present recent neuroscientific evidence that fundamentally undercuts the problem.

The Problem of Other Minds

The philosophical problem of other minds poses two questions. First, how is it possible to know that others think like I do? Second, how is it possible to be able to understand the mental states of others when they exist in minds of other people, hidden from view behind the cranium? John Austin frames the problem simply: "How do we know that another person is angry? . . . Do we ever know?"[44] We often take as a given that all humans experience states of mind similar to what we experience: pain, pleasure, fear, and so forth. Yet, what justifies that certainty? Assuming that we can be justified in thinking that others have similar minds to our own, how do actors, through the course of interaction, come to understand those states in others? The problem of other minds occurs precisely because there is a large disconnect between the access we have to our own inner experience and our access to the experience of others. We can often tell when we are in serious pain, yet we do not have access to the mental states of other humans to tell whether or not *they* are in serious pain; maybe they are being particularly stoic or faking the pain. This creates an information asymmetry between what we know about ourselves and what can be known about others.

To illustrate how this problem manifests itself in international politics, consider again the example of Alter/Ego above. Wendt's "mirror theory of identity formation"[45] has been critiqued from different angles, but the most forceful criticism has been made from the perspective of the other minds problem. For Dale Copeland, Wendt's theory of intention suffers from three "serious flaws."[46] First, while some states, perhaps even the majority, have a pretty good sense of the intentions of others today, history suggests that uncertainty has been a significant problem in the international system, accounting for war and widespread conflict. Second, the problem with the gestures that Alter/Ego engage in is that "behavior does not speak for itself."[47] Linking this criticism to the problem of other minds, Copeland argues that "[b]ecause leaders cannot observe directly what the other is thinking, they are resigned to making inferences from its behavior."[48] This critique amounts to a claim that the social learning aspects of the theory remain undertheorized and underspecified. At the end of the day, for Copeland, Wendt's

[44] Austin 1979, 76. [45] Wendt 1999, 407.
[46] Copeland 2006, 11. [47] Copeland 2006, 12. [48] Copeland 2006, 12.

formulation does not provide us with a basis for distinguishing sincere gestures from insincere ones.

Relatedly, Zehfuss and others have pointed out that while Wendt acknowledges the importance of language, the actors in the Ego/Alter interaction do not speak, rather, they signal to each other through behavioral actions.[49] According to Zehfuss, "a social act consists in sending a signal, interpreting it and responding on the basis of interpretation. A 'conversation of gestures' develops."[50] Ego/Alter are unlikely to share a common language, since they do not acquire characteristics from participating in a society. Therefore the symbolic interactionist framework, from which Wendt seeks to draw, and for which shared language may be vital, loses one of the core ways in which actors communicate. More generally, Zehfuss and Copeland agree that the "actors cannot communicate about their behavior; they communicate *through* their behavior."[51] On what basis can Ego or Alter's behavior be judged in a first encounter?

Finally, third, Copeland notes that there is a strategic aspect to interaction that Wendt does not discuss, namely the problem of impression management. The concern here is that a strategic actor will have strong incentives to make conciliatory gestures in order to deceive the other. "When we consider the implications of a Hitlerite state deceiving others to achieve a position of military superiority, we understand why great powers in history have tended to adopt postures of prudent mistrust."[52] Equally problematically, even if for some reason a particular actor could be trusted, the diachronic problem of future intentions remains ever present. This means that not only are deceptive intentions unknowable because of the problem of other minds, the problem of intentions persists into the future as well; solving the synchronic problem does not necessarily solve the diachronic problem. The philosophical problem of other minds thus constitutes the foundation upon which the problem of intentions is built.

Research into this philosophical problem falls into two theoretical camps and is broadly referred to as "theory of mind."[53] The first, termed "theory-theory" (TT) suggests that we rely on theories of mental states

[49] Zehfuss 2002; Booth and Wheeler 2008; Wheeler 2018. [50] Zehfuss 2002, 48.
[51] Zehfuss 2002, 49. [52] Copeland 2006, 13.
[53] The theory of mind literature is very diverse and too broad to be fully explored here. Each of the perspectives that follow, including TT and ST, has engendered significant literatures. There are also recent phenomenological accounts of social perception that reject many of the assumptions of the theory of mind literature, namely that mental states are not directly observable. On this view phenomenologists argue that we can indeed "directly perceive" the mental states of others through their behaviors. When we see people behaving in particular ways we are, in essence, seeing their minds in action. See, for example, the debate between Zahavi 2008; Krueger 2012; Bohl and Gangopadhyay 2014. See also Wendt 2015, 232 which engages with this literature from a quantum perspective.

that we have derived throughout life.[54] In this view we hold personal theories of psychology that we use to infer the mental states – including emotions, beliefs, intentions, desires, and so forth – of others. These inner theories allow us to understand the present actions of people as well as predict their future actions. A simple example illustrates this perspective. A person witnessing someone sitting in a restaurant crying, holding their face in their hands, may well infer from their own life experiences, such as being taught by a parent, that the crying individual is sad. TT suggests that in these instances we are essentially applying a folk psychology theory, honed and updated since childhood, of how others think and behave to the situation. The mechanisms of understanding others at work in this example are reasoning and observation, a form of "scientific theorizing" about the intentions of others.[55] While one can never be certain of one's reading of another, after all the person in the restaurant may be crying out of happiness or suffering from allergies, the TT perspective provides a probabilistic and approximating route to understanding the mental states of others; and, context and situational factors provide clues to help us refine our predictions.

Recently a second perspective on the problem of other minds has garnered significant attention by philosophers, psychologists, and neuroscientists. This approach, "simulation theory" (ST), proposes that we come to understand the mental states of others and their intentions not through theorization and approximation, but rather through simulation. ST stems from skepticism about TT's claim that individuals possess a database complete with vast theories of social behavior, which would not only be biologically expensive but seemingly impractical from an evolutionary standpoint. Rather, ST proponents suggest that understanding the mental states of others involves activating mental processes that, if actually carried out, would produce similar behaviors. In this sense they are "pretend" versions of the mental states of others that occur in one's own mind.[56] This simulation, for many neuroscientists, is the basis for a specific form of empathy, the ability for individuals to know what it feels like to be someone else and in another's position both analytically and emotionally.[57] The ST insight is that we often do not need to theorize, either scientifically or through folk psychology, about what someone else

[54] For an excellent synopsis of the theory of mind literature, see Goldman 2006 as well as Goldman and Jordan 2013.
[55] Wendt 2015, 231. [56] Goldman and Jordan 2013.
[57] The neuroscientific notion of empathy does not carry the positive normative bias that is often implied in common colloquial usage, as I mentioned above. Here the context of empathy is biologically self-serving and distinct from sympathy, which I conceive as

is experiencing – we simply know because we can experience it automatically for ourselves.

For example, individuals who watch a video of a spider crawling on the back of another human being often report that they get the "chills" watching the video. The participants do not need to theorize or think about what the experience feels like – they know, because they are simulating, or mirroring, in their own bodies, in real time, what the experience actually is. Similarly, individuals watching a horror movie in a theater often report physiological emotional changes, such as accelerated heart rates, as they experience the fear that the character in the film is experiencing. ST proponents point out that in these instances it seems unlikely that the individuals involved are *theorizing* about the experience others are going through; rather, they seem to be *simulating* the experience for themselves.

ST proponents also argue that this type of empathy, the very quick simulation of experience of others, is precisely the type of ability that is required in social life, where instantaneous evaluation of the social environment and quick decisions are necessary to get through life. As de Vignemont and Singer argue, "[e]mpathy might enable us to make faster and more accurate predictions of other people's needs and actions and discover salient aspects of our environment."[58] Much of what humans do on a day-to-day basis, indeed what social life requires, as Marco Iacoboni, a leader in investigating simulation and the neurological structures that support its processing argues, is the ability to quickly make important inferences about the actions and accompanying mental states of others:

> One glance at my eleven-year-old daughter at the breakfast table tells me to tread carefully and sip my espresso in silence. When a colleague reaches for a wrench in the laboratory, I know he's going to work on the magnetic simulation machine, and he's not going to throw his tool against the wall in anger. When another colleague walks in with a grin or a smirk on her face – the line can be fine indeed, the product of tiny differences in the way we set our face muscles – I automatically and almost instantaneously can discern which it is. We all make dozens – hundreds – of such distinctions every day . . . Nor do we give any of this a second thought.[59]

We know these things because we *simulate* them for ourselves, in our own brains, in automatic fashion. From a ST perspective, a smirk or grin are not theorized from past experience, but are simulated in present experience. Put another way, if social life required constantly checking behavior against a database of experiences and theorizing about the motivations of that behavior, it would be an exhausting and comparatively slow way of understanding others. Social life requires quick judgment. This does not mean that TT does not have a place in understanding the mental states of others, but ST provides another account, and arguably a more plausible one given the complexities of social life, of how intention understanding might occur.

