing fear. Amygdala lesions specifically impair detecting fear in the environment (Adolphs et al., 1999).

More recent research indicates the amygdala may have an even more broad-based role in mentalizing. Whalen (1998) asserts the reason the amygdala is most strongly involved in processing fear, especially from faces, is that the amygdala is involved in processing ambiguity surrounding the source of a threat and fear, at least when coupled with direct eye gaze, is an ambiguous emotion, because it only indicates the presence of a threat and does not indicate where the threat is. Anger coupled with direct gaze is a clear threat, revealing the source of threat and where it is directed (at you). Supporting this, when eye gaze is manipulated to make angry and fearful expressions either more or less threatening, the amygdala reacts most strongly to
averted gaze angry faces and direct gaze fear faces, as these are ambiguous threat-gaze signals (Adams, Gordon, Baird, Ambady, & Kleck, 2003).

Additional research suggests that the amygdala may be involved in processing different facial messages in different stages of processing. This work builds on research showing a dissociation between rapid processing of clear representations of threat and more sustained processing of ambiguous representations of threat (e.g., LeDoux, 1996). When presented quickly (33 or 300 milliseconds), the amygdala is more active for clear presentations of threat (averted eye gaze fear) while when presented for longer periods (1 or 2 seconds), the amygdala is more active for ambiguous displays of facial threat (Adams, Franklin, Kveraga, et al., 2011; Adams, Franklin, Nelson, et al., 2011).

**Summary**

Understanding the mental states of others is a unique mental challenge for humans, yet we are very adept at understanding that others have minds and that the processes within those minds may be totally different than our own processes. A dedicated neural system underlies this ability, including the mPFC, STS, TPJ, precuneus, and posterior cingulate cortex. That we are able to understand people so well and such a network exists for this underscores just how important detecting others’ thoughts and feelings is in social interaction.

Some evidence suggests that mentalizing using perceptual cues, especially involving the face, may be to some degree distinct than networks involved with mentalizing based on cognitive information about another. This social-perceptual mentalizing network includes regions known to be involved in decoding emotions, such as the STS and the amygdala, as well as the IFG, which includes mirror neurons. This social-perceptual network of mentalizing indicates just how
important the face is in understanding others, especially due to the emotional content that a face displays.

**Processing and Remembering Faces**

“There is in every human countenance either a history or a prophecy, which must sadden, or at least soften, every reflecting observer” – Samuel Taylor Coleridge

Our social environment plays an important role in the basic operation of our cognitive processes. This is the case for memory, as well as in perception. Even though memory at its most basic level does not need social context, our social environment affects how we remember the world. The collective experience of those around us shapes our memories, and as such, our very identity (Barnier, Sutton, Harris, & Wilson, 2008). Social factors impact memory, but our social behaviors also require a memory system too. We need memory to function socially. We have to remember complex social scripts for various situations, who others are, and sets of information about each individual we know to build relationships.

As the face is central as a social stimulus, the face plays a special role in social memory. We recognize and remember others efficiently and easily from facial cues (Duchaine & Nakayama, 2006). This ability to recognize faces so well has led to extensive debate if faces are innately “special” or if we are merely expert decoders of faces (Gauthier, Skudlarski, Gore, & Anderson, 2000; Kanwisher, McDermott, & Chun, 1997). Regardless of whether faces require unique neural architecture to process, they are nevertheless a unique social stimulus and we possess a unique ability to remember them.
Many studies have explored the mechanisms by which we remember faces, including how we recognize them, encode them, store faces in memory, and how we retrieve them. In the section below, I detail the cognitive and neural mechanisms that are thought to be part of decoding identity from faces and remembering faces. I also detail the social factors that influence face memory. A long tradition in social psychology on social factors that moderate face memory has existed separately from the cognitive and neuroscience research on face memory. Only recently have these two traditions of research been brought together.

**Cognitive Models of Face Memory**

Our ability to encode, remember, and recognize so many faces so well is a computational feat that is difficult to explain if we consider faces to be remembered in the same way as other visual stimuli. The visual differences that we use to distinguish between faces are extremely small. In fact, we can notice a variation of only 7 percent of the features of a face as a different identity (Lehky, 2000). Faces can also be easily recognized easily even with different visual presentations. We are able to recognize and remember faces shown from different angles (such as faces shown from profile or frontal views), with partial occlusions, and under conditions of high visual noise (Duchaine & Nakayama, 2006). All of this evidence suggests that the human ability to recognize and remember faces is more advanced than other visual stimuli.

In addition, some evidence suggests that faces are processed holistically; that is, as being represented in memory as a whole unit, rather than the separate parts of a face being remembered separately (Farah, Wilson, Drain, & Tanaka 1998). Configural information, the relationship between facial features, is extremely important in remembering faces. We are attuned to configurally process faces by quickly detecting the first-order relationships that define the presence of a face, such as two eyes over a nose and a mouth as well as processing the second-
order configural differences that distinguish between different faces, such as the distance between facial features (Maurer, Le Grand, & Mondloch, 2002). Supporting this, we are much better at discriminating and remembering faces presented upright than inverted (Yin, 1969) and manipulations that alter the relationship of a whole face, such combining the top half of a face with the bottom half of another face, interfere with how faces are recognized (A. Young, Hellawell, & Hay, 1987). Further, it is more difficult to recognize deviations in specific features of a face (such as determining if a nose is a different shape) when in the context of the whole of a face than when presented alone, or outside of the context of a face (Tanaka & Farah, 1993). These studies support the idea that faces are recognized as wholes and that we have a special cognitive mechanism for processing faces, regardless if such a model is inherently special for faces or a result of our extensive expertise with faces.

The multidimensional model of encoding faces in face-space is the main cognitive model used to explain just how we have such an amazing ability to recognize faces (Valentine, 1991). The face-space model is a model that describes faces as encoded as a point in multidimensional space, where each axis or dimension represents a different attribute on which faces vary perceptually. By encoding faces on their values on these dimensions, we are able to store many faces in memory in an efficient way, rather than remembering the features of each face as distinct items in memory. Further, such a model allows for fine-tuned recognition of new stimuli, as stimuli that are perceived can be explicitly compared on these dimensions to other previously-encoded items. However, such a model requires many initial inputs, as one must learn both which dimensions to use to encode items in this multidimensional space and one must have enough exemplars in order to make a dedicated model worthwhile.
Multidimensional face-space is a powerful model to explain how we can recognize faces, but within this framework there are two different ways that we can code identity. The first of these hypothesizes that faces are coded as points in an exemplar-based network, which encodes facial identity differences based on parameters uniquely defined by each individual face (Valentine, 2001). In this case, each dimension is coded as an absolute value that does not vary as a function of other faces. The other explanation is norm-based coding, which codes each dimension as norms in terms of differences from the prototypical or average face (Leopold, O’Toole, Vetter, & Blanz, 2001). In this model, norms represent dimensions in the multidimensional space framework and the face is coded on each norm by virtue of how much it differs from a prototypical face, or the average of all the faces in a person’s face-space. Norm-based coding is common in low-level visual processing, as neurons responsive for detecting brightness and motion adapt to average values in the environment and are maximally sensitive to deviations to this average (Clifford & Rhodes, 2005).

Extensive evidence supports the norm-based network for coding faces. Caricatures of faces, which deviate from an average face, are easier to identify than anti-caricatures, which are faces manipulated to be closer to average (Lee, Byatt, & Rhodes, 2000). By exposure, a person’s prototype in face-space can be changed as well. Repeatedly showing a series of faces that deviate in a specific way can change one’s perceptual average and change how well other faces can be recognized (Leopold, Rhodes, Müller, & Jeffery, 2005). Adapting a person to a specific face by repeated exposure facilitates the recognition of the computationally opposite face (Leopold, O’Toole, Vetter, & Blanz, 2001), an effect that is selective for this opposite face (Rhodes & Jeffery, 2006). None of these effects can be explained by exemplar-based coding network.
This face-space model also explains how faces are recognized in this model, specifically how perceived faces are matched with those encoded in identity. One model of recognition in face-space is the attractor field model (Tanaka, Giles, Kremen, & Simon, 1998; Tank & Hopfield, 1987). In this model, each face represents an encoded point in multidimensional space. Surrounding this point in multidimensional space is an associated attractor field, which is then used in order to recognize newly perceived faces. This face-space is used to judge if newly-perceived faces match faces in memory and if a perceived face is similar in enough dimensions to a specific encoded face in memory, it becomes part of that attractor field and is recognized as that face. This accounts for how we can recognize faces given such different perceptual representations of a face, such as seeing a person from profile versus frontal angles, seeing faces with glasses, hats, or other occlusions, seeing faces in different lighting conditions, or seeing faces with different perceptual changes such as showing different emotions or after aging-related changes. These different perceptual representations of a face should only affect some of the dimensions of a face, leaving enough resemblance on other dimensions of a face in order for a perceived face to still be coded in the attractor field of the memory representation of that face.

Attractor fields provide a model of how faces are encoded and remembered in face-space, yet little work has examined which dimensions are actually coded in face-space and how these dimensions and related attractor fields form. Corneille, Hugenberg, & Potter (2007) suggest that motivational influences may affect how faces are encoded in face-space, which suggests that the degree to which participants are motivated to encode a face affects how it is encoded in face-space. Supporting this, when they made pairings of races and emotions to be either stereotypic (i.e., angry Black face) or counter to stereotypes (i.e. fearful Black face), they found the more ambiguous, counter-stereotypic pairings were more easily discriminated as a separate identity,
purportedly because the expression took the face away from the stereotype in face-space. Interestingly, these findings of interactions between racial category and expression may work in other ways, as another study examining face memory found better memory for angry Black faces than neutral Black faces (Ackerman et al., 2006). They argue in this case that angry Black faces represent angry outgroup members and thus it is more functionally relevant for survival to remember angry outgroup members. Therefore, it is possible that other moderators, such as the interaction between racial category stereotypes, the overlap of facial cues associated with different races with perceptual features, and the cues involved with emotion affect the interaction between race and emotion in face space, but both of these studies show that face-space itself may be moderated by cues such as emotion or racial category.

Perceptual encoding of faces in face-space using attractor fields provides a cognitive model of how faces can be remembered. It also suggests that factors such as emotion can interact with how faces are remembered by changing how a face is encoded in face-space. This suggests a plausible mechanism of how mentalizing as a cognitive process can interact with face memory, which I discuss below.

**Neural Models of Face Memory**

Face memory engages a widespread network of brain regions that are involved in general memory processes that interact with brain regions that are specifically disposed to remember faces. Central in memory processes across different stimulus modalities are regions in the medial temporal lobe. The hippocampus and neighboring structures contribute to recognition and associative memory, as well as recollection and familiarity judgments (Mayes, Montaldi, & Migo, 2007; Simons & Spiers, 2003). Hippocampal lesions impair the ability to form memories in a variety of contexts (e.g., Corkin, 1965; Knight, 1996; Scoville & Milner, 1957). Further,
spontaneous hippocampal activation in a semantic classification task moderates memory in an upcoming surprise memory test (Dennis, Daselaar, & Cabeza, 2007). In contrast, the perirhinal cortex may be involved in familiarity-based recognition of stimuli, especially those that are visual in nature (Bussey, Saksida, & Murray, 2002).

These medial temporal lobe structures are linked with memory processing in the prefrontal cortex. The dorsolateral prefrontal cortex (DLPFC) is important in working memory and manipulating memory traces, especially in maintaining associations (Murray & Ranganath, 2007). Thus, the DLPFC is critical in encoding relations between items (Blumenfeld, Parks, Yonelinas, & Ranganath, 2011; Blumenfeld & Ranganath, 2006). Further, the DLPFC is found to be involved in successful encoding of emotional facial expressions, leading to the contention that this region is involved in integrating emotional memory (Sergerie, Lepage, & Armony, 2005), though the amygdala is more often implicated in emotional memory (see below). In contrast, the ventrolateral prefrontal cortex (VLPFC) is important in the maintenance of memory, especially for item-specific memory (Badre & Wagner, 2004; Simons & Spiers, 2003).

Regions of the parietal cortex are involved in memory, especially in retrieval processes (e.g., Cabeza, Ciaramelli, Olson,& Moscovitch, 2008; Vilberg & Rugg, 2008). Parietal activation in memory processing may reflect attentional processes critical in memory retrieval (Majerus et al., 2006), specifically with superior regions of the parietal lobe modulating top-down attentional allocation to retrieval while inferior regions are involved in bottom-up attentional capture by memory contents that are retrieved (Ciaramelli, Grady, & Moscovitch, 2008). Further, the inferior regions of the parietal lobe are involved in maintaining attention on memory contents that are held for retrieval (Vilberg & Rugg, 2009). The insula and anterior cingulate are also important in memory retrieval, as the insula (Paulus, Feinstein, Leland, & Simmons, 2005) and
anterior cingulate (Fleck, Daselaar, Dobbins, & Cabeza, 2006) may contribute to decision making processes in memory retrieval.

Brain regions are also involved in face memory due to their contribution to processing faces rather than memory more generally. The posterior fusiform gyrus (FG) is consistently implicated in face processing, so much that it is often referred to as the “fusiform face area” (Kanwisher et al., 1997). FG activity is stronger for encoding faces that deviate farther from the prototypical face, reflecting its role in the neural representation of face space (Loffler, Yourganov, Wilkinson, & Wison, 2005). Activity in the FG reflects encoding both feature-based and configural processing of faces (Harris & Aguirre, 2008). The FG also moderates face memory, especially during encoding processes and successful recognition (Zeineh et al., 2003). This also extends to social differences in memory. Differential activation of the FG moderates the own-race bias (Golby, Gabrieli, Chiao, & Eberhardt, 2001). Participants showed better memory for faces of their own race and this advantage in memory correlated with differences in fusiform activation between faces of their own race and another race.