One way to understand the key differences between TT and ST is to consider the activities associated with each proposed process by which we understand others. At root, TT is a detached theoretical activity, with emphasis on individuals observing others and interpreting through the use of inner databases and folk-psychology perspectives, not unlike the ways states interpret the intentions of other states in IR theory. ST, on the other hand, implies a much more *direct* attempt to replicate or mimic the mental states of others in one's own mind. ST is not detached in the sense that there is a "correspondence between the mental activity of the simulator and target,"[60] and not theoretical in the sense that individuals are not using honed theories of psychology to draw inferences about intentions. Some have suggested that the ST perspective implies something of a "shared circuit" or "coupling" between individuals where the distance between individuals is minimized.[61] Since the minds of individuals are actively experiencing what the other is experiencing, there is a sense in which the two become congruent and the perception of others' mental states is done automatically and directly. In that sense we often do not need a theory at all to understand the mental states of others.[62]

Importantly, these very different perspectives on understanding others imply important empirical differences that can be investigated. If ST is correct, for example, we should expect to find evidence of mental mimicry, the literal simulation of the other, in the body. Since TT does not claim to have a mimicry component to its process, evidence of mental mimicry would support ST. Lastly, while the two theories of mind are delineated analytically, it is likely that both are involved, at some level, in understanding others. I will return to this point in more detail later.

an emotional response of sorrow or concern for the condition of someone else. In this way empathy can be entirely egoistic or self-serving: successful chess playing arguably requires empathy in order to anticipate an opponent's moves but does not necessarily involve sympathy for the opponent. See also Bloom 2016.

[58] de Vignemont and Singer 2006, 440. [59] Iacoboni 2008, 3–4.

[60] Gallese and Goldman 1998, 497. [61] Hurley 2008.

[62] This is not to say that theorizing never comes into play about intentions. There are undoubtedly situations where we do not have a clue as to what someone else is thinking and therefore require theorizing about them. Or, perhaps more common, as I will discuss below, theory comes into play *after* interpersonal social cognition as part of reflection back on the interaction that took place.

While these different models of understanding actors exist, IR theory has largely been developed from a TT perspective, as has positivist social science in general. As behaviorism declined in the 1950s and 1960s, cognitive scientists looked inward rather than outward for explanations of mental states. Cognitivists argued that individuals possessed inner representations of the world and that the word *theory* described what these representations constituted and how they operated. If we knew how such representations and symbols were created, we could understand and deduce them through probabilistic rules. The differing views of behaviorists and cognitivists become clear in the following basic situation: a person leaves the house carrying an umbrella. A behaviorist might suggest "X believes that it is raining, if X takes an umbrella when X goes out." This logic works only if the rain believer does not like getting wet. A person might not take an umbrella, yet still believe that it is raining. Behavior does not necessarily reflect a person's underlying mental state. Cognitivists attack this problem by suggesting it is not solely behavior that matters, but that inner beliefs and desires explain mental states. There is no straight line from behavior to mental state. We invoke a theory of desire plus belief, in conjunction with behavior, to understand mental states. Therefore, from a TT perspective one might possess a folk theory about individuals who enjoy walking in the rain and those who do not, a theory constantly updated by experience.

Realist, liberal, and constructivist perspectives invoke, either implicitly or explicitly, these cognitivist perspectives.[63] The traditional rationalist model of desire + belief = action,[64] for example rests on this TT perspective of attributing mental states to actors in order to understand their behavior. Just as individuals engaged in a social interaction interpret the presentation of the other's behaviors, signals, and so forth against theories they hold of behavior, states do the same thing. When realists argue that states interpret costly signals as a way to reveal intentions, it means that states are using their own folk theories of signal sending/receiving to interpret the behavior and intentions of other states. Liberal institutionalists argue that states have a theory of state reputation that they use to interpret the actions of states embedded in institutions. Finally, constructivists, by suggesting that states can understand the meaning of the actions of others through identity, are arguing that states possess an innate theory of identity, and how it works, that has been honed through time and iterative interactions with salient others.

[63] Tetlock 1998.

[64] Fearon and Wendt 2002. It should be noted that there is some disagreement amongst the rationalist school regarding how accessible the beliefs and desires may be, independent of outcomes. See Frieden 1999.

What this suggests is that there has been a parallel between, on the one hand, the ways some psychologists and philosophers of mind approach the problem of other minds, and on the other, the ways in which IR theories have approached the problem of intentions. Both have largely adopted a TT perspective, where each side, be it an individual or a state, in an interaction responds to presentations of the other's signals, identity, and so forth, and deduces an explanation of behavior through a theoretical or (non-pejorative) folk-psychology perspective.[65] The incoming data is then subconsciously checked against the database of experiences that the individual or state has amassed over time to come to a conclusion about the present actions and future intentions of others. Put simply, IR theory is TT-dominated and, as of yet, has not incorporated the ST perspective into its theorizing. It is therefore not surprising that extant IR scholarship has viewed the problem of other minds as something of an unsurpassable problem: from a folk psychology or scientific theorizing (i.e. TT) perspective, other minds really are unknowable. And crucially this amounts to something of a wager. If the TT perspective is correct, and the only way to know intentions is to theorize about them, then this places the pessimists on firm ground.

In the next section, I review efforts by psychologists, neuroscientists, and philosophers of mind to understand when, and under what conditions, these two conceptions of theory of mind might be operating. In particular, face-to-face interactions have become the topic of very intense study when it comes to how we understand the mental states of others, in part because new findings suggest that simulation is abundant in face-to-face interactions, which supports a rich set of findings and arguments in social theory and psychology that face-to-face is a unique form of social interaction. It is to that discussion that I will now turn in order to construct a ST-based theory of intention understanding in world politics.

Face-to-Face as a Unique Simulation Opportunity

A Typology of Behavioral Face-to-Face Effects

Scholars in a variety of domains have long understood that face-to-face interaction is a particularly rich information environment that allows

[65] The term *folk psychology*, despite its unfortunate and misleading name, refers to a sophisticated and highly complex theory about the interaction of desires and beliefs, despite carrying an unfortunate label that implies amateurish post hoc rationalization of the crudest sort. In fact, "folk-psychology" is quite difficult to undermine theoretically. See Horgan and Woodward 1985.

individuals to communicate on multiple levels, including dialog, nonverbal actions, and emotions. Erving Goffman famously theorized this in his 1959 treatise on dramaturgical theory applied to social interactions, noting the importance of these various information levels in conveying overt and subtle clues:

> When an individual enters the presence of others, they commonly seek to acquire information about him or to bring into play information about him already possessed . . . many sources of information become accessible and many carriers (or "sign-vehicles") become available for conveying this information. If unacquainted with the individual, observers can glean clues from his conduct and appearance which allow them to apply their previous experience with individuals roughly similar to the one before them or, more important, to apply untested stereotypes to him. . . . [T]he "true" or "real" attitudes, beliefs, and emotions of the individual can be ascertained only indirectly, through his avowals or through what appears to be involuntary expressive behavior.[66]

One of Goffman's contributions was to highlight the ways in which face-to-face interactions are rich environments of information sharing, including information that may not want to be shared, but *is* through involuntary expression. Robert Jervis applied this insight to international politics and introduced the importance of indices, signals that bring with them some sign of inherent credibility, such as the expressive behavior in interpersonal interactions that may be difficult to conceal.[67] The insight here is particularly important for understanding the value of personal face-to-face interactions in diplomacy: since expressive behaviors are often involuntary and can reveal "true" or "real" mental states, the interaction itself brings with it a certain credibility that may elevate it to an index.[68] As Jervis notes, these are ultimately perceptual in nature, so they may be prone to error, but nevertheless are important to the actor because he or she views them as credible.

This framework has been applied specifically to the problem of intentions. Keren Yarhi-Milo argues that there are a variety of ways that states

[66] Goffman 1959, 1.

[67] Jervis distinguishes an index from a signal in the following way: "*Signals* are statements or actions the meanings of which are established by tacit or explicit understandings among the actors . . . In contrast to signals, *indices* are statements or actions that carry some inherent evidence that the image projected is correct because they are believed to be inextricably linked to the actor's capabilities or intentions." Jervis 1970, 18.

[68] Many authors implicitly endorse this perspective, even if they do not explicitly invoke the Goffman/Jervis perspective. Jenny Edkins (2015, 1) opens her book on face politics by noting "[w]e pay attention to reading each other's faces – reading people's moods, personalities and origins into their facial appearance. We search for clues as to who the person opposite us may be and what they may be thinking . . . face-to-face encounters are seen as potentially more honest and open than those conducted by other means."

may attempt to assess the intentions of other states, including interpreting capabilities, current and past behaviors, as well as the subjective experiences of individual decision-makers. In this latter process, individuals "tend to rely on kinds of information that are particularly *vivid* . . . [referring] to the 'emotional interest of information, the concreteness and imaginability of information, and the sensory, spatial, and temporal proximity of information.'"[69] Face-to-face interactions are particularly good sources of vivid information, for the reasons Goffman and Jervis allude to, namely the difficulty of manipulation. For Yarhi-Milo, decision-makers often rely on this information, even if it is costless, to derive conclusions about the intentions of others. The vividness theory is still in the realm of approximating and theorizing intentions since it relies, at root, on the "selection and interpretation of signals" in the face-to-face interaction, though the insight that face-to-face encounters provide credible information even if the interaction is costless, is of critical importance. One drawback of the vividness hypothesis is that it is difficult to predict *a priori* what types of information decision-makers will find to be vivid, since it is not clear what mechanism makes information vivid. That is, the emotional interest of information or the imaginability of information is highly subjective and largely dependent on individual personalities and prior-held beliefs, and therefore it becomes difficult in any one particular instance to determine whether decision-makers will find a face-to-face encounter relevant from an intention understanding perspective.