Interestingly, little research has linked brain regions involved with mentalizing, such as the STS and VMPFC, to face memory. However, some evidence suggests that activity in regions involved in face memory and mentalizing is related, as face-specific areas in the FG and STS are functionally connected (Turk-Browne, Norman-Haignere, & McCarthy, 2010). Additionally, mental state information contained within a face can affect brain activation that is linked to memory. For instance, smiling faces elicit activity in the orbitofrontal cortex (OFC), which is linked to better memory for those same faces (Tsukiura & Cabeza, 2008).

Likewise, the amygdala is involved in both processing social information from faces and also in emotional memory (Phelps, 2006). Lesions to the amygdala cause deficits in forming
emotional memories (Adolphs, Tranel, & Buchanan, 2005; Labar & Phelps, 1998). The role of the amygdala is thought to be due to the amygdala’s involvement in processing emotionally-valenced stimuli and its connections between the amygdala and hippocampus, regions which border each other (Dolcos, LaBar, & Cabeza, 2004; Richardson, Strange, & Dolan, 2004). Thus, it is thought the amygdala is important in modulating what is remembered in the hippocampus and in the prefrontal cortex.

The involvement of the OFC and amygdala in socially-relevant memory process has led to the suggestion that the OFC, amygdala, and hippocampus are a network disposed for social memory (LoPresti et al., 2008). As I discuss in the following section, research is beginning to explore how these regions are involved in memory for stimuli with social meanings, such as faces, and how this, along with other brain regions involved in mentalizing (such as the VMPFC and STS) may provide the basis for an underlying neural relationship between mentalizing and face memory.

**Mentalizing and Memory**

In the above sections, I reviewed the separate and extensive research that has examined mentalizing and memory respectively. Despite the long tradition examining each of these constructs, very little research has investigated if a direct relationship exists between these two processes. Nonetheless, some research has examined how memory can affect mentalizing, specifically examining how autobiographical memories and development of the sense of self and self-aware memories lead to mentalizing (e.g. Gallagher & C. Frith, 2003). Supporting this, many of the same brain regions involved with mentalizing are also involved with self-memories, (such as the mPFC and temporal poles; Maguire, 2001). In this chapter, I examine research that
links mentalizing and memory, reviewing literature suggesting how memory processes can influence mentalizing and the research that suggests a link between mentalizing and memory.

**Memory Affecting Mentalizing**

**Simulation and the Prospective Brain.**

Though it is not often discussed as being important in mentalizing, memory is a prerequisite to a complete understanding of another’s mental state in both a simulation and a theory-theory mentalizing perspective. Simulation requires a person to neurally represent another’s situation and perspective and then draw upon that representation to understand what another person is thinking. Thus, simulation requires us to use our own experiences in order to understand another (e.g. Goldman, 2006). Once we have representations of others that we have simulated, we need to use our own experiences to know what we would feel in that case, and then use that understanding to understand another’s feeling. Therefore, the memory processes that modulate autobiographical memory may then affect how we mentalize about another.

Recent research indicates that much of the ability to imagine and simulate the future is based on the same neural mechanisms that are involved with remembering the past (the default mode network; Schacter, Addis, & Buckner, 2007). Research examining this neural network has mostly investigated the link between remembering past episodic events and simulating potential future episodes (Addis, Wong, & Schacter, 2007; Szpunar, Watson, & McDermott, 2007). This does not necessarily imply simulation involved with mentalizing about others. However, the brain regions involved with what is termed the prospective brain involve many of the same regions that are involved in mentalizing, including the vmPFC and the TPJ (Spreng, Mar, & Kim, 2009) along with regions involved with memory such as the medial temporal lobe.
Though research on the prospective brain and simulating future episodes has not examined mentalizing per se, many of the same processes involved with simulating future events can be drawn to simulating mental states. Further the same default mode network is involved in mentalizing as well (including the vmPFC and posterior cingulate cortex). The mechanisms by which memory is used in simulating potential future events may thus be related to how memory is used in simulating others’ mental states. Little work has examined the role of memory in simulating others’ mental states, but some models of empathy have suggested the necessity of remembering our own experiences in feeling empathy for others (Preston & de Waal, 2002). Empathy is often defined as the ability to share others’ mental states and though this is a distinct construct from mentalizing about others’ mental states, empathy necessarily entails understanding another’s mental state, in order to share it. This research suggests that sharing experiences with others leads us to have greater empathy for them. Supporting this, deeper friendships lead to greater neural responses of empathy for social rejection, an effect which is mediated by activation in the MNS (Beeney et al., 2011). If sharing experiences does lead to greater empathy, then episodic memory may be contributing to how we empathize with others, possibly by leading to deeper simulation for others’ mental states.

**Person Knowledge and Stereotypes**

The knowledge we learn about others also affects how we mentalize about them. Our memory of our own previous experiences and the things we have learned allow us to have a rich understanding of what others may be thinking or feeling. In order to mentalize about others using a theory-based or folk psychology system, we have to develop theories we have about others’ mental processes. We develop these theories based on our experiences with others, using the situations and actions of others to create theories of others’ mental processes. Memory is
required in order to remember these experiences as well as these theories themselves that we then use to understand others. Thus, this requires a relationship between the neural mechanisms involved with mentalizing and memory processes.

Knowledge about others may be stored and processed like other memories, but some evidence indicates that person knowledge may be its own separate system of knowledge. Lesion studies indicate a dissociation between knowledge of animate and inanimate beings (e.g. Warrington & Shallice 1984; Caramazza & Shelton, 1998). Likewise, person knowledge may have separate neural substrates than knowledge about dogs, as the former involves more activation of frontal regions while knowledge of dogs involves more activation of medial temporal regions (Mason, Banfield, & Macrae, 2004). This suggests social knowledge about others may have its own specialized memory systems.

Stereotypes represent another form of learned knowledge that is used to mentalize about others. Stereotypes can be defined as a set of qualities that one associates with a certain group of people (Schneider, 2004). These qualities include both affective and cognitive sets of information about others (Fiske & Taylor, 1984). Stereotypes are learned sets of information about groups of others that affect how we attribute mental states to others. Stereotypes are not always accurate and are often completely inaccurate in understanding others, but they can be a useful way of simplifying the set of person knowledge we have about others.

Several learned sets of stereotypes affect how we mentalize about others. For instance, gender stereotypes exist for differences in women and men experiencing and expressing emotion. Women are thought to feel more joy, sadness, and fear than men and are expected to express these emotions in social settings (Shields, 2002). Conversely, men are thought to experience more anger and are thus expected to express it more often. These gender stereotypes
also affect how we interpret the mental states of others based on how we perceive them. For instance, the mental states ascribed to a person crying varies, depending on if that person is a woman or a man (Warner & Shields, 2007).

Using stereotypical information to mentalize about others also may use separate memory systems than individuating others, as stereotypes rely on semantic sets of memory while individuating others uses more episodic memory, which involves deeper cognitive processing (Sherman, 1996; Sherman & Bessenoff, 1999). Faces that are associated with knowledge-based personal information, in addition to what is gathered perceptually from the face, are encoded in different ways that faces that are not associated with knowledge (Cloutier, Kelley, & Heatherton, 2010). When person knowledge is available, faces are encoded in a different way, presumably reflecting the presence of this individual knowledge. Also, using learned stereotypical knowledge also induces activation in different brain regions than using non-stereotypical knowledge about others, namely the right frontal cortex (Mitchell, Ames, Jenkins, & Banaji, 2009). Activity in the right frontal cortex areas may reflect semantic retrieval of categorical knowledge that is then used for mentalizing about others.

**Mentalizing Affecting Memory**

Despite the evidence indicating the necessity of memory for mentalizing, very little research has examined how mentalizing about others may influence memory. In the section above, I detailed the various ways that memory is necessary for understanding others. This indicates a link between the two processes. In this dissertation, I assert that mentalizing actively influences memory. Those who we mentalize more extensively about are those we remember better. Very little research has directly considered this question, but several indirect lines of
evidence indicate that mentalizing might be a moderator of memory as several effects that moderate mentalizing moderate memory as well.

**Evidence for a Relationship**

**Parallel Ingroup Biases**

Memory and mentalizing are both affected by group membership and show evidence of biases for ingroup faces. A large literature details the existence of the own-race bias in face memory. This is a consistent bias where we have better memory for faces of members of one’s own race (Malpass & Kravitz, 1969; Meissner & Brigham, 2001). Specifically, we tend to have a more liberal response bias toward faces of other races, with more frequent false alarms for other-race faces (Sporer, 2001). The own-race bias generalizes to a variety of face recognition tasks, such as police photo lineups (Brigham & Ready, 1985) and reaction time studies of face recognition (Valentine, 1991). Indeed, over a hundred studies have shown a deficit in memory for other-race faces.

However, the own-race bias is not only present for own-race versus other-race faces. A more general own-group bias exists in face memory, such that faces of one’s own age group, (Wright & Stroud, 2002), gender (Lewin & Herlitz, 2002; Wright & Sladden, 2003), and even arbitrary ingroups are better remembered (M. Bernstein, S. Young, & Hugenberg, 2007; Shriver, S. Young, Hugenberg, M. Bernstein, & Lanter, 2008). Further, instructing people about the own-race bias (Hugenberg, Miller, & Claypool, 2007) or manipulating eye gaze (Adams, Pauker, & Weisbuch, 2010) can eliminate the own-race bias, suggesting that the own-race bias itself is based on motivational factors. This research supports the assertion that group membership, rather than perceptual expertise or more experience with same-race faces is the cause for ingroup advantages in face memory.
We also mentalize more accurately and more extensively about those in our own group than other groups. Social psychological research describes the outgroup homogeneity effect, or the conception that we tend to think members of outgroups are more similar to one another than members of our own group (Judd & Park, 1988; Ostrom & Sedikides, 1992; Tajfel, 1970). Related to this concept is infrahumanization, or the tendency for people attribute more complex and uniquely human mental states to their own group while attributing more basic mental states (which humans share with animals) to outgroup members (Leyens et al., 2000). People take longer attributing mental states to ingroup members versus outgroup members, use more characters, and use words rated as more complex (Adams et al., in preparation). Further, people more readily associate words that indicate uniquely human emotions to ingroup others implicitly (Boccato, Cortes, Demoulin & Leyens; 2007; Paladino et al., 2002).

Ingroup biases also affect the accuracy of decoding others’ mental states based on the face. People show a slight, but significant advantage in accuracy of decoding basic emotions of members of their own culture than of other cultures (Elfenbein & Ambady, 2002). Likewise, both Japanese and U.S. Caucasian participants are more accurate at mentalizing about faces of their own culture than of another culture (Adams et al., 2010), and this intragroup advantage predicts the cross-race memory effect for these same stimuli (Adams, Franklin, & Stevenson in preparation). Ingroup biases in emotion recognition are not limited to different racial or cultural groups, as participants show an ingroup deficits recognizing emotional expressions in faces of their same race in arbitrarily assigned outgroups (S. Young & Hugenberg, 2010).

Parallel ingroup biases exist for memory as well as mentalizing about others. Indeed, the ingroup bias in mentalizing predicts the ingroup bias in face memory. This indicates a relationship between ingroup biases in these two processes, suggesting a relationship between
mentalizing and face memory more generally. If more extensive mentalizing leads to greater memory, as I propose, then the own-group bias in face memory may be because we mentalize deeper about those who are in our own group, thereby generating more elaborate associations. Thus, we remember them better. As I discuss below, such depth of processing through elaboration is known to yield better memory for a variety of stimuli, including faces.

**Facial Emotion and Memory Processing.**

Even though little evidence exists linking the cognitive processes of mentalizing and memory from the perspective of a perceiver, several studies have examined the relationship between remembering facial identity from the face and the emotional content expressed in a face. Though some of the prevailing models of face perception above (e.g., Bruce & A. Young, 1986; Haxby et al. 2000) suggest face identity encoding may have separate neural mechanisms than emotion processing, it is a stretch to assert that emotion has no impact on remembering faces. In fact, emotional expressions cause variations in the very same facial cues that are encoded when remembering faces. If this was the only impact emotional expression had on memory (i.e., if the message communicated by an expression is unrelated to memory), then the only effect that emotion should have on memory is to impair face memory overall. However, if processing another’s mental state (which includes processing their emotional expression) can affect face memory, then emotional content should meaningfully affect memory, facilitating memory for some expressions and disrupting memory for others.

Indeed, the type of facial emotion one is displaying affects face memory. Several studies have shown a bias for remembering happy faces versus negative expressions, such as angry or sad faces (D’Argembeau, van der Linden, Comblain, & Etienne, 2003; Ridout, Astell, Reid, Glen, & O’Carroll, 2003). This may be because happy faces are more rewarding than other
expressions (Tsukira & Cabeza, 2008) or because happy faces evoke positive mood, which leads to greater holistic processing of the faces and thus better encoding (Bridge, Chiao, & Paller, 2010).