In addition to signaling and information gathering, face-to-face contact has been of particular interest to IR scholars for some time because of the perceived benefits it produces with respect to humanization, emotional bonding, and, potentially, reduction of prejudice. Roland Bleiker's study of reconciliation in Korea, between North and South Koreans, highlights the importance of face-to-face encounters in "removing entrenched stereotypes and threat images," contributing "to the creation of a culture of reconciliation, which is an essential – and so far lacking – precondition for a significant diplomatic breakthrough on the Korean peninsula."[70] The Sunshine Policy by South Korea in the late 1990s and early 2000s seemed to internalize this logic, attempting to increase face-to-face contacts among ordinary citizens in an effort to, as Bleiker puts it, find a *sustainable* way to accept difference between the two Korean identities without appealing to violence.[71] While the

[69] Yarhi-Milo 2014, 3.

[70] Bleiker 2004, 152.

[71] Bleiker 2005, *xlii*; also Bleiker 2004, 153–55 and Bleiker 2010, 235–55 connecting the North Korea case with literature on "sustainable diplomacies" which seek to focus on

precise mechanism by which face-to-face encounters work to reduce threat images and increase constructive dialogue remains unexplored, Bleiker's perspective resonates with a large research paradigm in sociology on the "contact hypothesis," which provides relatively clear conditions, such as the existence of equal status and common goals among individuals, that foster improved relations between groups.[72] Finally, a critical product of the humanization process may be the growth of trust between actors. I address trust more directly in the concluding chapter of the book, but at this point it is worth noting that scholars have identified face-to-face contact as that which "makes possible the growth of trust between . . . key actors."[73] Precisely how, and under what conditions, this trust development occurs remains in dispute, though it is likely the case that face-to-face interactions play a strong role in building trust between adversaries at the interpersonal level.[74]

Face-to-face interactions have also featured prominently in both the relational and practice "turns" in IR theory. From a relational perspective, IR has embraced a substantialist tendency, the conflation of an object with the outcomes of its actions.[75] As Patrick Jackson and Daniel Nexon argue, this tendency or bias assumes particular entities existed prior to interaction, "or that entities are already entities before they enter into social relations with other entities rather than being created and shaped by the process of interaction . . . "[76] A relational perspective, in contrast, suggests that we take seriously the notion that it is social *relationships* that make up world politics. States, for example, do not come into being *qua* states and then have interactions with other states; rather, states are made through relations with both other states and non-state actors. With reference to diplomacy, Rebecca Adler-Nissen argues that " . . . most diplomats know . . . that world policy is deeply relational. Their job is to make those relationships 'work.'" Face-to-face interactions, as the most ubiquitous form of diplomatic interaction in many environments, often are the starting point for relational diplomatic studies.[77]

Practice theorists, who highlight the quotidian processes of building and maintaining relationships, have similarly highlighted the importance of face-to-face interactions in structuring diplomatic practice. Pouliot,

"long-term reconciliation and/or coexistence of competing entities and ways of living." Constantinou and Der Derian 2010, 2. On reconciliation more generally see Hutchison and Bleiker 2013.

[72] There is a very large literature on the contact hypothesis, beginning with Allport 1954. For more recent developments, see: Pettigrew and Tropp 2006; Pettigrew and Tropp 2013. While there are obvious connections between the theory in this book and the contact hypothesis, my interest is specifically on intention understanding rather than prejudice reduction.

[73] Wheeler 2013, 479. [74] See Wheeler 2018. [75] Holmes and Rofe 2016, 6.

[76] Jackson and Nexon 1999, 293. [77] Adler-Nissen 2015, 286.

for example, argues that it is the "thick face-to-face environment" of multilateral diplomacy, specifically in relation to UN Security Council diplomacy, that provides structures of appropriate behavior, calling to order failure to conform with norms of the game.[78] Permanent representation follows a similar logic. "The small world of permanent representation, feeding on regular and face-to-face interaction, heightens the probability that breaching the rules will spark an immediate social reaction."[79] Mitzen finds similar behavior in the forum effects of face-to-face in the Concert of Europe.[80] Face-to-face interactions thus serve to aid the creation and maintenance of routine and expected behaviors.

Parallel to these developments in IR and sociology, psychologists developed face-to-face interaction research programs, investigating in great detail at the behavioral level how interpersonal social interactions are conducted. As Wilson wrote in 1972, the study of human communication "surely must be one of the most important of all emerging scholarly fields," with the importance of face-to-face as a discrete type of interaction serving as a cornerstone.[81] Duncan and Fiske systemized this face-to-face research and cataloged the early effects of face-to-face, including empirical studies in the ability to pick up on emotions, the ability to send and receive signals through nonverbal communication, gender differences in the ways face-to-face interactions are conducted, the tendency of individuals to mimic each other and so forth.[82] The development of these studies in the 1970s, both in IR and psychology, laid the foundation and inspiration for sophisticated experimental designs that could both delineate what it is precisely about face-to-face that makes it unique as well as predict under what conditions its effects obtain.

One of the first disciplines to investigate this empirically was business administration and organizational behavior, where understanding the dynamics of negotiation, particularly face-to-face negotiation, may lead to more beneficial (and profitable) outcomes. Early experiments in the 1980s and 1990s demonstrated that face-to-face negotiation, relative to negotiation conducted via computer terminals, for example, increased the prospects for cooperation and minimized the pernicious aspects of the prisoner's dilemma.[83] One of the reasons hypothesized for this was that face-to-face allows for the reading of emotions, which serves to "*provide information* to help individuals know others' emotions, beliefs, and intentions."[84] Morris and Keltner conclude that "emotional expression helps individuals solve one of the basic problems of social

[78] Pouliot 2016, 61. [79] Pouliot 2016, 125. [80] Mitzen 2013.

[81] See, for example, Duncan and Fiske 1977. [82] Duncan and Fiske 1977.

[83] Raiffa 1982. [84] Morris and Keltner 2000, 16.

interaction: reliably knowing the thoughts and feelings of others."[85] Put another way, they are alluding to the possibility of face-to-face interactions undercutting the problem of other minds through emotional signaling.

Face-to-face signaling has been shown to be particularly important because of its *unconscious* information exchange. As Alex Pentland has argued, building on Goffman, Jervis, and this initial wave of psychology studies, signals become "honest" when they are processed unconsciously and are otherwise uncontrollable.[86] These signals include mimicry, "the reflexive copying of one person by another during a conversation."[87] Individuals engaged in face-to-face interactions will often begin to copy each other. If one person nods their head frequently, the other will do so; if one smiles frequently, the other will often do so as well.[88] "Laughter is contagious," but so are many other behaviors, including emotions, nonverbal physical behaviors, tone of voice, mannerisms, and so forth.[89] Studies suggest that face-to-face interactions, all else being equal, also increase the amount of the empathy displayed between participants. For instance, in experiments where individuals interact face-to-face versus in online environments, individuals in face-to-face report more favorable impressions of the other (greater rapport) as well as greater perceived union (or "oneness," and coupling) with the other individual. Put simply, there was more self–other overlap in identity in the face-to-face condition than in the online condition, suggesting increased empathy in the face-to-face condition.[90] This greater overlap with the other allows for better understanding of the intentions of the other. For instance, in a business negotiation experiment where the participants were asked to predict the pricing intentions of others in the experiment, accuracy of intention-reading was much higher in the face-to-face negotiating group than other groups, such as computer-mediated groups.[91]

Similarly, comparisons between face-to-face environments and online environments suggest behavioral differences in face-to-face interactions relative to other modalities, such as e-mail and online auctions.[92] E-negotiators have a harder time trusting each other than those who engage in face-to-face negotiation, possibly because of the difficulty in understanding intentions in the online context.[93] The cues associated

[85] Morris and Keltner 2000, 16. [86] Pentland 2008, 4. [87] Pentland 2008, 4.
[88] Chartrand and Bargh 1999; Chartrand, Maddux, and Lakin 2005; Tummolini et al. 2006.
[89] For a review, see Chartrand, Maddux, and Lakin 2005.
[90] See, for example, Okdie et al. 2011. [91] Rockmann and Northcraft 2008.
[92] See, for example, Moramarco et al. 2013.
[93] Naquin and Paulsen 2003; Rockmann and Northcraft 2008.

with cooperative intentions, including pitch and cadence of voice for example, are difficult to convey in non face-to-face communication modalities.[94] Studies also indicate that e-negotiators engage in more deception and less truth telling than their counterparts interacting face-to-face.[95] In the aggregate, face-to-face interaction "is at the heart of a process of building trust, mutual respect... [and] shared understanding."[96] At this point it is important to note that while these studies mainly focus on the salutatory effects of mimicry and face-to-face, they should not imply anything about a normative dimension to this type of intention understanding. As will be made clear in subsequent chapters, face-to-face interaction and the intention understanding that engenders from it can also be strategic, or deceptive, in nature.