Additionally, ambiguous emotional expressions, created by combining a face displaying one emotion with the top half of the face with the same identity displaying a different emotion on the bottom half, are better remembered than faces that contain just a single emotion (LaPlante & Ambady, 2002). In this case, the effect was most salient for mixed emotions that were both positive and negative, i.e., a positive expression on the top of a face combined with a negative expression on the bottom of the face of the same identity. These mixed emotions were remembered better than mixed emotions that were both positive or both negative. The authors argued that the mixed emotions are more common in everyday life (rather than pure emotional expressions) and thus more familiar, and then more remembered.

Other contextual factors can lead to better memory for other emotions. Threatening displays such as anger can increase memory for outgroup faces, even offsetting the own-group bias discussed above (Ackerman et al., 2006). This is ostensibly because it is more important to remember outgroup members who are potential threats than outgroup members who may not be potential threats.

**Direct Evidence**

Studies showing a relationship between emotional content and face memory suggest mentalizing and face memory are linked. However, these studies do not provide any direct evidence that mentalizing actually affects how faces are remembered. In this dissertation, I propose that mentalizing can affect memory and those who we mentalize more about are better remembered than those we mentalize less about.
If a relationship exists between mentalizing and memory, there are two ways that mentalizing can influence memory. First, individual differences in mentalizing ability could be related to individual differences in face memory. If mentalizing affects face memory, then those who mentalize more extensively or more accurately in general about a face should also show better face memory. In previous work we conducted, we found that individual differences in a test of mentalizing ability (the Diagnostic Analysis of Nonverbal Ability, or DANVA) was indeed correlated with individual differences in face memory, such that people who scored higher on the DANVA also remembered faces better (Franklin & Adams, 2010).

Stimulus differences in how much a certain face evokes mentalizing could predict memory of faces on a stimulus level. Faces, even with ostensibly neutral expressions, vary consistently in the degree to which participants attribute either basic or complex mental states to them. If mentalizing is related to memory, then faces that evoke more complex mental states should be better remembered. Supporting this contention, we also found previously that faces that were higher in emotional state complexity (that is, people attributed more complex mental states to them) were better remembered than faces that were lower in rated mental state complexity (Franklin & Adams, 2010). These findings suggest that mentalizing about a face can influence how a face is remembered, but only provide correlational evidence of a relationship.
Chapter 2. Behavioral Manipulations of Mentalizing and Memory

In the preceding chapter, I outline evidence for the possible relationship between mentalizing and face memory, including parallel ingroup biases for the two processes and facial emotion affecting face memory. I also described correlational evidence that the ability to decode mental states from faces is correlated to face memory. Further, faces that elicited more complex mental state judgments were better remembered. This is evidence for a relationship between mentalizing and memory and that it is possible that mentalizing affects memory; however the correlational approach of this evidence does not directly support the assertion that increased mentalizing causes greater memory.

A stronger test for the role of mentalizing in face memory would be to experimentally manipulate mentalizing and examine if it affected how faces were remembered. Manipulations that elicited more extensive mentalizing should lead to better subsequent memory. In the first part of this dissertation, I examined this by having participants rate how much people appeared to be thinking or feeling mental states that were previously judged as representing either very basic or very complex emotions or mental states. I theorized that when people judged if a person was feeling a more complex mental state, they would engage in more extensive mentalizing processes in order to make this assumption. In Study 1, using a surprise memory test, and in Study 2, using an intentional memory test, I found greater memory for faces when participants were asked to judge if a person was thinking or feeling a complex versus a basic mental state.
Study 1

Participants

Eighty-one Caucasian undergraduate participants (42 male, 39 female, mean age = 19.39, SD = 1.34) participated in this study for course credit.

Stimuli

168 grayscale facial images of Caucasian individuals (half female, half male) were used as the face stimuli. Faces were selected from the AR Face Database (Martinez & Benavente, 1998), the Color FERET database (Phillips, Wechsler, Huang, & Rauss, 1998), and from photographs of Penn State students taken 4 years before this experiment, thereby reducing the possibility of participants seeing a familiar face. Faces were cropped using an ovoid region to include the face and to exclude hair cues, due to the fact that hair cues are particularly salient in face recognition (O’Donnell & Bruce, 2001). Images were resized to a resolution of 320 x 240 pixels, and displayed at a size of 7.5 x 5 cm, subtending a visual angle of 7.2 x 4.8 degrees with participants seated approximately 60 cm from the screen.

Twelve words were selected to be the mental state adjectives used in the encoding phase, half of which were pre-rated as being high and half as low in perceived mental state complexity. In order to arrive at these 12 words, 15 separate undergraduate participants completed a task in which they viewed 114 mental state adjectives and rated them on a seven point scale indicating how complex of a mental state each word represented with 1 = no mental state and 7 = complex mental state. We also had them indicate whether they understood what the word meant. From their ratings, the highest and lowest rated six words that were understood by all of the pre-raters were used in this study (see Appendix 1). The words used for the low mental state complexity were lazy, sleepy, tired, bored, disinterested, and relieved, and the words used for the high
mental state complexity were depressed, reflective, regretful, loving, thoughtful and sympathetic. High complexity words \((M = 5.49, SE = 0.22)\) were rated significantly more complex than the low complexity adjectives \((M = 2.68, SE = .27; t(14) = 9.36, p < .0001)\). Further, these words were used in the English language at relatively the same frequency (according to word frequency data for published works in 2000 provided by Michel et al., 2011; high complexity words mean frequency \(= 9.7 \times 10^{-7}, SD = 6.1 \times 10^{-7}\) low-complexity words mean frequency, \(6.75 \times 10^{-7}, SD = 6.9 \times 10^{-7}\), \(t(14) = .22\)).

**Procedure**

Participants completed an incidental memory encoding task followed by a distracter math task and a retrieval memory test where participants had to choose between previously seen faces and novel faces. In the encoding task, participants saw 84 faces (42 male, 42 female), each of which was paired with a basic or complex mental state adjective. Participants completed one of two versions of the experimental task, counterbalanced across participants, such that for approximately half of the participants, each face was paired with a basic mental state adjective and for the other half, each face was paired with a complex mental state adjective. For each face, participants saw the adjective for 1 second, followed by the face for 2 seconds, a blank screen for 1 second, and then was presented with an untimed rating task where participants rated on a 7 point scale how much the stimulus face appeared to be displaying the mental state adjective. Participants were not told that there would be a subsequent memory test.

Following the encoding task, participants performed a math distracter test that they were told was part of a different experiment. This task lasted 3.5 minutes. Following a 30 second introduction screen, participants saw each of 12 math problems for 15 seconds each.
problem was displayed on the screen and participants typed in a response. Problems were created such that all the numbers used were 2 digits and that the answers were 2 digits (e.g. 34 + 27).

Following the distracter task, participants saw an instruction screen for the retrieval task that lasted 30 seconds, thereby leaving 4 minutes between the end of the encoding task and the beginning of the retrieval task. In the retrieval task, participants saw 168 faces in random order, with 84 novel faces added to the 84 previously seen faces. In the retrieval task, participants were asked to indicate if a face was previously seen or if it was new using a 4 choice response. Participants chose 1 on the keyboard if a face was definitely a previously seen face, 2 if the face was possibly a previously seen face, 3 if the face was possibly new, and 4 if the face was definitely new.

**Results**

To explore the effect of the manipulation of basic versus complex adjectives on memory, I assessed if there was a significant difference in memory for faces encoded while being assessed for basic versus complex mental states. In order to assess recollection memory, I examined responses for faces that were either encoded as high confidence hits (HCH) or high confidence misses (HCM) in order to remove responses related to guessing (e.g. Yonelias, 1994). HCH responses were responses where a face was previously seen and marked as definitely seen by the participant while HCM responses were responses where participants marked a previously seen face as definitely new. Performance was calculated by a difference score for HCH minus HCM for faces encoded using complex or basic adjectives separately for each participant. This analysis was done in contrast to signal detection approaches due to the specific interests in high-confidence responses and because the manipulation here was only done for the encoding manipulation. Therefore, calculating signal detection scores (e.g. D-prime) would not be
additionally informative as those assess response biases by calculating ratios using responses for
target faces versus responses to novel foils shown in the retrieval task. In this case, the mental
state manipulation did not affect the foils but was only an encoding manipulation; therefore it
would not be informative to use signal detection scores.

Using this as the dependent variable in a paired sample t-test, faces encoded while
assessing complex mental state adjectives were better remembered than basic adjectives ($t(80) = \text{2.152}, p = .034, d = .156$, see Figure 1, top pane). Complex adjectives elicited a greater
proportion of high confidence hits minus high confidence misses ($M = .105, SE = .024$) than
basic adjectives ($M = .072, SE = .023$). When comparing the percentage of HCH for complex
versus basic adjectives, complex adjectives elicited significantly more HCH ($M = .292, SE = .018$) than
basic HCH ($M = .270, SE = .017; t(80) = 2.27, p = .026, d = .150$). When comparing
HCM, however, there was no significant difference between complex ($M = .187, SE = .016$) than
basic adjectives ($M = .198, SE = .017; t(80) = 1.23, p = .22, d = .071$).

In addition, I examined responses for LCH and LCM, even though they were more likely
to be affected by guessing. Interestingly, there were more LCM for complex adjectives ($M = .282, SE = .016$) than basic adjectives ($M = .261, SE = .016; t(80) = 2.45, p = .016, d = .167$)
There was no difference between complex ($M = .259, SE = .016$) and basic ($M = .249, SE = .015$)
adjectives for LCM ($t(80) = 1.40, p = .17, d = .085$).

Further, I explored if there were any differences in the encoding task for basic versus
complex adjectives. In this case, I measured differences in the time it took to respond to the
rating for how much each face resembled the mental state word during the encoding task as well
as the ratings of how much the adjectives resembled the faces as dependent variables. These
analyses were done on the items level examining responses for complex versus basic mental state
adjectives as independent variables, using paired sample t-tests with the response latency and rating for each face as a dependent variable. There was a trend for participants taking longer to rate faces displayed with basic mental state adjectives ($M = 1054.2$ ms, $SE = 28.2$ ms) versus complex mental state adjectives ($M = 993.8$ ms, $SE = 23.8$ ms; $t(167) = 1.64$, $p = .10$). Further, participants rated faces to resemble the adjectives more when presented with basic mental state adjectives ($M = 3.87$ out of 7, $SE = .12$) versus complex adjectives ($M = 3.56$, $SE = .10$; $t(167) = 1.952$, $p = .053$).
**Discussion**

Study 1 revealed that the same faces encoded while participants assessed complex mental states were better remembered than those same faces assessed for basic mental states. This indicates that the complexity of mental states searched for affected the subsequent memory of faces. These effects were likely not due to any additional time processing the faces, as faces were on the screen the same time for each group and there were no significant differences in the time spent rating the faces. Indeed, there was even a trend that participants took longer rating the faces with the basic adjectives. These effects are necessarily due to the search for a mental state, rather than any differences with the faces themselves, as the same faces were paired with both complex and basic mental states across an equal number of participants.
Interestingly, the effects were only in the predicted direction for HCH and HCM, as there were more LCM for complex adjectives than basic adjectives. Given that LCM and LCH represent guessing responses, this pattern could be due to guessing, but it is also possible that mentalizing leads to more recognition memory versus familiarity. Recognition memory is associated with higher confidence than familiarity, which is a less confident judgment (Wixted, 2007). If complex mental state encoding is associated with higher levels of recognition versus familiarity memory, it is possible that it would yield a pattern of better performance in high confidence judgments than low-confidence judgments.

**Study 2**

Study 2 was a replication of the effects of Study 1 using an intentional memory test paradigm where participants were told that they had a memory test upcoming. Further, participants completed two separate runs of the encoding and retrieval paradigm, where participants completed the encoding test using the same methodology as in Study 1 (except with fewer faces), a math distracter test, and then a retrieval task. After this first run of the paradigm, participants completed the same series of encoding, distracter, and retrieval tasks using the same methodology as the first run but with different faces. We examined this method in order to examine if these effects generalize to an intentional memory task, using fewer faces in each encoding block, as a pilot test for parameters required using functional magnetic resonance imaging in Study 3.

**Participants**

Sixty Caucasian undergraduate participants (32 female, 28 male, mean age = 18.67, $SD = 3.72$) participated in this study for course credit.
Stimuli

184 images were selected from the same face databases as used in Study 1 and cropped in the same manner. The same mental state words were used as in Study 1.

Procedure

Participants completed two identical and sequential runs of a memory task, each of which consisted of an encoding task, a brief distracter task with math problems, and a retrieval task where participants judged if faces were previously seen or novel. In the encoding task, participants saw 48 faces, each of which were paired with either a basic or complex adjective, and participants were asked to rate on a 7 point scale how much the mental state adjective applied to each person depicted in the photographs. The encoding and retrieval task used the same timing and methodology as in Study 1. During the retrieval task, participants saw all 92 faces in random order and had as long as they needed to respond to each face. Following the first run of the encoding and retrieval task, participants immediately completed the second run, which followed using the same encoding, distracter, and retrieval pattern.

Results

To explore the effect of the manipulation of basic versus complex adjectives on memory, we assessed if there was a significant difference in memory for faces encoded as basic versus complex adjectives using the same analysis (HCH – HCM) as in Study 1 (see also Figure 2). Faces encoded while participants assessed complex mental state adjectives were better remembered than when they assessed basic adjectives when comparing each with a within-subject t-test ($t(59) = 2.778, p = .007, d = .212$). Complex adjectives elicited a greater proportion of high confidence hits minus high confidence misses ($M = .223, SE = .023$) than basic adjectives ($M = .185, SE = .024$). When comparing the percentage of HCH for complex versus basic
adjectives, complex adjectives elicited marginally more HCH \((M = .375, SE = .020)\) than basic HCH \((M = .361, SE = .020; t = 1.728, p = .089, d = .10)\). However, basic adjectives \((M = .177, SE = .013)\) elicited significantly more HCM than complex adjectives \((M = .153, SE = .013; t(59) = 2.775, p = .007, d = .227)\). However, when comparing LCH, basic adjectives \((M = .358, SE = .020)\) elicited significantly more LCH than complex adjectives \((M = .243, SE = .015; t(59) = 3.447, p = .001, d = .837)\). Additionally, when comparing LCM, complex adjectives elicited more LCM \((M = .234, SE = .014)\) than basic adjectives \((M = .173, SE = .023; t = 2.688, p = .009, d = .588)\).

In order to examine if there were any effects of block, I used a 2(encoding manipulation: basic or complex) X 2(experimental block: first or second) within-subjects, repeated-measures ANOVA using the difference score between HCH minus HCM as a dependent variable. This analysis only revealed a main effect of manipulation \((F(1, 58) = 7.58, p = .008)\), as complex words were remembered better than basic words. There was no main effect of order or an interaction \((Fs < 1.1)\).

Further, we examined if there were any behavioral effects during the encoding rating paradigm for how long participants took to rate each face and for the ratings themselves between basic versus complex adjectives in how much participants said each face resembled the mental state adjective. These analyses were done on the items level, using paired sample t-tests with the response latency and rating for each face when rated with a complex adjective versus a basic adjective as a dependent variable. No significant differences existed in how much time participants spent judging faces in the encoding task as well (basic adjective mean = 1237 ms, \(SE = 34.2\) ms; complex adjective mean = 1174 ms, \(SE = 34.8\) ms; \(t(191) = 1.28, p = .20\)). Further, no differences existed in how much participants rated faces as displaying the adjectives when
comparing basic ($M = 3.90$ out of $7$, $SE = .12$) versus complex adjectives ($M = 3.86$, $SE = .10$; $t(191) = .19$).

**Figure 2.** Proportion of high-confidence hits and misses for complex and basic adjectives in Study 2.

![Figure 2](image-url)

**Discussion**

Study 2 replicated the findings of Study 1, indicating that faces encoded while assessing complex mental state adjectives were better remembered than faces encoded while assessing basic mental state adjectives, when comparing high confidence responses. Further, Study 2 indicated that this effect generalized in an intentional memory paradigm, where participants knew a memory task was coming. This indicates that the effects of reading complex or basic mental states into faces is present even when participants know they should remember all faces,
which offers further evidence that this effect is not probably not simply due to greater depth of processing for complex mental state faces; thus, the effect is more likely due to differences in the way that faces were being encoded during complex mental state reasoning.

Study 2 also found higher levels of LCM and lower levels of LCH for complex versus basic adjectives, which is counter to the prediction of better overall memory for greater mentalizing. As mentioned above, LCH and LCM are often not used in judgments of memory due to their lower certainty, but the replication of this pattern more strongly suggests greater use of recollection memory for faces encoded with complex adjectives.
Chapter 3. Neural Mechanisms of a Relationship

Studies 1 and 2 demonstrated that manipulating mentalizing caused differential impacts in how faces were remembered. In this chapter, I present two fMRI studies examining the neural relationship between mentalizing and face memory, examining if differential levels of mentalizing led to different patterns of neural activation in regions involved with face processing, mentalizing, and face memory. In Study 3, I present a preliminary examination in how variations in mental state complexity in faces affect brain activation during an encoding task. In Study 4, I extended this by experimentally varying how much mentalizing occurred while participants perceived a series of faces in an encoding test in a similar way to Study 2. Participants were then surprised with a recall task and I examined on a subject-level and item-level how mentalizing affected successful and unsuccessful encoding.

Study 3

In Study 3, I examined if neural evidence existed for a relationship between mentalizing and memory. Participants completed an encoding task while they were being scanned. Participants perceived a series of faces which were divided into blocks based on their pre-ratings of mental state complexity by a separate group of raters. Using these ratings, I examined if faces that previously elicited high levels of mentalizing led to different patterns of neural activity in brain regions involved with memory as well as brain regions involved with mentalizing during a encoding task.
Participants

Ten participants (4 women, 6 men) completed this study. Participants were recruited from the Hershey, PA area, and were between the ages of 18 and 35 years old, and screened to have normal or corrected to normal vision and be free of neurological damage.

Stimuli

A set of 96 grayscale images of White individuals were used in this study. Faces were collected from the AR Face Database (Martinez & Benavente, 1998), the NimStim, and Color FERET Database. Pictures were cropped to be a size of 312 x 240 pixels and displayed on MRI compatible goggles at a resolution of 1024 x 768 pixels, subtending a visual angle of 9.1 x 7.0 degrees. All hair cues and extraneous cues were also cropped out, as in Studies 1 and 2.

Faces were pre-rated by a set of 30 raters (with 14 raters rating one half of the faces and 16 raters rating the other half, Cronbach’s alpha of group 1 = .61, alpha of group 2 = .78) using the same mental state complexity scale used in Preliminary Study 1. Based on these ratings, faces were divided into four quartiles of mental state complexity separately for male and female faces. From each quartile, 6 female and 6 male faces were randomly chosen to be in the encoding part of the experiment while the other 12 faces from that quartile were used as distracters, leading to 48 faces in the encoding portion of the experiment. The other forty-eight faces were shown as distractors with the original forty-eight in the retrieval part of the experiment after scanning.

Design and Procedure

Participants completed a memory encoding task while being scanned in an MRI scanner and were tested on their memory following scanning using a recognition test. Prior to scanning,
faces were divided into four quartiles of 12 faces based on pre-ratings of the emotional state complexity of the faces. These quartiles were further divided into four blocks of 3 faces, leading to sixteen blocks of faces. These blocks formed the block design of the experiment, with 8 blocks of baseline fixation also included.

Each block consisted of three faces, with each face shown for 4 seconds, with a fixation of 1 second following each face. Each block lasted 15 seconds and the fixation blocks were the same length. Blocks were shown in a pseudo-random order in regard to the complexity of the face, with the order of the faces within the block shown randomly. Total run time was 6 minutes.

Immediately following the encoding portion of the experiment, participants were removed from the magnet and taken to a separate room and given a memory posttest where they were shown a series of faces and asked if the face shown was previously seen or a novel face. There was approximately between a 5 to 8 minute interval between the end of the encoding task and the beginning of the retrieval test. The 48 faces shown in the scanner were mixed with 48 other faces from the same set. Participants identified if the face was previously seen with a left mouse click and if it was novel with a right mouse click. Faces were shown in random order and remained on the screen until the participant made a response.

**Scanning Parameters**

The study was carried out on a 3.0 Tesla Philips MRI scanner (Philips Medical Systems, the Netherlands). Visual stimuli were administered using MRI compatible goggles.

Anatomical and functional data was acquired by the MRI scanner. A T1–weighted three-dimensional image (3-D TFE) was acquired for anatomical structure with time of repetition (TR)/ echo time (TE)/ flip angle (FA)=8.05ms/ 3.7ms/ 8°, FOV=256 x 150 mm³, acquisition matrix=256 x 256x150 mm, and a SENSE factor =2. Echo planar imaging (EPI) was used for
data acquisition with TR=3600ms, TE=30ms, FA=90°, 28 axial slices, slice thickness=2.5 mm,
FOV=210 x 210 mm³, acquisition matrix=80 x 80, reconstruction resolution=128 x 128, and
SENSE factor=2. Two tasks were administered in each imaging session. For each session, 64
images were acquired during 8 blocks of stimulation and baseline.

MRI data were pre-processed and analyzed using SPM5 software. Functional data for
each individual subject by using a rigid six-parameter motion correction, and then coregistered to
the anatomical image. The anatomical image was segmented using a T1-weighted template in
MNI space and normalized based on these parameters. These normalization parameters were
then applied to the functional images to normalize data into the same brain space. Functional
data were then smoothed with an 8mm³ full-width, half maximum Gaussian kernel.

Data were analyzed using a mass-univariate GLM approach using statistical parametric
mapping. For each subject, individual first-level fixed analyses were used. The effects of interest
were modeled as a block design comparing brain activation for blocks that varied in the level of
pre-rated complexity of each face from the lowest quartile to the highest quartile. Further, the six
motion parameters derived from the realignment algorithm were used as regressors as well to
control for motion-induced artifact. This analysis led to the generation of first level-contrasts for
each participant for each condition of interest that were then used as the basis of a second-level
random effects analysis across participants. These contrasts formed the basis of the contrasts
used for the second-level group analyses reported below. Given the exploratory nature of this
task, data are thresholded at $p < .005$, 10 voxels (see Lieberman & Cunningham, 2009), except
for exploratory analyses, which are reported at $p < .01$, 10 voxels.

**Results**

**Behavioral Results**
Behavioral results were collected for 9 of the 10 participants, with one participant’s behavioral data not recorded due to a computer error. Participants had an overall accuracy of 72.5% and each participant scored significantly above chance (binomial probability of a participant’s accuracy to be at chance level $p < .04$ for all participants). For each quartile of pre-rated complexity, the means of accuracy showed the predicted pattern where the most complex faces were better remembered (mean $D'$ of 1.53) than the least complex (mean $D'$ of 1.28), thus replicating research showing higher memory for faces rated as more complex (Franklin & Adams, 2010). However with only 9 participants’ data, this difference was not significant ($t(8) = .738$, $p > .4$, $d = .30$).

**fMRI Results**

In order to examine the effects of emotional state complexity on memory, I computed first-level contrasts with the linear effect of increasing complexity and evaluated these using a group-level whole-brain analysis. This parametric analysis revealed activation in the right hippocampus, left superior temporal sulcus (STS), and left lateral prefrontal cortex at the set threshold of $p < .001$, 10 voxels. At the reduced threshold of $p < .01$, the medial PFC and right fusiform gyrus were also active (see Figure 3. For additional regions active, see Table 1).

Additionally I assessed if brain regions involved with mental state complexity were related to brain regions associated with memory. These regions were pulled because they were significant at the original threshold and also critically implicated in memory or reading emotional cues from faces. For the clusters that were significant in the right hippocampus and left STS, I pulled beta coefficients from each individual’s activation in those voxels significant in the group level contrast. Using these beta coefficients, I found that subjects’ mean activity in the left STS was correlated with their activity in the right hippocampus ($r = .716$, $p = .020$).
Table 1. Regions active for the parametric increase of mentalizing in Preliminary Study 1.

Regions are reported posterior to anterior. Italicized regions are significant at an exploratory threshold of \( p < .01 \), 10 voxels.

<table>
<thead>
<tr>
<th>Parametric Increase</th>
<th>Coordinates</th>
<th>T-Value</th>
<th>Cluster Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. Cuneus</td>
<td>-14, -88, 2</td>
<td>3.89</td>
<td>176</td>
</tr>
<tr>
<td>L. Precuneus</td>
<td>-16, -80, 34</td>
<td>3.62</td>
<td>9</td>
</tr>
<tr>
<td>L. Cerebellum</td>
<td>-22, -74, -44</td>
<td>4.13</td>
<td>51</td>
</tr>
<tr>
<td>R. Cerebellum</td>
<td>12, -56, -24</td>
<td>4.1</td>
<td>11</td>
</tr>
<tr>
<td>L. STS</td>
<td>-64, -52, 16</td>
<td>5.28</td>
<td>63</td>
</tr>
<tr>
<td>R. Fusiform Gyrus</td>
<td>48, -42, -12</td>
<td>4.01</td>
<td>39</td>
</tr>
<tr>
<td>R. STS</td>
<td>62, -40, 18</td>
<td>4.54</td>
<td>107</td>
</tr>
<tr>
<td>L. Temporoparietal Junction</td>
<td>-54, -34, 28</td>
<td>6.49</td>
<td>254</td>
</tr>
<tr>
<td>L. Posterior Cingulate</td>
<td>-18, -34, 44</td>
<td>5.17</td>
<td>193</td>
</tr>
<tr>
<td>L. Somatosensory Cortex</td>
<td>-56, -20, 30</td>
<td>4.51</td>
<td>58</td>
</tr>
<tr>
<td>R. Hippocampus</td>
<td>34, -18, -12</td>
<td>4.92</td>
<td>326</td>
</tr>
<tr>
<td>R. Precentral Gyrus</td>
<td>56, -2, 36</td>
<td>5.81</td>
<td>522</td>
</tr>
<tr>
<td>L. Globus Pallidus</td>
<td>-10, 0, 2</td>
<td>5.1</td>
<td>366</td>
</tr>
<tr>
<td>R. Cingulate</td>
<td>18, 2, 46</td>
<td>5.24</td>
<td>119</td>
</tr>
<tr>
<td>L. Inferior Frontal Gyrus</td>
<td>-62, 4, 26</td>
<td>11.39</td>
<td>226</td>
</tr>
<tr>
<td>R. Anterior Insula</td>
<td>32, 10, 12</td>
<td>8.53</td>
<td>1337</td>
</tr>
<tr>
<td>L. Putamen</td>
<td>-14, 12, -4</td>
<td>5.18</td>
<td>155</td>
</tr>
<tr>
<td>L. Cingulate</td>
<td>-12, 12, 36</td>
<td>4.34</td>
<td>73</td>
</tr>
<tr>
<td>R. Caudate</td>
<td>16, 14, 2</td>
<td>4.45</td>
<td>41</td>
</tr>
<tr>
<td>R. Inferior Frontal Gyrus</td>
<td>40, 30, 20</td>
<td>5.38</td>
<td>223</td>
</tr>
<tr>
<td>L. Ventrolateral Prefrontal Cortex</td>
<td>-24, 32, 2</td>
<td>5.28</td>
<td>121</td>
</tr>
<tr>
<td>R. Ventrolateral Prefrontal Cortex</td>
<td>40, 38, -4</td>
<td>6.9</td>
<td>122</td>
</tr>
<tr>
<td>R. Dorsal Medial PFC</td>
<td>10, 40, 26</td>
<td>3.94</td>
<td>30</td>
</tr>
<tr>
<td>L. Medial Prefrontal Cortex</td>
<td>-10, 44, 8</td>
<td>3.88</td>
<td>33</td>
</tr>
<tr>
<td>L. Ventral Medial Prefrontal Cortex</td>
<td>14, 60, 6</td>
<td>3.62</td>
<td>15</td>
</tr>
</tbody>
</table>
**Figure 3.** Activity to parametric increase of mental state complexity. Graphs represent beta values for activity for predicted activation in R. Hippocampus (left) and L. vmPFC (right) to faces in each of the four quartiles of complexity. Regions active include the R. hippocampus (A, $p < .001$), medial PFC (B, $p < .01$), L. STS (C, $p < .005$), R. fusiform gyrus (D, a threshold of $p < .01$), and L. lateral prefrontal cortex ($p < .005$).
Discussion