In sum, as Jane Mansbridge suggests, face-to-face contact "seems to increase the actual congruence of interests by encouraging the empathy by which individual members make one another's interests their own."[97] These insights, particularly the automatic unconscious mimicry in face-to-face interactions, also lend support to the ST theory of mind perspective. If we develop a type of empathy with each other and understand the perspectives and intentions of the other through simulation, then mimicry suggests behavioral simulation of the other. Mimicry and development of empathy, in other words, are a reflection of the internal simulation that is occurring in the mind of the other. This important insight has not only been studied behaviorally but also in the brain through a series of experiments which have important ramifications for theory of mind, to which I will now turn.

The Mirroring System in the Brain

As Marco Iacoboni argues, we rarely give much thought to our ability to read and understand quite clearly the mental states of others in our day-to-day lives. Indeed "[i]t all seems so ordinary."[98] Yet for centuries philosophers have been befuddled by the problem of other minds – how could something so extraordinary feel so ordinary? As Iacoboni points out, this "befuddlement was reasonable: they had essentially no science to work with." Now, however, sophisticated tools have changed that baseline. "No one could begin to explain how it is that we know what

[94] Ekman 1992. [95] Rockmann and Northcraft 2008.
[96] Ansell and Gash 2007, 558. As they note, and I will discuss later in the book, face-to-face provides the potential for trust-building and empathy-building, but it certainly is not a panacea, as the possibility for deception looms large. See also Naquin & Paulson 2003, Purdy et al. 2000, Thompson & Coovert 2003.
[97] Mansbridge 1983, 33. [98] Iacoboni 2008, 4.

others are doing, thinking, and feeling... Now we can achieve our very subtle understanding of other people thanks to certain collections of special cells in the brain called mirrorneurons."[99] The mirror neurons that Iacoboni refers to have an interesting story behind their discovery.

Recently, neuroscience researchers in Parma, Italy were amazed at a discovery involving macaque monkeys, a species of monkey whose brain, while much smaller than that of the human, corresponds well with the human brain. The researchers were investigating motor actions and were specifically interested in the brain functioning involved in planning and executing motor acts, such as grasping, holding, bringing objects to the mouth, and so on.[100] One day between experiments, with a monkey sitting quietly waiting for the next experiment to begin, neurophysiologist Vittorio Gallese reached out for an object and was startled by the burst of activity coming from the computer that was monitoring the brain activity of the monkey. The monkey was sitting quietly, not grasping for anything, yet the neurons in the monkey's brain associated with hand grasping were firing all the same. The neurons were firing based on the mere perception of the researcher reaching out to grasp an object. This was an amazing discovery. As Iacoboni points out, "[c]ells in the monkey brain that send signals to other cells that are anatomically connected to muscles have no business firing when the monkey is completely still, hands in lap, watching somebody else's actions. And yet they did."[101]

Subsequent research revealed that a distinctive class of neurons fires when a monkey executes a motor act and also when it observes another monkey, or a human researcher, performing the same motor act.[102] If a monkey saw someone else eating an ice cream cone, neurons would fire in the monkey's brain as if the monkey itself was eating an ice cream cone. These neurons have been termed "mirror neurons" because of the functional role they play in the brain: they actively replicate, or mirror, the actions of others. Subsequent research has demonstrated that humans possess these mirror neurons as well.[103] One of the reasons this discovery was remarkable is that it demonstrated that the very same neurons could be involved with both perception (such as seeing objects) and action (grasping objects). Rather than these two domains remaining separate, as received neuroscientific knowledge suggested,[104] they are "two sides of the same coin, inextricably linked to each other." This is a very important finding because it links action, and models of action in the brain, with abstract behaviors such as perception. Put another way, the

[99] Iacoboni 2008, 3–4. [100] Iacoboni 2008. [101] Iacoboni 2008.
[102] See Gallese et al. 1996; and Iacoboni 2009a and 2009b.
[103] For a review of the literature, see Iacoboni 2009a and 2009b.
[104] Keysers 2011, 14.

mirror neuron discovery suggested a more holistic interpretation of how the brain works, with motor actions and the goal of those actions, linked together. Action is not just action or physical behavior; rather, actions are linked with perceptions and intentions.

The discovery of the mirroring system in the brain provides a physiological basis for the ST perspective approach to the problem of other minds. Since perception and action are coupled together in the functioning of the neuron, there is not a separate theorizing activity that needs to occur in order for an individual to gain access to the mental states of another; they are instead simulating the experience for themselves. One may wonder, however, whether this type of experience is the same as being able to *read* the mental states of others. The initial mirror neuron studies did not ask monkey participants about their beliefs regarding the mental states of others, of course, nor did they ask participants to predict what their future actions might be. Put simply, is mirroring the same thing as mindreading?

Neuroscientists attacked this problem by examining the experience of witnessing pain in others. Individuals were shown depictions of hands and feet in two conditions: one where it was obvious that the hands and feet were in a painful position and another in a neutral position. They asked participants to attribute the intensity of pain the other was experiencing. The study illustrated that watching others in painful situations triggered the neural network involved in self-pain processing. This confirms the mirroring of experience. But more importantly, the attributions that individuals had regarding the level of pain in others were strongly correlated with the level of activity in the self-pain processing network, confirming the idea that individuals not only mirror the experience of others but also use that mirroring to produce mindreading.[105]

In addition to supporting the ST perspective on other minds with respect to present actions and behavior, mirror neurons are also critical to intentions. The studies mentioned so far suggest that the neurons allow for individuals to mirror the experiences of others and make attributions about their mental states based on that experience, but it may be that the neurons simply are mirroring the actual motor acts, i.e. the grasping or bringing to mouth, but do not mirror the intention of that act. Neuroscientists have identified the "what" and the "why" as two distinct, but linked, elements that characterize each intentional action and are processed by mirror neurons. The "what" refers to a simple observation of the action. Actor A grabs a basketball. The "why" represents an inference of intention. Actor A grabs a basketball because A

[105] Goldman 2006, 138.

intends to shoot it at the basketball hoop. Since our actions are normally associated with a particular intention, it is important to know whether the mirroring system is involved with the mere action or the intention.

In a seminal experiment that illustrates many of the nuances of the mirroring system, Leo Fogassi assessed the neural activity of monkeys during the execution of a grasping task as well as the observation of a grasping task, but in different contexts with respect to the goal of the action. In one execution condition the monkey reached for a piece of food and then brought it to its mouth in order to eat it. In another condition the monkey reached for an inedible object and placed the object in a container. The experimenters attempted to ensure that the actions were as close as possible by placing the container as close to the mouth of the monkey as possible, thereby matching the grasping-to-eat and grasping-to-place-in-container conditions (the experimenter even rewarded the monkey with a treat after placing the inedible object in the container so that there would be little difference in reward between the two conditions). Fogassi wanted to know whether the mirror neurons would fire in the same way with these very similar actions, or whether the neurons would be able to pick up on the intention difference. Does the intention of the act, in other words, matter? Fogassi found that while a small percentage of the neurons fired equivalently in both conditions, the majority of the neurons fired differently between the two setups. Grasping in order to eat caused approximately seventy-five percent of these particular neurons to fire vigorously, whereas grasping to place caused twenty-five percent of the neurons to fire vigorously.[106] The same was found when the experimenters changed the placing condition such that the monkey was placing food in the container and not an inedible object. This suggested that the difference in firing was not about the object itself but rather had something to do with the goal of the action.[107]

Fogassi's team then tested the monkeys as they observed the same experimental setup, but this time the human experimenter conducted the action. In one condition, with the container visible to the monkey, the experimenter grabbed food and placed it in a container. In another

[106] The actual percentage of neurons firing is less important (most of the neurons seemed to "prefer" eating (Iacoboni 2008, 32)) than the *difference* in firing, which is attributable to the distinction in specific intention.

[107] The differential firing between grasping-to-eat and grasping-to-place may seem inefficient, since "a large number of neurons with similar properties are required for executing different types of actions," but as Rizzolatti and Fogassi (2014) point out, it is like "that neurons encoding specific motor acts within an action form pre-wired intentional chains, in which a neuron encoding a motor act is facilitated by the neuron encoding the previous one."

condition, with no container, the experimenter grabbed food and ate it. The container provided a clue to the monkey about the future movement of the experimenter. The results were striking. The intention of the experimenter made a difference. The pattern of firing during observation very closely mirrored the pattern of firing during the monkey's execution of the same act.[108] The same was true if the monkey was observing the human grasping food to place it in the container. These results are striking because they demonstrated that the mirroring system was not merely aiding in recognizing action, but they provided a way of recognizing intentions as well. Importantly, the neuron discharges are measured during grasping, *before* the monkey knows whether eating or placing of the food is going to take place. When there is no intention associated with a particular act, such as with an experimenter pantomiming a grasping action, the neurons do not fire, suggesting that they are able to discern intentional from non-intentional behavior.

The neurons also play a role in inferring specific intentions even when intentions are shielded from view. In one of the most influential mirror neuron studies, experimenters demonstrated that the neurons in a monkey fired when the monkey observed an experimenter's hand moving toward a piece of food to grasp it. They did not fire nearly to the same extent, however, when the hand made the same motion without the food present, as this was a pantomiming action. This replicated the finding mentioned above, where the neurons correctly identified if there was a specific intention present, such as reaching for food. The researchers then placed an opaque screen in front of the food. While the monkey could no longer see the food, the memory of the food being there was enough to cause a subset of the neurons to fire when witnessing a hand reach behind it to grab the food. Similarly, if the screen was placed in front of the bare table and the experimenter reached their hand behind the food then the neurons would not fire. This finding suggests that the mirroring system is involved not only in discriminating intentions, but also in inferring intentions that are partially shielded or hidden from the participant. The only difference between the conditions was the monkey's knowledge of the object on the table; the monkey was still able to discern a pantomime from an actual intentional action. Put another way, the neurons are filling in gaps when information is omitted. Or, as the authors argue, "For these neurons, 'out of sight' was therefore not 'out of mind.'"[109]

Contributing to this "out of sight" aspect of the mirroring system may be the ability to predict future unknown actions, or the diachronic

[108] Iacoboni 2008, 32. [109] Umilta et al. 2001, 156.

problem of intentions. One study at Harvard Medical School placed monkeys into a prisoner's dilemma situation where cooperation was rewarded with apple juice and defection punished by a lack of juice. When the monkeys played against a computer program, they rarely chose to cooperate. Yet when they played the game with another monkey that they were able to see, they were more likely to cooperate. This finding replicates the earlier finding I discussed above, where face-to-face interaction helps to overcome the prisoner's dilemma. In this case, however, the scientists were also measuring brain activity. As the monkey was deciding what to do, a particular set of neurons that predict an actor's unknown decisions during social interaction was activated. In effect the neurons were coding unknown actions, anticipating cooperation based merely on the fact that the monkeys could see each other, even though the action of cooperation/defection remained in the future.[110] As Haroush and Williams argue, this means that a particular set of neurons is able to predict the covert intentions of others.[111]

For decades there was limited evidence of the mirroring system in human brains since studying single neurons in the human brain requires invasive procedures that are not typically available to neuroscientists interested in mirroring. Mirror neurons were therefore traditionally studied in humans from a distance, using brain-wave activity and functional magnetic resonance imaging (fMRI), to identify particular areas of the brain associated with mirroring. For example, participants that are scanned using fMRI while observing and executing actions showed intense activity in the regions of the brain where mirror neurons are believed to reside,[112] particularly the inferior frontal cortex and superior parietal lobe. More recently, researchers have been able to study the mirror neurons directly. Using brains of twenty-one patients who were being treated for intractable epilepsy, and therefore required implanted electrodes to identify where the seizures were originating, the researchers were able to use the same electrodes to study mirror neurons. The researchers found specific neurons that fired when individuals performed a task and observed the task, providing evidence that the mirroring system exists in humans as well.[113] I will now turn in more detail specifically

[110] Haroush and Williams 2015.

[111] This is the first set of findings with respect to these particular neurons, though it remains unclear if these should be considered mirror neurons or a different set altogether. As Mukamel et al. 2010, 1 argue, there is increasing evidence that there are "multiple systems in humans [that] may be endowed with neural mechanisms of mirroring," and therefore thinking about a mirror neuron system as one discrete entity may be misleading. For this reason I prefer to use the language of "mirroring system" to capture the holistic and multi-region essence of mirroring in the brain.

[112] Keysers and Gazzola 2009.

[113] Mukamel et al. 2010.

to what these neurons might mean for human communication and interaction.

Face-to-Face Interaction and Mirroring: Blurring the Distinction between Self and Other

The prominent neuroscientist V. S. Ramachandran has termed these mirror neurons in humans "'Gandhi neurons' because they blur the boundary between self and others – not just metaphorically, but quite literally, since the neuron can't tell the difference."[114] As we have seen, the mirroring system in the brain actively simulates the actions and intentions of others, providing a physiological mechanism by which it is possible for individuals to be able to understand and read the mental states of others. This is the basis for the version of empathy discussed earlier, the ability to feel what it is like to be someone else and understand their cognitive mental states. It also provides philosophical support for a ST understanding of reading others minds. And, importantly, if Ramachandran is right, then empathy is not causing the neurons to fire, but rather the neurons firing *constitutes* empathy. Put another way, empathy has an observable mechanism in the brain.[115]

One of the questions that arises, however, is why we do not *literally* feel the pain of others or literally mimic the actions of others when we see action being performed. One explanation is that mirror neurons are operating in a bodily system comprised of many other types of neurons and processing systems. The mirroring system is not the only one in play in social interactions. For example, when we watch someone being touched the mirror neurons in the brain will fire as if we were being touched, but our skin receptors are sending "null signals" that tell the brain that we are not literally being touched. This suggests a further sophistication of the mirroring system: it helps individuals to empathize with others and be aware of their experiences without having to always literally feel what another is experiencing. Ramachandran supports this interpretation by noting that individuals with phantom limb syndrome, situations that occur when an individual loses a limb but still feels pain in that region as if the limb were still present, when watching another person who still has the limb, are able to feel what occurs with the limb that they are observing. If the observed limb is touched the individual without the limb will feel it. This suggests that the mirroring system is

[114] Ramachandran 2011, 124.

[115] See also Singer 2006; Corradini and Antonietti 2013 on this issue of the neural underpinnings of empathy and philosophical debates connecting mirror neurons to empathy.

actively simulating the literal feeling on the limb but there are no "null signals" that are sent because the limb is physically missing.

As the preceding studies indicate, there is a strong role for visuality in the mirroring system. Much of what is occurring is replication of action and intention based on *observing* that behavior. This has led neuroscientists to become particularly interested in face-to-face interactions in humans, particularly since the behavioral studies mentioned above seem to suggest a coupling or congruence between individuals in face-to-face encounters, ostensibly suggesting that face-to-face is a prime candidate for mirror neuron activity. As Schulte-Rüther and colleagues have shown in a study of the mirroring system involved in face-to-face interaction, the process of recognizing the other's affective and cognitive states strongly invokes the mirroring system and, as they hypothesize, this system allows for interpersonal cognition.[116] They point out that face-to-face interaction involves constant firing of mirror neurons, likely because each side in the interaction, both the self and other, are simulating each other in an effort to understand the other person's overt and inferred meanings.[117] This finding suggests a significant difference between face-to-face and other communication modalities: during face-to-face interaction we move from private to shared experiences.[118] "During conversation, the participants focus or orient toward the other person's mind, inferring meanings and relevancies rather than just decoding the verbal messages. The interaction involves, as probably the most important part, the recognition of the other person's affective and cognitive states."[119] Verbal communication is clearly part of the exchange, but it is not just the verbal aspect that is relevant.

One study explored the relative importance of verbal communication between the face-to-face modality and other modalities by asking participants to communicate in a variety of conditions, including face-to-face and back-to-back. The study utilized a new hyper-scanning technique, where multiple subjects can interact with each other while simultaneous fMRI is conducted.[120] The researchers found in this study that there were significant differences in "neural synchronization" between these conditions. Put simply, the same biological brain processing occurred simultaneously between individuals. This finding suggests, as the authors point out, that compared to other forms of communication, face-to-face "is characterized by a significant neural synchronization between partners based primarily on multimodal

[116] Schulte-Rüther et al. 2007, 1369. [117] Schulte-Rüther et al. 2007, 1369
[118] Hari and Kujala 2009. [119] Hari and Kujala 2009, 461.
[120] Montague et al. 2002.

sensory information integration."[121] In reviewing the relevance of this study for our understanding of communication, Yun points out that the findings suggest that this was not simply about mirroring the action of speaking, such as mouth movements, but rather a much greater and broader neural synchronization.[122] This testing method is important because it allows researchers to isolate various action-related movements from other forms of intentions, creating an index of indicators. Face-to-face therefore includes neural features that other forms of communication do not have, leading the authors to conclude "that people should take more time to communicate face-to-face."[123] Therefore there is some neural truth to the notion of being "on the same wavelength."[124]

Therefore in addition to allowing individuals to understand another person's "*visual* vantage point," they also enable "us to adopt the other person's *conceptual* vantage point."[125] Put another way, as Ramachandran argues, "[t]he only thing separating your consciousness from another's might be your skin!"[126] While we should not take this metaphor too far, since individuals are indeed separate autonomous beings with their own sentience and consciousness, the notion of sharing and simulating experiences helps to drive home the point of shared experiences that occur when individuals are observing each other, and arguably the potential for shared identity as well. Neuroscientists have referred to this as the "shared network" hypothesis, connoting the extent to which the brains of individuals are linked by the mirroring system. These findings point to what Leonhard Schilbach and colleagues have termed a "second-person" neuroscience.[127] The implication here is that in face-to-face interaction, and the social cognition that goes along with it, individuals form a congruent "I–You" relationship rather than a detached "I–She" relationship, suggesting that the *second person* stance is the appropriate one for thinking about face-to-face interpersonal cognition, including intention understanding.[128]

Scholars have also found evidence to suggest that the mirroring system is activated not only with simple low-level instrumental action understanding, but through higher-level abstract thinking, such as propositional attitudes and links to consciousness, as well.[129] Abstract thinking may be derived from the same simulation and mirroring processes involved with perception of action, making abstract thinking a

[121] Jiang et al. 2012, 16069. [122] Yun 2013. [123] Jiang et al. 2012, 16069.
[124] Yun 2013. [125] Ramachandran 2011. [126] Ramachandran 2011, 125.
[127] Schilbach et al. 2013. [128] See Schilbach et al. 2013; Wendt 2015, 232.
[129] See Iacoboni 2009a and 2009b; Keysers and Gazzola 2007; and Rizzolatti and Craighero 2004.

form of an inner motor action.[130] "Activation of mirror neurons in a task relying on empathic abilities without explicit task-related motor components supports the view that mirror neurons are not only involved in motor cognition but also in emotional interpersonal cognition."[131] Additionally, researchers have found that empathic accuracy, the ability to *accurately* judge the mental states and emotions of others, is associated with increased activity of the mirror neuron system, suggesting an important role for mirror neurons and higher-level mentalizing.[132] As I will suggest below, these findings suggest that unconscious signaling, of the type Pentland suggests, enters conscious reasoning somewhere in the process of interpersonal exchange.