Study 3 shows evidence of a relationship between mentalizing and memory. Those faces that were rated as having more complex mental states evoke more activation in brain regions involved with memory as well as mentalizing in a memory encoding task. Further, activation in the STS, which is involved in decoding mental states from the face, significantly correlated with activation in the hippocampus. This provides initial evidence for the relationship between mentalizing and memory processes and brain regions in a memory encoding task. However, this study does not indicate if manipulating mentalizing affects memory, as it could be that faces that happen to be more memorable (due to a factor like distinctiveness), could also elicit more processing in mentalizing-related areas.

Study 4

The memory difference for complex versus basic mental state adjectives suggests that the way faces are encoded is modulated by the mental states read from them, suggesting a relationship between the processes of mentalizing and face memory. However, these behavioral effects do not show direct evidence for the link between the mechanisms involved in mentalizing and face memory. Several other explanations exist which can explain this effect that do not necessitate a direct relationship between mentalizing and face memory. For instance, greater mentalizing can cause participants to pay greater attention to faces independent of any mentalizing-specific processes. Further, the valence of the mental state words can affect how mental states are read from faces. In this thesis, I predict that mentalizing affects the neural, as well as behavioral, encoding of faces.

Study 3 provided preliminary evidence for the relationship between mentalizing and face memory showing the involvement of mentalizing areas in perceiving faces with higher mental
state complexity during a face memory task. This is consistent with the supposition that these regions are involved in face memory and this may be due to their role in mentalizing about faces, but does not provide direct experimental evidence for a relationship. Study 4 extended this, experimentally investigating the relationship between mentalizing and face memory.

The purposes of Study 4 were as such: I sought to examine if regions involved with mentalizing and face memory were both recruited during an encoding task that varied the degree of mentalizing that was present for each face. Further, I examined if the activation in neural regions involved with mentalizing was related to behavioral differences in memory, specifically investigating if mentalizing was a moderator of the relationship between activation in mentalizing and memory regions and successful encoding.

Study 4 employed a similar paradigm as in Study 2 in order to examine whether encoding faces while assessing them for basic versus complex mental states leads to differences in neural processing in a subsequent surprise memory paradigm. In this study, I hypothesized that regions involved with face memory as well as mentalizing would both be active to encoding faces with complex adjectives versus basic adjectives. Second, I also hypothesized that successful encoding would lead to greater memory-related activation than unsuccessful encoding. Critically, I hypothesized that mentalizing would moderate the relationship between encoding and memory-related activation, as hits encoded with complex adjectives would elicit greater mentalizing and memory activation than hits with basic adjectives or misses. Such a result would indicate a meaningful function of mentalizing regions in face memory.

**Participants**

30 participants completed this study (16 female, 14 male). Participants were between 20 and 30 years old. Participants were screened to be right-handed, free of any history of
neurological problems, and have normal or corrected to normal vision. Of those who completed
the study, 4 participants were excluded (2 female, 2 male), one for excessive movement (> 5 mm
across the run), while three participants were excluded due to equipment failure that prevented
collection of their behavioral data. Of the participants that remained, 24 participants were of
Caucasian descent and two participants were of Asian descent. Participants were recruited from
advertisements and from participation lists from previous fMRI studies and were all either
undergraduate or graduate students. All participants reported being naïve of the study hypotheses
when asked both before and after performing the study. Study and recruitment methods were
approved by the Pennsylvania State University Institutional Review Board.

Stimuli

160 stimuli (80 female, 80 male) were used in this paradigm. These were randomly
selected from the same set as used in Studies 1 and 2 and displayed at the same size as in Studies
1 and 2. 96 stimuli were used as target stimuli whereas 84 stimuli were used as foils and only
displayed in the retrieval phase of the study. The same stimuli were used as targets versus foils
for each participant.

Design and Procedure

Participants completed two sets of an encoding and retrieval task while being scanned,
with each set consisting of an encoding task, followed by a 6-8 minute period, and a surprise
retrieval task. This design was very similar to Study 2, except that participants were not told that
a memory test was following each rating task (i.e., incidental encoding task). Participants were
given two four-button response boxes to use, one for each hand, with buttons corresponding to
their pinky, ring, middle, and index finger of each hand.
Each encoding task consisted of a series of 48 faces displayed with accompanying mental state adjectives. Half of the faces were presented with basic mental state adjectives and half with complex mental state adjectives (the same words used in Studies 1 and 2). Each trial consisted of a mental state word presented for 1 second, followed by the face, which was presented for 2 seconds. A 1 second blank screen followed, which was subsequently followed by a 3.5 second period where the participant rated how much the mental state adjective applied to the face, using the same 7 point scale as used in Studies 1 and 2. This scale used a rating where the left pinky represented 1, the left index finger represented 4, the right index finger represented 5, and the right ring finger represented 7. After each trial, either a 0, 2.5, or 5 second jitter followed in a pseudorandom order. These jitter periods were determined using a stochastic method in order to allow for maximum power and the best possible assessment of the hemodynamic response (e.g. Burock, Buckner, Woldorff, Rosen, & Dale, 1998; Wager & Nichols, 2003). Each jitter period consisted of a blank screen with a fixation cross for the period of the jitter, between each trial. There were 12 jitters of 0 seconds (the next trial followed immediately after the previous trial), 24 jitters of 2.5 seconds and 12 jitters of 5 seconds. There was a 10 second fixation following the end of the paradigm.

Following the encoding task, participants had a 6 to 8 minute break before the retrieval task, where they were told that they were going to undergo an anatomical scan and they had no task to perform. During the first break between encoding and retrieval tasks, a high resolution anatomical T1 scan was collected, while during the second break, a diffusion tensor imaging scan was collected (which was not analyzed). Following this break, participants completed a surprise retrieval task, where they viewed previously seen faces and novel faces and responded if the faces were novel or previously viewed using a 4 choice response. Participants responded if
the face was definitely previously viewed by pressing their right index finger, if the face was probably previously viewed using their right middle finger, if the face was probably new using the right ring finger, and if the face was definitely new using the right pinky.

Immediately following the first encoding and retrieval task, the second encoding and retrieval task began. Participants were told prior to completing the second encoding task that the second encoding task would not have a memory component following it so that the second encoding task would be an incidental encoding task as well. No participants reported expecting that the second encoding task would have a following retrieval task. Following the second set of encoding and retrieval tasks, participants were removed from the scanner and debriefed.

**Scanning Parameters and Analysis**

Functional data were collected for both the encoding and retrieval task; however, only data from the encoding task was analyzed as the current study was focused on the influence of mental state reasoning on encoding. Functional data were collected using a 3T Siemens Tim-Trio using echo-planar imaging with 2 runs of 188 T2* images (TR: 2500ms, TE: 30ms, flip angle = 80 degrees, 44 interleaved slices, transverse orientation, 2.5x2.5x2mm voxels, with 2 mm slices and .5mm skip). SPM8 software was used to preprocess and analyze data (Wellcome Institute, London, UK). Differences in slice-timing were corrected by interpolating data in time to the time of the acquisition of the first slice of each functional volume. After this, data were realigned by a six-parameter rigid realignment which generated movement parameters used as a regressor for the first-level analysis to control for artifact activation caused by movement. Following this, the mean functional realigned image was coregistered to a T1 anatomical image. This anatomical image was then segmented and normalized to the Montreal Neurological Institute (MNI) template. The normalization parameters generated by this normalization were then applied to
each functional image and functional voxels were resliced to have 2.5mm isotropic voxels. Finally, functional images were smoothed using an 8mm full-width, half-maximum 3D Gaussian kernel.

Functional data were analyzed using statistical parametric mapping using a multi-level approach implemented in SPM8 software. This approach involved fitting a first-level fixed-effects model within each subject followed by a second-level random-effects analysis across individual subjects. For the first-level fixed-effects model, stimulus conditions were modeled as delayed boxcar functions convolved with a standard hemodynamic response function. These conditions were modeled as 2 second blocks for each face during the period which the face was presented. Additionally, the scale response was modeled as a regressor. All other conditions (such as fixation and jitter periods) were unmodeled. Low frequency signal components less than 0.007Hz were filtered out using a high-pass filter. Contrasts were generated from these fixed-effects model and were subsequently used into a second-level random-effects model representing summary measures of subject responses. For each subject, we generated first-level contrasts predicting activation for each contrast (e.g., responses to complex HCH) and difference contrasts (e.g., complex faces minus basic faces). These contrasts were then used as the basis for group-level activation.

Activation maps were thresholded at $p < .005$ uncorrected, with a cluster extent of 10 voxels. Lieberman and Cunningham (2009) assert that this threshold represents a reasonable tradeoff between Type-1 and Type-2 errors. This is especially germane to this study given the novelty of comparing mentalizing processes in memory. However, using this data, I also conducted a Monte Carlo simulation using AlphaSim included in the AFNI (Analysis for Functional Neuroimaging) program (Cox, 1996). This simulation was performed using 1000
iterations using the voxels based on the acquisition mask generated from the group-level random effects model. This simulation revealed that any activations meeting a threshold of $p < .005$, 73 voxel extent, can be treated as equivalent to a threshold of $p < .05$, corrected.

Results

Behavioral Results

In order to assess if the mental state adjective manipulation affected memory, I examined memory performance in the same way as in Studies 1 and 2 by performing a difference score between high confidence hits (HCH) minus high confidence misses (HCM) separately for faces encoded with complex mental state adjectives and basic mental state adjectives (see Figure 3). Using this difference score, faces encoded with complex mental state adjectives ($M = .214, SE = .032$) were better remembered than those encoded with basic mental state adjectives ($M = .150, SE = .039$; $t(25) = 2.61, p = .014, d = .356$). Complex adjectives led to marginally more high-confidence hits ($M = .314, SE = .029$) than basic adjectives ($M = .285, SE = .039$; $t(25) = 1.84, p = .077, d = .202$). The reverse pattern was present for high confidence misses, as basic adjectives ($M = .135, SE = .023$) led to more HCM than complex adjectives ($M = .100, SE = .015$; $t(25) = 2.54, p = .018, d = .359$). There was no difference in LCH for complex ($M = .276, SE = .020$) and basic hits ($M = .285, SE = .029$; $t(25) = .028$), but there was significantly more LCM for complex adjectives ($M = .280, SE = .025$) than basic adjectives ($M = .135, SE = .023$; $t(25) = 4.21, p < .001, d = 1.17$).
**Figure 4.** Proportion of high-confidence hits and misses for complex and basic adjectives for the behavioral data in Study 4.
fMRI Results

Encoding Manipulation. First, I examined the effects of encoding manipulation by comparing activation for faces encoded with a complex mental state adjective versus a basic mental state adjective. This analysis, as reported in Table 2 and Figure 5 led to activation in several regions involved with mentalizing, including bilateral regions of the ventromedial and dorsomedial prefrontal cortex, the left IFC, and right temporal pole (see Table 2, Figures 5 and 6). Activation in these regions likely reflected increased processing of mental states for the complex adjectives. Additionally, the left fusiform gyrus was active for this condition, likely reflecting deeper face processing. Regions involved in memory processing were also active for this contrast, including bilateral activation in the hippocampus. The reverse contrast led to activation in the bilateral insula and right superior temporal sulcus (See Figure 5).
Table 2. Activations for the contrast comparing complex minus basic mental state adjectives, including cluster extent.

<table>
<thead>
<tr>
<th>Region</th>
<th>Coordinates</th>
<th>T-Value</th>
<th>Extent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
<td>y</td>
<td>z</td>
</tr>
<tr>
<td>L. Anterior Insula</td>
<td>-26</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>L. vmPFC</td>
<td>-1</td>
<td>43</td>
<td>-20</td>
</tr>
<tr>
<td>L. Thalamus</td>
<td>-16</td>
<td>-30</td>
<td>15</td>
</tr>
<tr>
<td>L. mPFC</td>
<td>-8</td>
<td>46</td>
<td>48</td>
</tr>
<tr>
<td>R. mPFC</td>
<td>12</td>
<td>61</td>
<td>13</td>
</tr>
<tr>
<td>L. dmPFC</td>
<td>-3</td>
<td>56</td>
<td>23</td>
</tr>
<tr>
<td>R. Temporal Pole</td>
<td>47</td>
<td>1</td>
<td>-25</td>
</tr>
<tr>
<td>L. Lingual Gyrus</td>
<td>-38</td>
<td>-80</td>
<td>-10</td>
</tr>
<tr>
<td>L. Cuneus</td>
<td>-11</td>
<td>-102</td>
<td>-5</td>
</tr>
<tr>
<td>R. dmPFC</td>
<td>25</td>
<td>38</td>
<td>10</td>
</tr>
<tr>
<td>R. ACC</td>
<td>20</td>
<td>31</td>
<td>20</td>
</tr>
<tr>
<td>L. Fusiform Gyrus</td>
<td>-36</td>
<td>-45</td>
<td>-23</td>
</tr>
<tr>
<td>L. Interior Frontal Cortex</td>
<td>-31</td>
<td>21</td>
<td>-23</td>
</tr>
<tr>
<td>R. Lingual Gyrus</td>
<td>15</td>
<td>-95</td>
<td>-5</td>
</tr>
<tr>
<td>L. Hippocampus</td>
<td>-33</td>
<td>1</td>
<td>-25</td>
</tr>
<tr>
<td>R. Hippocampus</td>
<td>25</td>
<td>-17</td>
<td>-25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Region</th>
<th>Coordinates</th>
<th>T-Value</th>
<th>Extent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
<td>y</td>
<td>z</td>
</tr>
<tr>
<td>R. Posterior Insula</td>
<td>30</td>
<td>-20</td>
<td>18</td>
</tr>
<tr>
<td>L. Posterior Insula</td>
<td>-43</td>
<td>-25</td>
<td>20</td>
</tr>
<tr>
<td>R. Cuneus</td>
<td>10</td>
<td>-90</td>
<td>13</td>
</tr>
<tr>
<td>L. Lateral Occipital Gyrus</td>
<td>-23</td>
<td>-82</td>
<td>23</td>
</tr>
<tr>
<td>R. Precuneus</td>
<td>17</td>
<td>-60</td>
<td>50</td>
</tr>
<tr>
<td>R. Precentral Gyrus</td>
<td>65</td>
<td>-17</td>
<td>40</td>
</tr>
<tr>
<td>L. Precentral Gyrus</td>
<td>-26</td>
<td>-7</td>
<td>60</td>
</tr>
<tr>
<td>R. STS</td>
<td>55</td>
<td>-32</td>
<td>3</td>
</tr>
<tr>
<td>L. Posterior Cingulate</td>
<td>17</td>
<td>-60</td>
<td>18</td>
</tr>
<tr>
<td>L. Anterior Insula</td>
<td>-36</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>R. Anterior Insula</td>
<td>47</td>
<td>-2</td>
<td>10</td>
</tr>
<tr>
<td>R. Lateral PFC</td>
<td>47</td>
<td>48</td>
<td>15</td>
</tr>
</tbody>
</table>
Figure 5. Activations to complex mental state adjectives minus basic mental state adjectives. T-maps are overlayed on the average of participants’ T1 anatomical image.
Figure 6. Activation to complex minus basic mental state adjectives in the medial prefrontal cortex (above) and hippocampus (below).
**Hits versus Misses.** Second, I examined responses for hits versus misses. In order to increase power due to the low number of faces in the design, I compared responses for HCH versus all misses. HCH minus misses led to activation in the left hippocampus and parahippocampal cortex (PHC), along with increased activation in the bilateral vmPFC (see Table 3 and Figures 7 and 8). The reverse contrast led to activation in the precuneus, ventrolateral PFC, and parietal regions including the precuneus and posterior cingulate (see Table 3 and Figure 7).
Table 3. Regions active to the contrast comparing HCH minus misses and misses minus HCH.

<table>
<thead>
<tr>
<th>Region</th>
<th>Coordinates</th>
<th>T-Value</th>
<th>Extent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>HCH Minus Miss</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L. Cuneus</td>
<td>-11</td>
<td>-102</td>
<td>3</td>
</tr>
<tr>
<td>R. Inferior Temporal Gyrus</td>
<td>45</td>
<td>-72</td>
<td>-10</td>
</tr>
<tr>
<td>L. Parahippocampal Gyrus</td>
<td>-26</td>
<td>6</td>
<td>-18</td>
</tr>
<tr>
<td>L. Putamen</td>
<td>-23</td>
<td>-7</td>
<td>5</td>
</tr>
<tr>
<td>R. vmPFC</td>
<td>10</td>
<td>41</td>
<td>-10</td>
</tr>
<tr>
<td>L. vmPFC</td>
<td>-3</td>
<td>51</td>
<td>-18</td>
</tr>
<tr>
<td>Miss Minus HCH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L. Precuneus</td>
<td>-18</td>
<td>-72</td>
<td>15</td>
</tr>
<tr>
<td>R. Precuneus</td>
<td>17</td>
<td>-70</td>
<td>20</td>
</tr>
<tr>
<td>L. Occipito-Parietal Junction</td>
<td>-33</td>
<td>-67</td>
<td>30</td>
</tr>
<tr>
<td>R. Posterior Cingulate</td>
<td>30</td>
<td>-32</td>
<td>43</td>
</tr>
<tr>
<td>R. Temporoparietal Junction</td>
<td>62</td>
<td>-32</td>
<td>28</td>
</tr>
<tr>
<td>L. Postcentral Gyrus</td>
<td>-58</td>
<td>-27</td>
<td>15</td>
</tr>
<tr>
<td>R. Thalamus</td>
<td>5</td>
<td>-25</td>
<td>13</td>
</tr>
<tr>
<td>R. Precentral Gyrus</td>
<td>25</td>
<td>13</td>
<td>55</td>
</tr>
<tr>
<td>R. ventrolateral PFC</td>
<td>37</td>
<td>48</td>
<td>23</td>
</tr>
</tbody>
</table>

* Indicates peak is part of preceding cluster
Figure 7. Regions active for HCH minus misses (orange to yellow) and regions active for misses minus HCH (blue to green).
Figure 8. Activation in the vmPFC (above) and hippocampus (below) for HCH minus misses.
Interaction between Complexity and Memory. In order to see if level of mentalizing interacted with successful encoding, I conducted a subject-level 2 (level of mentalizing: complex versus basic) X 2 (encoding success: HCH versus misses), within-subject, whole-brain ANOVA. This was conducted using a planned contrast evaluating Complex HCH versus each of the other conditions (Basic HCH and Complex and Basic misses). This analysis revealed an interaction in activation in regions involved with mentalizing, including the dmPFC and vmPFC, as well as regions involved with face processing and memory, including the right FG and left hippocampus and parahippocampal cortex (PHC) (see Table 4 and Figures 9-10).
Table 4. Regions active to the interaction between complex HCH minus other encoding conditions.

<table>
<thead>
<tr>
<th>Region</th>
<th>Coordinates</th>
<th>T-Value</th>
<th>Extent</th>
<th>Beta Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
<td>y</td>
<td>z</td>
<td>Complex HCH</td>
</tr>
<tr>
<td>R. Fusiform Gyrus</td>
<td>30</td>
<td>-52</td>
<td>-20</td>
<td>3.04</td>
</tr>
<tr>
<td>L. Precentral Gyrus</td>
<td>-31</td>
<td>-30</td>
<td>55</td>
<td>3.82</td>
</tr>
<tr>
<td>L. Temporal Pole</td>
<td>-66</td>
<td>-17</td>
<td>-18</td>
<td>3.63</td>
</tr>
<tr>
<td>L. PHC</td>
<td>-16</td>
<td>-12</td>
<td>-15</td>
<td>3.14</td>
</tr>
<tr>
<td>L. PHC</td>
<td>-33</td>
<td>-7</td>
<td>-20</td>
<td>3.17</td>
</tr>
<tr>
<td>L. IFG</td>
<td>-28</td>
<td>13</td>
<td>-23</td>
<td>4.38</td>
</tr>
<tr>
<td>L. Caudate</td>
<td>-13</td>
<td>26</td>
<td>5</td>
<td>3.56</td>
</tr>
<tr>
<td>L. dmPFC</td>
<td>-11</td>
<td>31</td>
<td>58</td>
<td>4.37</td>
</tr>
<tr>
<td>R. mPFC</td>
<td>20</td>
<td>31</td>
<td>20</td>
<td>3.3</td>
</tr>
<tr>
<td>R. dmPFC</td>
<td>15</td>
<td>43</td>
<td>45</td>
<td>3.28</td>
</tr>
<tr>
<td>L. dmPFC</td>
<td>-8</td>
<td>46</td>
<td>45</td>
<td>4.57</td>
</tr>
<tr>
<td>L. vmPFC</td>
<td>-1</td>
<td>48</td>
<td>-18</td>
<td>4.31</td>
</tr>
</tbody>
</table>

* Indicates Peak is part of previous cluster
Figure 9. Regions active to the interaction between mentalizing and face memory comparing complex HCH versus all other conditions.
Figure 10. Activation in the parahippocampal cortex (top left), dmPFC (bottom left and top right) and vmPFC (bottom left and bottom right) for the interaction between complex HCH and all other conditions.
Connectivity Analyses. In order to assess the relationship of activation between regions involved in mentalizing and face memory, I also conducted a connectivity analysis. This analysis assessed the relationship between activation in regions in face memory and mentalizing regions. Given that complex minus basic adjectives led to activation in regions involved in face memory and mentalizing, including the mPFC and hippocampus, I chose to create regions of interest (ROIs) based on these regions. For the hippocampus, an ROI was created using the anatomical definition as provided by the AAL atlas template. For the dmPFC and vmPFC, I selected peak voxels for these regions in Mitchell, Macrae, & Banaji, (2005) in order to create an unbiased functional neuroanatomic definition. Mitchell et al used a similar task as used here, which involved assessing how much various mental states applied to a series of faces. I chose three peaks of activation that were significant in their study for comparing mentalizing to baseline, one for the left dmPFC (-9, 51, 36) and left (-6, 45, 3), and right (9, 57, 3) vmPFC. Those peaks were then used as the center to create spherical 8 mm. Then, I correlated the mean signal change in these regions for complex minus basic adjectives between subjects, using these unbiased ROIs. This analysis showed a relationship between activation in the left hippocampus and left vmPFC ($r(24) = .56, p = .003$) and the right hippocampus and right vmPFC ($r(24) = .41, p = .036$) and left vmPFC ($r(24) = .53, p = .006$). The dmPFC was uncorrelated with either hippocampus cluster ($r(24) > .12, p > .5$).

Discussion

Study 4 examined the neural relationship between mentalizing and memory by manipulating encoding of faces either paired with complex or basic mental state adjectives. In this study, I replicated behavioral findings in Studies 1 and 2 showing that faces paired with complex mental state adjectives were remembered better. Further, when examining neural
activation associated with encoding faces, I found several differences in encoding complex mental state adjectives minus basic mental state adjectives. This comparison elicited activation in several brain regions associated with face memory, including the FG and hippocampus. Additionally, this comparison elicited activation in regions involved in mentalizing, including the medial prefrontal cortex. Further, I investigated how neural activation predicted performance in the memory task by investigating which brain regions contributed to HCH versus misses. Replicating common findings in the memory literature, I found that this contrast led to greater activation in memory-related regions, including the hippocampus and parahippocampal cortex.

In order to assess if mentalizing meaningfully influenced neural activation involved with successful memory, I examined the interaction between mentalizing and face memory, using a planned contrast to examine activation to complex HCH versus all other conditions. This analysis targeted activation that was specific to successful encoding of faces paired with complex adjectives. This analysis revealed activation in mentalizing regions, including both the dmPFC and vmPFC. Further, this analysis led to activation in the PHC, indicating greater involvement in memory-related regions for complex encoded HCH.

In addition to these analyses, I also examined the relationship between the mPFC and memory-related activation in the hippocampus for encoding complex versus basic adjectives. This analysis found that subject-level activation in the hippocampus was positively associated with vmPFC activation. This analysis suggests that mentalizing-related activation may be in fact related to the memory-related activation in the hippocampus and provides further evidence of the role of the dmPFC in face memory.
Chapter 4. Summary and Discussion

Knowing who others are and understanding their mental states are two essential skills for social interaction. Though social psychological and neuroscientific research has investigated each of these skills extensively, very little research has examined how mentalizing and face memory may relate to each other. In this dissertation, I took a functional approach to face perception (e.g., Adams, Franklin, Nelson, & Stevenson, 2010; Zebrowitz, 2006), with the assumption that the mechanisms involved in face perception are active perceptual processes that exist to process information relevant to survival. Such an approach would suggest a relationship between mentalizing and face memory, as it would be important to remember those whom we mentalize more extensively about. In the three studies discussed above, I show that manipulating the complexity of mental state reasoning about a face affects behavioral aspects of face memory as well as the neural mechanisms underlying face memory and mentalizing.