Mirroring Emotions and Deception

One area of particular importance for empathy is the affective component, specifically the ability to recognize and understand the emotions of others. Neuroscientists hypothesize that the mirroring system plays an important role in allowing individuals to "feel" the emotions of others when looking at their faces by imitating the emotion in their own body. On this view, emotions are not only reflected in the face, but individuals are able to feel the emotions of others by simulating, through observation, the emotional experience another is going through. As Iacoboni argues, "this simulation process is not an effortful, deliberate pretense of being in somebody else's shoes. It is an *effortless*, automatic, and unconscious inner mirroring." Or, put another way, "[h]umans are 'walking mood inductors,'" continuously resonating with others at a neural level.[133] Much of this simulation occurs at a "micro" level.

[130] Keysers and Gazzola 2007, 4. This type of argument is typically referred to as embodied cognition. With respect to mirror neurons it would suggest that "it is the embodied imitation of the observed body in action that directly enables us to recognize other as persons like us, not an abstract, inferential and theory-like process. The hypothesis that action understanding is based on a resonance mechanism does not exclude the possibility that other processes, based on movement descriptions, could influence this function. It simply highlights the primacy of a direct, automatic and prereflexive matching between the observation and the execution of action. By accepting this premise the traditional tension between acting and thinking considerably shrinks, as the capacity to detect the meaning of the behaviors of others consists in employing the same resources used to model our motor behavior." See Wilson and Foglia 2011.

[131] Schulte-Rüther et al. 2007, 1354. [132] Zaki, Weber, Bolger & Ochsner 2009.

[133] Christov-Moore and Iacoboni 2016.

Microexpressions are very brief involuntary facial expressions that accompany an emotional experience.[134]

One of the early questions raised regarding this type of shared experience that individuals may engender when observing each other is to what extent that experience allows individuals to perceive deception, bad faith, or flat-out lying. Can the mirroring system help individuals differentiate sincere intentions from deceptive intentions? Might the ability to read microexpressions aid in detecting deception since the expressions reveal a true emotional state? Just as there is an evolutionary reason to need to read others' intentions, there is a similar evolutionary need to differentiate sincerity from deceit. As Putnam has noted in his review of the values of social interaction, "[i]t seems possible that the ability to spot nonverbal signs of mendacity offered a significant survival advantage during the long course of human evolution."[135] Or, as Jay explains, "[S]pecies who can't tell the difference between what is true and what is not are unlikely to prosper for very long. The ability to detect deception is, after all, just as functional in evolutionary terms as the ability to deceive."[136] Clearly, deception-detection capabilities help to ensure the survival of the species. This insight led researchers to ask whether discrete brain architecture exists that has evolved for that particular purpose.[137]

A seminal experiment in face-to-face deception detection conducted by Grèzes has provided important answers. Her findings demonstrate that individuals routinely perform better than chance in the detection of deceptive intentions and that distinct brain architecture is invoked when making judgments about those intentions.[138] The study asked participants to view other individuals lifting boxes and assess how heavy the boxes appeared to be. By watching how the individuals picked up the boxes, the observers could infer the boxes' weight, because the movements required for heavy boxes differed from light boxes. The researchers then instructed the individuals lifting the boxes to pretend

[134] There is a large literature on microexpressions which were first described in the 1960s investigating nonverbal communication. Paul Ekman and colleagues later connected microexpressions with deception. For an overview see Ekman and Rosenberg 1997; Ekman 2009; Porter and ten Brinke 2008.

[135] Putnam 2001, 175. See also Mehrabian 1981. [136] Jay 2010, 24.

[137] It should be noted that there is some disagreement as to what constitutes a lie or deception. Sissela Bok in a seminal book on lying referred to a lie, for example, as "any intentionally deceptive message which is stated" (1978, 13). Ekman and O'Sullivan (1991) and Ekman and Rosenberg (1997) added an important notification criteria: lies require the target to be unaware that there is an intention to mislead, thereby explicitly linking deception to intention.

[138] Grezes, Frith, and Passingham 2004.

they were lifting a heavy box when it was actually light. The participants who recognized the deception were able to do so because they inferred the mental state of the person trying to deceive. The finding allows neuroscientists to begin theorizing about the ways in which the brain processes deceptive intentions. Ekman and O'Sullivan, whose research specializes in deception detection, identify this type of mirroring as "emotional." They suggest that human beings detect lies in face-to-face encounters by understanding the emotions on display.[139]

This research is particularly important for refining the conditions under which individuals can simulate sincere intentions and detect deception. Psychologists have long known that, while it is difficult to detect deception, face-to-face interaction makes it easier to do so, because individuals can utilize nonverbal behavior in making judgments about deception.[140] For example, in economic dyadic negotiations individuals display better accuracy in detecting deception face-to-face than in computer-mediated conditions.[141] Building on the work of Ekman and others, Buller and Burgoon synthesize an "Interpersonal Deception Theory" (IDT), where deception is modeled as a dyadic and dynamic process between individuals. Deceivers expend considerable cognitive resources in an attempt to mask deception and unwittingly perform unconscious nonverbal actions during those attempts. It is these nonverbal actions that can provide evidence of deception. These "behavioral markers" suggest that deception literally plays out in the deceiver's physical expression.[142] This "leakage" of intention is part of the, according to some estimates, roughly ninety percent of meaning that is conveyed nonverbally.[143]

Economic models of exchange implicitly adopt these IDT insights and often emphasize face-to-face engagement as a truth-detection device. As Storper and Venables put it, "[F]or complex context-dependent information, the medium *is* the message. And the most powerful such medium for verifying the intentions of another is direct [face-to-face] contact."[144] People with the best deception-detection skills rely on nonverbal cues more than verbal ones. This may be one reason why deception studies, using strangers as participants, often find a weak ability to detect deception. As Frank argues, many of the expressive clues to deception become available through repeated exposure. While individuals have difficulty detecting deception among strangers based on short

[139] Ekman and O'Sullivan 1991.
[140] See Frank 1988, Chapter 6; Bond et al. 1992; Ekman and O'Sullivan 1991; Frank and Ekman 1997; Vrij et al. 2004; Giordano et al. 2007.
[141] Giordano et al. 2007. [142] Barry and Rehel 2014. [143] Mehrabian 1972.
[144] Storper and Venables 2004, 356.

exposure, such as in an experimental setup, "[i]t takes time to recognize a person's normal pattern of speech, gesturing, and other mannerisms," suggesting that separating honest cues from deceptive ones requires experience with that individual, something that is typically not present in experimental designs involving deception detection.[145] In sum, while deception detection is certainly not easy nor foolproof, the evidence suggests that it is easier in a face-to-face context than it is in other interaction modalities. The discovery of brain regions associated with that detection helps to define the precise conditions where detection is successful.[146]

There is another important link between face-to-face interactions and deception: individuals tend to engage in deception *less* when they are interacting face-to-face versus other modalities. As Rockmann and Northcraft show in an innovative experiment where they conduct business negotiations between strangers, individuals who wish to engage in deception when interacting face-to-face confront an uphill battle since they have to control both verbal and nonverbal signals, any one of which may divulge the intention to deceive. As such negotiators may intuitively be aware, or "have a feel," that allows them to pick up on deception and realize that "face-to-face is the most difficult environment" to deceive, since "the deceiver must control all aspects of communication and must present a consistent and believable story even when questioned. The increased cognitive load present when attempting to accomplish such a task impairs one's ability to project a believable lie."[147] The effect of more honesty in face-to-face interactions is seen even with very minimal interaction where there is no back-and-forth conversation; reading from a script in a distributive economic game face-to-face leads participants to be more honest.[148] Face-to-face interactions therefore might serve as an important check on deception, by decreasing the likelihood that it is attempted at all.[149] This is particularly true in "high-stakes" environments, presumably such as high-level diplomacy, where the deceiver has a strong motivation to be successful and consequently becomes *less* effective at controlling nonverbal behavior and therefore is, all else being equal, worse at deceiving.[150] This also resonates with Mearsheimer's claim that lying is not particularly ubiquitous in international politics due to fear of being caught.[151]

Lastly, individuals vary in their ability to perceive deception through expressive cues, just as they vary in their ability to empathize more

[145] Frank 1988, 136–37. [146] Langleben, Willard, and Moriarty 2012.
[147] Rockmann and Northcraft 2008. [148] Van Zant and Kray 2014.
[149] See also Valley et al. 1998 on individuals telling the truth more in face-to-face negotiations versus telephone and written communications.
[150] Depaulo et al. 2003. [151] Mearsheimer 2011.

generally. Deception detection is a manifestation of emotional and social intelligence.[152] This type of intelligence refers to the ability to understand one's own, and others', emotional states and the capacity to successfully navigate social relationships.[153] Indeed, the individuals best at detecting deception are those that score highly on emotional intelligence measures and consequently are able to interpret nonverbal cues accurately.[154] Interestingly, some studies suggest that successful leaders in general possess high levels of emotional intelligence.[155] On the other hand, individuals routinely express overconfidence in their ability to detect lies. In fact, in a range of tasks, including sensitivity to lies and reading emotions, across multiple dimensions, including controlled video stimuli and face-to-face interactions, the individuals with the *least* sensitivity tended to substantially overestimate their capabilities. For example, those in the lowest quartile on actual empathic ability rated themselves at or above average.[156] Importantly, overestimation is somewhat predictable based on individual characteristics. Narcissism, for example, predicted self-estimated performance. Put simply, individuals with narcissistic tendencies toward self-aggrandizement overestimated their ability to read others.[157]

All of this suggests that there is variation in deception detection and there is reason to believe that leaders, all else being equal, may be in a better position than most to pick up on deception unless they possess particular characteristics, such as high levels of narcissism in their personality. Deception is therefore not as devastating to a theory of intention understanding as the problem of dissembling might initially suggest, though as the Munich case to be discussed later illustrates, it remains a difficult problem for face-to-face diplomacy.

What Kind of "Knowledge" is Brain-Based Knowledge?

The preceding discussion has suggested that brain systems are involved in the understanding of others' intentions, and these integrated systems allow for individuals to experience what the other experiences, suggesting a way around the problem of other minds, such that others are conceived "not as bodies endowed with a mind but as persons like us."[158] A question remains, however, regarding how we should think about mirroring from an epistemological perspective. Put most simply, what do these studies *mean*? Since brain data is confined to the laboratory, how do we fill in the gaps between fMRI experiments and real-world politics?

[152] O'Sullivan 2005, 237. [153] O'Sullivan 2005, 237 [154] O'Sullivan 2005, 248.
[155] Goleman 2005. [156] Ames and Kammrath 2004, 205.
[157] Ames and Kammrath 2004, 205–08. [158] Gallese 2001, 43.

This is the scientific realism challenge: unable to observe mirror neurons in actual politics, how can we know whether or not they are indeed a causal mechanism of import? The first step in answering these important questions is understanding how we should conceive of the type of inferences that are supplied through face-to-face interaction. After all, it may be that the brain is simulating the intentions of others, but if that simulation remains inaccessible to the individual's conscious thoughts, then it is less clear how useful the mirroring is in political situations, though current research points toward mirroring occurring in conscious thought as well as subconscious thought.[159]

While it may be the case that what we call the simulation is less relevant than what it allows individuals to do, what is important is determining the extent to which the mirroring is accessible in consciousness, since political decisions in diplomacy are presumably conscious ones.[160] As Corradini and Antonietti point out, neuroscientists refer to the type of intention understanding that occurs quite variably, including: "non-predicative,"[161] "without verbal mediation,"[162] "without the need of theorising,"[163] "without propositional attitudes,"[164] "non-inferential,"[165] "without any knowledge operation,"[166] "not needing cognitive processes,"[167] "pre-reflective."[168] Further, the adjectives used to describe these types of understanding also vary, including: "direct,"[169] "immediate,"[170] "effortless,"[171] "automatic,"[172] "implicit,"[173] "unconscious,"[174] "subpersonal."[175] One curious observation regarding this survey of terminology is that "knowledge" is absent; on this view the mirroring system is providing the self with a pre-reflective unconscious and automatic mental picture of what is going on in the mind of the other, but we should stop short of calling that sense knowledge, even if it does influence decision-making.

There is reason to believe, however, that the pre-reflective picture does constitute, at least for some, a special type of knowledge. The problem

[159] See, for example, Christov-Moore et al. 2016 that demonstrates a correlation between conscious decision-making in economic games and brain activity that occurs during mirroring-related tasks, such as the use of facial headshots.
[160] Though there is a growing recognition in IR that many decisions may be habitual, intuitive, or otherwise automatic in nature. See, for example, Pouliot 2008; Hopf 2010; Holmes 2015; Holmes and Traven 2015.
[161] Gallese 2001, 44. [162] Rizzolatti and Sinigaglia 2006, 120.
[163] Gallese 2001, 41. [164] Gallese 2001, 41.
[165] Gallese 2001, 41; Rizzolatti and Sinigaglia 2006, 174.
[166] Rizzolatti and Sinigaglia, 2006, 127. [167] Rizzolatti and Sinigaglia, 2006, 174.
[168] Iacoboni 2009a, 666. [169] Gallese 2001, 41.
[170] Gallese 2001, 41; Rizzolatti and Sinigaglia 2006, 127. [171] Iacoboni 2009a, 666.
[172] Gallese 2001, 41; Iacoboni 2009a, 666. [173] Gallese 2001, 41.
[174] Gallese 2001, 41. [175] Gallese 2001, 42 and 46.

with traditional knowledge terminology, as Alvin Goldman points out, is that "knowledge is a thick epistemological concept, connoting more than just attribution or belief."[176] Philosophers tend to view knowledge "as something like justified true belief, or reliably formed true belief."[177] Since it would be difficult to objectively justify what engenders from the mirroring system without first checking up on its accuracy empirically, it is difficult to see how the mental picture constitutes traditional knowledge and this is likely the reason that neuroscientists do not refer to the picture in knowledge terms. Yet, many philosophers argue that the traditional view of knowledge does not account for the distinctiveness of "self-knowledge." Self-knowledge is knowledge of one's own mental states, including one's own beliefs, desires, sensations, and so forth.[178] What makes self-knowledge unique is that the self has "privileged access" to what occurs in one's own mind. This doctrine suggests that there is a fundamental difference between experiencing something for oneself, in one's own mind, and relying on something that comes from the outside, such as a third-person report. In self-knowledge, the justification for having the belief is that one is experiencing it for oneself. From this perspective it may be that the mirroring system provides a type of knowledge that is quite unique and, as some argue, potentially the basis for consciousness.

I conceptualize the direct and automatic product of the mirroring system as an intuition. Intuition is often viewed pejoratively in decision-making as the source of bias or irrationality; after all, *homo economicus* is a rational cost-benefit calculating individual,[179] not one prone to gut reactions and decisions based on instincts. Work in psychology and philosophy of mind takes a different view, however, suggesting that automatic intuitions play a vital role in social interaction and decision-making. While there is disagreement on exactly how to conceptualize intuitions, there is widespread agreement that intuitions serve a *belief-like* function that is likely antecedent to belief and other cognitive states. The functions include "the speed of knowing something, knowing without knowing how you know, and also knowing without a conscious step-by-step process."[180]

Intuition can have many sources, including learned experience, emotional or affective responses, or gut reactions that may not have identifiable sources, and is often referred to in different forms and is linked to a number of similar concepts in the literature. For instance, practice theory and the logic of habit, which have recently been applied

[176] Goldman 2006, 223–24. [177] Goldman 2006, 223–24.
[178] Gertler 2003 for review of this literature. [179] Frantz 2004. [180] Frantz 2004.

to international politics, and suggest non-conscious non-deliberative sources of behavior, have strong resonance with intuitions. On both views individuals do not act on conscious reasoning but rather on unthinking practices and habits.[181] These practices and habits, which are automatic in nature and resistant to change, share many of the same properties of intuitions, though they are not necessarily referred to as such. Therefore when leaders, for example, have a "feel for the game" that guides their decision-making, they are likely operating on an intuition about how to proceed in a given context.[182] Recent work in psychology also suggests a role for intuition in creating moral judgments. Haidt has argued that contra rationalist models of moral judgment, individuals often use quick automatic affective evaluations, or intuitions, to come to moral conclusions.[183]

One way of understanding intuition is to view it as a heuristic that bypasses conscious rational thought processes but is nevertheless accessible by consciousness. A familiar example of this concept is found in Kahneman and Tversky's research program in prospect theory. As Kahneman notes, "[our research] was guided by the idea that intuitive judgments occupy a position – perhaps corresponding to evolutionary history – between the automatic operations of perception and the deliberate operations of reasoning."[184] The different types of processing that Kahneman refers to are exemplified by the difference between "System 1" and "System 2." System 1 is the fast and automatic processing that provides quick judgments; System 2 is the more deliberate, reasoned, and reflective processing that is much slower in nature.

The intuition framework is helpful because it captures much of what the mirroring system is doing, as articulated by neuroscientists, and how it is useful to individuals engaged in social interactions, as articulated by psychologists and philosophers of mind. Intuition represents a very general category of pre-analytical non-reasoning based on "knowing without knowing how you know." This "self-knowledge" intuition that is produced by the mirroring system constitutes a belief-like mental state that is antecedent to further cognitive processing and therefore is in the realm of System 1 automatic processing, but can move to System 2 upon reflection. What begins as an automatic experience-based *intention intuition* becomes an *intention belief* once the content of the intuition is reflected upon and the individual reasons and deliberates about its content. This formulation also helps to bridge the divide between

[181] Hopf 2010; Pouliot 2008; Adler and Pouliot 2011a; Pouliot 2010; Adler-Nissen and Pouliot 2014.
[182] Pouliot 2010, 35; Holmes and Traven 2015. [183] Haidt 2001.
[184] Kahneman 2003, 697.