Previous work has shown that faces that elicit highly consensual ratings of emotional complexity are better remembered, and that those who are better at emotional state reasoning are better at remembering faces (Franklin & Adams, 2010). That previous work suggested that mentalizing may contribute to face memory but these studies were purely correlational in nature, and thus cannot speak to causation. The studies herein provide direct experimental support for the relationship between mentalizing and face memory by demonstrating greater memory for faces when encoded while assessing a complex mental state versus a basic mental state. Across each of these studies, faces assessed for a complex mental state were remembered better by both increasing the amount of high confidence hits associated with each face as well as reducing the amount of high confidence misses (see Appendix 2). This pattern indicates that the encoding
involved with each face was not merely a result of an increase in the response criterion associated with each face but was a true indicator of increased discrimination.

Studies 3 and 4 also provided evidence for the involvement of both memory and mentalizing networks in memory tasks. Study 3 showed that faces that elicit more complex mental states elicit greater activation in mentalizing and face memory regions in an encoding task. Further, Study 4 showed that experimentally manipulating mentalizing led to greater mentalizing and memory-related activation. Further, successful encoding interacted with mentalizing, as complex HCH compared to basic HCH and all misses, led to greater activation in mentalizing and face memory regions, which is consistent with the hypothesis that increased mentalizing leads to deeper encoding and thus greater memory for faces encoded with complex mental state adjectives.

**Mechanisms for a Relationship**

**Depth of Processing.** Several possibilities exist here to explain how mentalizing can influence face memory. One possibility is that more extensive mentalizing leads to better face memory by virtue of increasing the depth to which people processed faces. One possible mechanism by which mentalizing about a face can lead to better memory for a face is by eliciting deeper processing of that face. Classic memory research shows that deeper processing of stimuli leads to greater memory for those stimuli. For instance, processing the semantic meaning of words leads to better memory than merely deciding if the words are written in all capital letters (Craik & Lockhart, 1972). In this way, more extensive mentalizing could elicit deeper processing than less extensive mentalizing and thus lead to better memory for that face. Supporting this explanation, judging if a person is likable or honest by their face leads to better memory than
making simpler judgments about a person, such as a person’s sex (Bower & Karlin, 1974; Wenger & Ingvalson, 2002).

Depth of processing also leads to differences in the neural activation involved with encoding items. More semantic processing of words leads to greater activation of the hippocampus and other memory-related structures (Wagner et al., 1998). Likewise, deeper encoding of faces also leads to more extensive neural processing in the medial temporal cortex (L. Bernstein, Beig, Siegenthaler, & Grady, 2002). Depth of processing thus leads to better memory not by eliciting a different set of neural processes but by eliciting more extensive activation in the same memory networks.

Depth of processing may improve memory for faces by virtue of increasing the amount of dimensions that are used in face-space to encode a face. As mentioned above, face-space is a series of dimensions in which faces meaningfully vary. Within face-space, faces can be encoded on only a few or many dimensions. Building upon the attractor field model, the more dimensions with which a face is encoded, the more refined attractor fields become, such that faces can be recognized more accurately with higher specificity (Corneille et al., 2007). Though it is untested if better memory for faces is a direct result of the number of dimensions with which a face is encoded in face-space, the idea that depth of processing may increase the number of dimensions on which faces are encoded in face-space is consistent with current models of norm-based encoding.

This evidence here indicates that mentalizing may increase the depth to which faces are remembered, thus leading to greater memory for these faces. This is supported by the evidence showing that this memory effect is based both on hits and misses as well as the neural evidence indicating increased activation in the hippocampus and FG for faces processed with complex
adjectives. Both of these are consistent with the depth of processing approach, which shows greater hits and fewer misses for deeper-encoded adjectives as well as increased activation in memory regions for deeper processing.

**Qualitative Differences in Memory.** A second explanation for the difference in memory for increased mentalizing is that mentalizing causes a qualitative difference in how faces are processed and remembered. There are two ways this can be the case. First, increased mentalizing can cause faces to be remembered using more associative memory versus semantic memory. Second, increased mentalizing can cause faces to be processed more configurally, causing them to be remembered in a qualitatively different way in face-space. Both of these explanations are discussed below.

One possible mechanism for the relationship between mentalizing and face memory is that mentalizing qualitatively changes the type of memory used to encode faces. This is consistent with dual-process memory models that distinguish between recollection and familiarity (e.g., Yonelinas, 1994). These models distinguish between the effortful process of recognition, which is a dichotomous judgment of remembering an item in context (either an item is recognized or not), and familiarity, which is a continuous variable of how familiar an item seems to be in memory. More complex mental state information may provide an associative context that a face can be encoded in and thus increase the recognition memory of faces, whereas encoding faces with only basic mental state information leads to a face to only be encoded using the face itself and thus is more likely to use familiarity-based encoding and recognition.

The evidence here provides some support of this explanation. The studies presented here consistently showed a behavioral pattern for more HCH and fewer HCM for complex versus basic adjectives, while the reverse pattern was true for LCH and LCM. This suggests a different
type of encoding for faces paired with complex mental states, which leads to more accurate judgments in high confidence decisions but not just a criterion shift where memory is improved for both high and low-confidence decisions. This is consistent with a distinction between recognition versus familiarly memory. Recognition memory is thought to be a high-threshold dichotomous judgment (i.e., requiring high certainty) of an item being remembered or not remembered (Wixted, 2007). This leads to the assumption that a series of items with higher levels of recognition memory would have a different pattern of memory judgments than a series of items with higher levels of familiarity. This pattern would lead to a greater number of HCH for recollection judgments versus familiarity judgments, which was found in these studies. However, if mentalizing contributed to face memory by virtue of increasing recollection, this memory effect should be present much more strongly for HCH than for LCH, as models of recognition memory models assume items that are not recollected are subjected to familiarity memory. Additionally supporting increased recognition is evidence that implicates the hippocampus in recognition memory and perirhinal cortex in association memory (e.g., Brown & Aggleton, 2001, Yonelinas, 2002). However, other observers have found hippocampal involvement in both recognition and familiarity memory (Jeneson, Kirwan, Hopkins, Wixted, & Squire, 2010; Kirwan, Wixted, & Squire, 2010; Squire, Wixted, & Clark, 2007), which would indicate that the evidence found in these studies would not support the assertion of increased recognition memory for faces encoded with complex mental state adjectives.

The evidence here comparing high and low confidence judgments suggests this is the case but a more precise way to examine whether increased recognition memory for mentalizing would be to examine mentalizing using confidence judgments and generate receiver operator characteristic plots (ROC) for memory for faces encoded with basic and complex mental state...
adjectives. These plots compare differences in hits and false alarms at different levels of confidence. This type of analysis uses a larger confidence judgment as to how confident a participant is in seeing if a face if new or old (instead of a 1 to 4 scale, using a 1 to 6 or greater scale). Recognition memory is associated with a curvilinear ROC whereas familiarity memory is associated with a linear ROC (e.g. Wixted, 2007; Yonelinas, 1997).

It is possible that increased depth of processing leads to better memory for faces by enhancing the degree to which configural information is used in encoding a face into memory. Both configural and feature-based information are used to remember faces (Tanaka & Farah, 1991; Wenger & Ingvalson, 2002). Thus it is possible that mentalizing induces increased levels of depth of processing by inducing a greater amount of configural processing. This may be because the configural nature of features is necessary in order to determine another’s mental state. However, this hypothesis is untested.

This depth of processing explanation is consistent with the face-space model discussed above, where depth of processing leads to more dimensions of a face being encoded in face-space. These findings suggest that mentalizing may moderate how faces are encoded in face-space, such that mentalizing increases the amount to which faces are encoded within face-space. Thus, faces that evoke greater mentalizing or are more extensively mentalized about are more deeply encoded by increasing the amount of dimensions that are used to encode a face, which then leads to greater face memory.

Attention. A third possibility for the relationship between mentalizing and face memory is that attentional processes may explain the increased memory for more extensive mentalizing. With these attentional explanations, mentalizing may not directly contribute to how faces are remembered. It is possible that increased mentalizing leads to increased attention paid to faces
which then leads to greater memory via greater attention paid to the stimulus. These possible attentional differences may be due to differences in words that were used for complex versus basic adjectives. Additionally, differences in the words used for complex versus basic mental states could cause differences in the attention paid to the faces. Complex mental states may be more intrinsically interesting than the basic mental state words and lead to increased interest which would causes participants to pay more general attention to the complex parts of the task. Further, the complex mental state words could have a different level of valence (positive versus negative) than basic mental states, which could cause participants to pay greater attention to them. Though it is impossible to rule out this explanation in the current data, it is not likely given the evidence presented here. First, there were few behavioral differences in the time it took to rate the faces in the encoding task, and those behavioral differences found indicated participants took longer to rate faces associated with basic mental state complexity adjectives. This is inconsistent with greater attention paid to complex mental state adjectives. However, since attentional explanations cannot be ruled out, additional studies are needed to ensure that this is not merely an attentional effect.

There are a couple of possible ways to test these attentional accounts. One direction would involve examining how differences in mental state complexity within a face affects the way faces capture attention. If faces that evoke more mentalizing also capture attention to a greater extent and if this degree of attention allocation then mediates the relationship between mentalizing and face memory, this would be a strong indicator for attention playing a critical role, then it is possible this may be an attentional effect. This would involve combining two different emotions (especially one of positive and one of negative valence) and assessing if this increases the mental state complexity attributed to faces. If these faces evoke deeper mental
states, then they should be better remembered. Another way to examine this is to vary the length faces are shown during encoding. If attention plays a role in this effect, it may be stronger for longer encoding durations (where participants’ attention is more likely to wander). Another future direction could involve selecting words that vary on mental state complexity but are highly standardized in order to match for how interesting the words are, the valence of the words themselves, and any other systematic difference in the words.

**Familiarity.** Another possible mechanism for the relationship between mentalizing and face memory may be familiarity with complex mental states. LaPlante & Ambady (2002) found better memory for mixed expressions versus pure emotional expressions, which they argued may be because mixed expressions are more common in everyday life and thus easier to remember. It may be the case that complex mental states are more common in everyday life, either in faces themselves (faces that evoke more mentalizing are more common) or in the words used to attribute to the faces (people are more often attributing complex versus basic emotions to faces). Thus it may be that these faces that evoke more complex mentalizing or the emotions that are marked by deeper mentalizing may be familiar. However, it is not likely this is an explanation here as the faces were counterbalanced and shown with both complex and basic adjectives, to remove any systematic difference in the faces themselves. Also, many studies show that distinctive faces are better remembered than basic ones (e.g. Valentine, 2001; Lee et al., 2000). However, it is still possible that if complex mental states are more common and familiar than basic ones, then they would lead to faces being easier to remember. The n-gram data reported in Study 1 suggests that there are no differences in the frequency of the use of the words used as complex versus basic mental states, which suggests that complex mental states are not more common. However, one way to directly examine this account is to examine how commonly
participants attribute complex versus basic mental states to faces. In Franklin and Adams (2010), we found that among ostensibly neutral faces similar to the ones used in this study, participants attributed mental state complexity below the midpoint of the scale to faces (with a mean rating of complexity of 3.2 out of 7 point scale, with a range from 2.5 to 4.7). This suggests that neutral faces, like the ones in the study here, evoke low overall levels of mental state complexity. However, it would be informative to examine how the trait of mental state complexity is attributed to faces and examine if the way this is attributed to faces is normally distributed. Further, participants can rate how common they think certain mental states are and see if that varies by how complex the mental states are or if individual ratings of the commonality of specific mental states predicts how easy it is to remember faces that are associated with that mental state.

**Neural mechanisms of a relationship**

Study 4 showed that networks involved in both encoding and mentalizing were involved in encoding complex versus basic adjectives. Face memory networks including the FG and hippocampus were robustly active to more mentalizing. Activation in these regions may reflect increased depth of processing related to increased mentalizing, as studies examining the neural correlates of depth of processing indicate that increased processing leads to more robust activation in the same memory systems that are active with more shallow processing (e.g., L. Bernstein et al., 2002; Wagner et al., 1998). Memory-related regions were also active in comparing HCH to misses, including the left hippocampus and parahippocampal cortex. These included the parahippocampal cortex as well as the left FG.

Mentalizing-related regions, including the left IFG and right temporal were also more involved in more extensive mentalizing. IFG activation may reflect increased simulation required
in processing more complex mental state activation, given the possible involvement in this region in the MNS. The temporal pole is also found in mentalizing studies, and is thought to represent a region involved with binding perceptual information with emotional content (Olson, Plotzker, & Ezzyat, 2007). This explanation is entirely plausible when considering deeper mentalizing, which is likely linked to more extensive emotional content associated with perceptual information.

Interestingly, however, comparing complex minus basic adjectives revealed robust activation in the mPFC, including both dorsal and ventral regions. This is interesting as although these regions have been found during socioperceptual mentalizing (e.g., Mitchell et al, 2005), they are more commonly involved in sociocognitive mentalizing (e.g. Amodio & C. Frith, 2006; C. Frith & U. Frith, 2006). Activation in the dorsal mPFC may reflect the use of theory-based person knowledge in assessing mental states while the ventral mPFC may reflect the use of perspective-taking and simulating others’ mental states. In addition, the correlation between these activations in the mPFC and hippocampus indicate that these processes may be actively contributing to how faces are being encoded.