ST and TT approaches. The evidence presented in this chapter suggests that ST is a largely automatic mirroring process in the body, creating a quick intuition about the mind of the other, and TT involves reflection about that intuition, forming a belief. Crucially, in this move from the automatic to the reflective there is an opportunity for other psychological mechanisms, or what Wheeler calls "psychological drivers" to enter the equation, providing an important scope condition for the theory. Doubt, latent mistrust, bad-faith models and images, stress, anxiety, hubris, and so on, all likely have an effect on the formation of *intention beliefs*. The privileged access doctrine discussed above implies that individuals will be likely to privilege intention beliefs over other types of information. As Jervis puts it, leaders may believe "that their rise to power was partly dependent on a keen ability to judge others."[185] This does not imply that the individual's reading of the other will be accurate, however, only that it will be privileged.

Intention Beliefs: Accuracy and Change

One of the greatest challenges that psychologists, neuroscientists, and philosophers face when researching intention understanding is delineating why it sometimes engenders, or "works," and why in other cases it does not. For every example of President Ronald Reagan and Russian President Mikhail Gorbachev creating an emotional bond of common understanding, what Gorbachev called "the human factor," and former US diplomat Jack Matlock said was essential in overcoming the Cold War,[186] there are salient examples of the precise opposite occurring. President George W. Bush visited Russia's president Vladimir Putin and argued that by looking him in the eye he was able to "get a sense of his soul." Whereas Reagan would find that his intention beliefs regarding Gorbachev's intentions to end the Cold War were accurate, Bush later found that his reading of Putin was flawed. Was this a case of poor empathic accuracy or a lack of mirroring in the first place?

Psychologists have found a variety of factors that can affect empathic accuracy. First, the perceiver's familiarity with the target is important.[187] Practice with a particular target can increase accuracy.[188] The intuition that ambassadors often have, for example, that it is their familiarity with counterparts on the ground that helps to explain their understanding of the political dynamics in a given culture,[189] is supported by these studies. Gender differences have also been suggested to play an important role in

[185] Jervis 1970, 33. [186] Matlock 2004. [187] Stinson and Ickes 1992; Ickes 1997.
[188] Marangoni et al. 1995. [189] See, for example, Pouliot 2016.

empathic accuracy, though interestingly only if the woman is made aware that her ability to empathize is being measured: "If a woman is aware that the task she is completing is assessing her empathic capabilities, it may be important for her to perform well. She therefore may be more successful than a man completing the same objective measurement of empathy because of her increased level of motivation."[190] Put simply, it seems that it is not a difference in ability, but rather a difference in motivation, that accounts for higher empathic accuracy in women.[191]

Empathy is also affected by group dynamics.[192] Participants in empathy studies show less simulation for members of different races, thus supporting the existence of a so-called "empathy gap" that suggests empathic accuracy may be mediated by social groups.[193] In one study individuals showed significantly less simulation of actions and intentions when viewing members of the out-group when compared to members of the in-group (in this study the groups were defined by ethnic identity).[194] Others studies have suggested that it is more difficult to recognize faces of those in the out-group,[195] including recognition and interpretation of facial expressions.[196] On the other hand, the mirroring system is also implicated in being able to *reduce* the empathy gap. In an innovative study Michael Inzlicht, Jennifer Hutsell, and Lisa Legault found that they could reduce racial prejudice in experiment participants by having them mimic the movements of a member of the group that they are prejudiced against.[197] The authors hypothesize that part of what defines prejudice is the lack of empathic connection with others; by explicitly invoking the mirroring system individuals are able to empathize more with others. Put another way, prejudice reduces the ability to resonate with out-groups, but resonating with the actions of the out-group can also reduce prejudice. Visuality reduces prejudice. Similarly, as mentioned earlier, emotional intelligence is correlated with empathy and deception detection while narcissism is inversely correlated with deception detection. Further, as Hall and Yarhi-Milo argue, readings of others through empathy may be particularly important in times of information scarcity, ambiguity, or cognitive stress, where leaders are searching for ways to understand one another.[198]

[190] Klein and Hodges 2001, 721. [191] Klein and Hodges 2001, 727.
[192] See, for example, Emile Bruneau's work on parochial empathy, the difference between ingroup empathy and outgroup empathy. Cikara, Bruneau, and Saxe 2011; Bruneau et al. 2012.
[193] See Gutsell and Inzlicht 2011; and Xu et al. 2009; Bloom 2016.
[194] Gutsell and Inzlicht 2010. [195] Sporer 2001.
[196] Elfenbein and Ambady 2002. [197] Inzlicht et al. 2012.
[198] Hall and Yarhi-Milo 2012.

The mirroring system is, however, also universal in a number of important, and surprising, respects. Neuroscientists and philosophers of the mind have suggested that the simulation circuit created between individuals is a universal phenomenon, with most individuals possessing the architecture required for simulating the mental states of even dissimilar others. Deficits to the mirroring system limit empathic capabilities, and it is also likely that individual differences on the other side of the spectrum, individuals with more robust mirroring systems, are capable of a type of "super-empathy," with most individuals falling somewhere in the middle of the empathy bell curve.[199] As mentioned above, empathy can, to a certain extent, also be practiced and learned. Lutz et al. found that Buddhist monks, who had years of training, practice, and meditation devoted to developing empathy and compassion for others, displayed greater activation of empathy networks than untrained control groups;[200] long-term meditation practice has also been linked to thicker brain regions associated with empathy tasks.[201] Even individuals who are very different from those they are observing, such as individuals born without hands and feet, are able to mirror the actions of those born *with* hands and feet.[202] More generally, however, subjects who score highly on standard empathy measurement scales also tend to show higher activation with empathy-circuit and mirroring systems in the brain.[203]

Finally, returning to the early mirroring experiments, specificity of intention plays a very important role in the ability to successfully intuit intentions. Mirror neurons fire the most when they are engaged in simulating a specific intention act, rather than a vague one, even if that specific intention is shielded or deceptive in nature. This implies an important scope condition for empathy generated in face-to-face interactions: we should expect the level of intention specificity to matter for intention understanding. Put another way, one reason why Bush may have read Putin incorrectly has to do with specificity. The mirroring system does not mirror "senses," or general trustworthiness; it mirrors specific intentions and specific actions. Whereas Gorbachev and Reagan iterated over a series of meetings where very specific intentions were discussed and conveyed, Bush's gleaning of a sense from Putin did not involve a similar level of intention specificity. Bush thought that he could look Putin in the eye and judge his trustworthiness; the error was not in thinking that others can be read, but rather *what* can be read. Reading specific intentions is a different activity than reading the generalized soul.

199 Baron-Cohen 2011, 177-181. 200 Lutz et al. 2004.
201 Lazar et al. 2005. 202 Keysers and Gazzola 2007.
203 Singer et al. 2004; Singer et al. 2006.

The iteration of interaction becomes particularly important when it comes to belief change. As noted earlier, since intuitions are the realm of System 1 and beliefs the realm of System 2, belief represents a stickier, and less conducive to change, mental state. Further, because System 2 includes the "psychological drivers" discussed above, intuitions face a particularly uphill battle in changing beliefs. A question emerges from this discussion as to how intention beliefs may change or update over time. While the question of belief change, particularly with respect to how it comes about, is far from settled in psychology and philosophy of mind, there is often a distinction made between belief updating and belief revision. With belief updating new information is taken into account in the present and old beliefs are changed to take account of the new information. Revision, on the other hand, occurs when new information shows an inconsistency between new and old information and the belief is changed in such a way to account for the inconsistency. Belief revision represents more minimal change than belief updating. In the case of intention intuitions changing intention beliefs, it is likely the case that we are more often dealing with belief revision than updating. Since beliefs are sticky, and intuitions may represent inconsistencies between old beliefs and new information, they are likely to require several iterations of revising before they are fully changed. Returning to the Gorbachev and Reagan example, which will be the focus of the next chapter, the iteration that occurred in their meetings led to a series of belief revision over time. Beliefs were not updated and changed completely when the first face-to-face intention intuition was engendered, but rather the beliefs required incremental revising in order to change.[204]

Special Problems for Individual-Level Psychological Arguments

Theories of international politics that are based in the individual-level of analysis, or "first image," necessarily run up against special problems of causal inference. These include the problem of aggregating individuals upward to the state decision-making apparatus, identifying the relative causal weight of individuals versus structural factors such as power, the

204 In this way the incremental process of System 1 intuitions changing System 2 beliefs is consistent of an interpersonal socialization perspective where individuals come to understand each other better through iterative interactions, though importantly in my case the intention intuitions are accessible without socialization. Iterative interaction and the socialization that often follows, in other words, is not necessary for the creation of intention intuitions, though as discussed above with deception may indeed make those intuitions more accurate.