It is interesting that complex versus basic mental state decoding led to differences in these regions as no studies to date have examined mentalizing about adjectives that are complex mental states versus basic ones. This data supports the contention that certain mental states may reflect or elicit more extensive mentalizing in order to decode them. As such, this supports the assertions made by infrahumanization work which suggests that we deny certain groups human-like mental states. The evidence here suggests additional processing is involved with complex mental states and the lack of this additional processing may be part of the reason for infrahumanization. However, this contention is speculative and requires further examination.
Most interesting, however, was the involvement of memory and mentalizing-related regions in the interaction between mentalizing and face memory. The activation in the mPFC, including both dorsal and ventral regions, underscores the relationship of those regions to successful encoding. In addition, the correlation between these regions and the hippocampus suggests that mPFC activation may be directly influencing hippocampal activation. This suggests that greater mentalizing modulates activity in mentalizing-related regions which then contributes to activation in memory-related regions. More research is necessary to examine if these regions actively contribute to and are functionally connected with memory-related regions; however the activation of these regions for complex HCH specifically likely indicates their involvement in successful memory.

**Future Directions.** Future research examining the neural relationship between mentalizing and face memory will involve assessing the functional relationship between regions involved in mentalizing and face memory. The data in Study 4 suggest such a relationship may exist, as Study 4 showed that activation in both memory and mentalizing was present for the interaction between mentalizing and face memory. Further, activation in mentalizing and memory-related regions was correlated assessing a correlation between subjects. However, a more powerful test of this would be to examine if activation in mentalizing-related regions on an individual level both was correlated with activation in memory regions and predicted activation in memory-related regions.

The correlations I report in Study 4 do not describe the direction of causality; however effective connectivity techniques can be used in order to assess this relationship. One technique that can be used to assess this relationship is unified structural equation modeling (Gates, Molenaar, Hillary, Ram & Rovine, 2010; Gates, Molenaar, Hillary, & Slobounov, 2011).
method is a combination of structural equation modeling and vector auto-regression in order to assess causal connectivity paths in brain data while accounting for the temporal relationships in fMRI data. This method also allows for the examination of moderators of these relationships which will allow for the examination of if depth of mentalizing moderates the relationship between mentalizing and face memory regions.

**Implications**

**Models of Face Processing.** The interaction between mentalizing and face memory found here offers insight into how faces themselves are processed. As discussed above, faces convey a variety of sources of social information. Classic face processing models (e.g., Bruce & A. Young, 1986; Haxby, Hoffman, & Gobbini, 2000) assert that the decoding of these different forms of information is dissociated to some degree. Specifically, these models propose a dissociation between decoding facial identity versus expressive information. However, more recent research has shown interactions between the decoding of identity and expressive information from faces can interact with each other (e.g. Adams et al., 2010; Calder & A. Young, 2005; Zebrowitz, 2006). The relationship between mentalizing and face memory examined herein supports this latter assumption, as I argue here that mentalizing (which is based on expressive information derived from faces) moderates to face memory. The evidence here supports a relationship between the two processes both behaviorally and neurally.

The interaction between mentalizing and face memory can also contribute to understanding why expressive information influences the processing of face memory. Many studies have shown that varying expression affects how identity is processed, yet these studies do not indicate why expression affects identity processing. Building upon the ecological models of face perception discussed above, Adams and colleagues (2010) propose a compound social cue
approach to face perception, asserting that the different sources of information contained within a face share underlying social messages and that face processing is disposed to decode these social messages. Applying this approach to the interaction between mentalizing and face memory, it is possible that mentalizing influences face memory more strongly depending on the social cues present in mental state information. For instance, mentalizing about a mental state that is more important to one’s social context should lead to a greater effect on memory than mentalizing about a mental state that is less important to one’s social context.

**Group Dynamics and Face Memory.** One of the most replicable effects in face memory research is that which shows better memory for faces of one’s own race. As discussed above, the own-race bias extends to faces of one’s own social group, including one’s age, gender, and to arbitrarily assigned social groups. These findings indicate the own-group bias is not due to perceptual expertise associated with the faces but is more likely due to social categorization involved with the faces (Hugenberg & Sacco, 2008). However, it is less clear how social categorization leads to impairments in face memory for outgroup members.

One possibility is that we mentalize less about social outgroups and this deficit in mentalizing leads to deficits in memory for outgroup faces. If this is the case, the processes that are thought to lead to less mentalization for outgroup faces may be the processes that drive the own-group bias. This explanation remains untested, however. If differences in mentalizing do indeed lead to the own-group bias, then mentalizing should mediate the own-group bias in a face memory task. Additionally, forcing mentalizing (by using a paradigm similar to the complex mental state adjective paradigm here) should reduce the own-group bias.

A second possibility is that the relationship between mentalizing and face memory is different for same-group and other-group faces. In this case, same-group faces would show
increased memory for increased mentalizing whereas other-group faces would not show a relationship between mentalizing and face memory. One study supports this contention. In a memory task, the race of faces moderated the relationship between mentalizing and memory. The degree to which faces evoked greater mentalizing predicted memory (i.e., mental state complexity) for faces for White faces but not for Black faces in White individuals (Franklin & Adams, in preparation). This indicated that mentalizing was related to memory for only same-race faces. Thus, it is possible that mentalizing leads to deeper encoding in face-space for same-race faces because there are fewer dimensions with which to encode other-race faces. Further, if mentalizing increases recollection memory, it is possible that mentalizing leads to greater memory for same-race faces because same-race faces more often encoded using recollection memory. This is supported by evidence that shows greater recollection versus familiarity memory for same-race faces (Marcon, Susa, & Meissner, 2009).

Future research is necessary to evaluate these possibilities. One potential future direction is examining face memory using a similar paradigm as used here but using different racial or social groups. If mentalizing contributes to memory equally for same-group versus other-group faces, then mentalizing about faces using complex mental state adjectives should lead to improved memory for faces regardless of face race. However, if the effect is only present for same-race faces, then complex mental state adjectives should only lead to an increase in memory for same-group faces.

The relationship between mentalizing and face memory can also help in the remediation of the own-race bias. If differences in mentalizing cause the own-race bias in face memory, then interventions intending to ameliorate this bias should be targeted at addressing mentalizing rather than face memory per se. These interventions can involve ways to increase the degree to which
participants attribute complex mental states to outgroups. Interventions may involve increasing contact with outgroup members, as increased contact decreases infrahumanization (Tam et al., 2007). Further, manipulations to take the perspective of outgroup others may also increase mentalizing for outgroup others. Increased perspective-taking of outgroup members reduces prejudice associated with outgroups (Vescio, Sechrist, & Paolucci, 2003), improves feelings toward outgroups (Batson et al., 1997) and increases the use of non-stereotype information when judging others (Galinsky & Moscovitz, 2000). These all suggest that manipulations to take the perspective of others could increase mentalizing in regard to outgroup others and possibly reduce the own-race bias.

**Conclusions**

The work in this dissertation provides direct evidence for the involvement of mentalizing in face memory. This is true in considering both the behavioral and neural implications of this relationship. As such, this work offers the foundation for understanding why we remember who we do and how reading others’ contributes to this process. Further, this work offers additional understanding as to why we remember some people and not others. Group-based differences in memory may be driven by differences in how we read others. We more extensively mentalize about those in our social groups and those who we find more relevant, as such it is not surprising that we remember them better.

This work also has the potential to contribute to how social psychology affects our visual system. Social influences are critical in the development of the visual system and therefore it is not surprising that we see others differently based on social context. Additionally, the mechanisms that we use to see the world are a product of and are disposed toward the social contexts that we are formed within. The relationship between mentalizing and face memory is
another example of this, showing that the social meaning that we read from others impacts how we process and remember them. As such, this provides more evidence of social influences being critical in basic mental systems, including perception and memory. In order to fully understand these systems, it is necessary to consider how social contexts moderate their processing and how those systems themselves are built in order to negotiate the necessities of a complex social world.
Appendix 1: List of mental state words used for pre-rating task.

<table>
<thead>
<tr>
<th>adjective</th>
<th>adjective</th>
<th>adjective</th>
<th>adjective</th>
<th>adjective</th>
</tr>
</thead>
<tbody>
<tr>
<td>accusing</td>
<td>Defiant</td>
<td>horrified</td>
<td>proud</td>
<td>reasserting</td>
</tr>
<tr>
<td>agitated</td>
<td>depressed</td>
<td>hostile</td>
<td>reflective</td>
<td>regretful</td>
</tr>
<tr>
<td>angry</td>
<td>despondent</td>
<td>impatient</td>
<td>imploaring</td>
<td>relaxed</td>
</tr>
<tr>
<td>annoyed</td>
<td>disappointed</td>
<td>incredulous</td>
<td>indecisive</td>
<td>relieved</td>
</tr>
<tr>
<td>anxious</td>
<td>disinterested</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>apathetic</td>
<td>dispirited</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>apologetic</td>
<td>distrustful</td>
<td>indifferent</td>
<td>resentful</td>
<td></td>
</tr>
<tr>
<td>apprehensive</td>
<td>dominant</td>
<td>indignant</td>
<td>sad</td>
<td></td>
</tr>
<tr>
<td>arrogant</td>
<td>doubtful</td>
<td>inspired</td>
<td>sarcastic</td>
<td></td>
</tr>
<tr>
<td>ashamed</td>
<td>dubious</td>
<td>interested</td>
<td>satisfied</td>
<td></td>
</tr>
<tr>
<td>assertive</td>
<td>eager</td>
<td>intrigued</td>
<td>scared</td>
<td>serious</td>
</tr>
<tr>
<td>baffled</td>
<td>elated</td>
<td>jealous</td>
<td>skeletal</td>
<td></td>
</tr>
<tr>
<td>bewildered</td>
<td>embarrassed</td>
<td>joyful</td>
<td>sleepy</td>
<td></td>
</tr>
<tr>
<td>bored</td>
<td>encouraging</td>
<td>lazy</td>
<td>stern</td>
<td></td>
</tr>
<tr>
<td>brave</td>
<td>energetic</td>
<td>lonely</td>
<td>surprised</td>
<td></td>
</tr>
<tr>
<td>calm</td>
<td>excited</td>
<td>loving</td>
<td>suspicious</td>
<td>sympathetic</td>
</tr>
<tr>
<td>cautious</td>
<td>exhausted</td>
<td>mad</td>
<td>tentative</td>
<td>terrified</td>
</tr>
<tr>
<td>charged</td>
<td>fascinated</td>
<td>melancholy</td>
<td>thoughtful</td>
<td>threatened</td>
</tr>
<tr>
<td>cheerful</td>
<td>fearful</td>
<td>motivated</td>
<td>nervous</td>
<td>tired</td>
</tr>
<tr>
<td>comforting</td>
<td>frenzied</td>
<td>overjoyed</td>
<td>pensive</td>
<td>uneasy</td>
</tr>
<tr>
<td>concerned</td>
<td>frightened</td>
<td>offended</td>
<td>perplexed</td>
<td>unhappy</td>
</tr>
<tr>
<td>confident</td>
<td>frustrated</td>
<td>overjoyed</td>
<td>playful</td>
<td>upset</td>
</tr>
<tr>
<td>confused</td>
<td>glad</td>
<td>panicked</td>
<td></td>
<td>worried</td>
</tr>
<tr>
<td>contemplative</td>
<td>gleeful</td>
<td>pensive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>contented</td>
<td>guilty</td>
<td>perplexed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>curious</td>
<td>happy</td>
<td>pleased</td>
<td></td>
<td></td>
</tr>
<tr>
<td>deciding</td>
<td>hateful</td>
<td>pleased</td>
<td></td>
<td></td>
</tr>
<tr>
<td>decisive</td>
<td>hopeful</td>
<td>preoccupied</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Studies 1, 2, and 4 each indicated a pattern showing greater memory for faces encoded with complex mental state adjectives versus basic mental state adjectives. However, each of these studies showed a different pattern in regard to if this effect is driven by a greater number of hits for faces encoded with complex mental state adjectives or a greater number of misses for faces encoded with basic mental state adjectives. The difference in memory in Study 1 was driven by more HCH for complex faces while in Study 2, this pattern was driven by more HCM for basic faces. In Study 4, a marginally greater number of HCH for complex faces was present as well as a greater number of HCM for basic faces.

In order to assess if the effects for HCH and HCM in the behavioral studies were significant across all three studies, I combined the results meta-analytically, using the significance level for each t-test using the approach detailed in Rosenthal and Rosnow (1984). Each of these meta-analyses were performed by deriving Z-values for the significance of the effects of each study and weighting these effects from each study proportionally by the number of people in each study. Using this method, the effects comparing complex minus basic adjectives were significant across the studies for HCH ($Z = 3.12, p = .0018$), HCM ($Z = 3.51, p = .0004$), and the difference for HCH minus HCM ($Z = 4.23, p = .00002$).
References


Calder, A. J., Lawrence, A. D., Keane, J., Scott, S. K., Owen, A. M., Christoffels, I., & Young, A. W. (2002). Reading the mind from eye gaze. *Neuropsychologia, 40*(8), 1129-1138. doi:16/S0028-3932(02)00008-8


doi:16/j.neuropsychologia.2005.07.001


doi:10.1037/0033-2909.128.2.203


Vita – Robert G. Franklin, Jr.

The Pennsylvania State University, University Park, PA
Ph. D., Psychology, May 2012
Dissertation Advisor: Dr. Reginald Adams

The College of William and Mary, Williamsburg, VA
B.S. (Magna cum Laude), Psychology; History, May 2006

Manuscripts


Chapters