Mindreading by Simulation:

The Roles of Imagination and Mirroring

Alvin I. Goldman and Lucy Jordan

1. Criteria of Adequacy for a Theory of 'Theory of Mind'

There is consensus in cognitive science that ordinary people are robust mindreaders and that mindreading begins early in life. Many other questions concerning mindreading, however, remain in dispute, including the four that follow:

(1) By what method(s) do cognizers read other people's minds -- that is, attribute mental states to them? Which cognitive capacities, mechanisms, or processes play pivotal roles in mindreading?[1] Call this the *task-execution* question.

(2) How did the human species, and how do individuals, acquire mindreading capacities? What are the phylogenetic and ontogenetic parts of the story? Call this the *acquisition* question.

(3) What neural substrates underlie mindreading? In other words, does the proposed story pass neuroscientific muster? Call this the *neural plausibility* question.

(4) How does the proffered story of mindreading mesh with the general story of human cognition? Is mindreading a typical example of cognition, or is it a singularity -- a one-off piece of cognitive hardware? Call this the question of *mesh*.

[1] People read their own minds as well as the minds of others. How first-person mindreading is executed is a question of equal importance and difficulty as the third-person mindreading question. For this reason it cannot be addressed within the confines of this chapter, so it is left for another day. (See Goldman 2006, chaps. 9-10 for an earlier foray into this territory.)

Any theory or approach to mindreading must answer these questions, or most of them. It should provide systematic answers that address the entire scope of mindreading: the full range of states that get attributed and the full range of contexts or cues on which mental attributions are based. The mental states imputed to others include at least three types: emotions (e.g. fear, anger, disgust), sensations (pain, touch, tickle), and propositional attitudes (belief, desire, intention). Attribution of all such states needs to be covered by an adequate theory. Our own approach will offer plausible answers to all of the foregoing questions. In that sense it constitutes a "full scope" approach. Some of its rivals, by contrast, don't pass this test of adequacy. The rationality or teleological approach, for example (cf. Dennett 1987; Gergely et al. 1995; Csibra et al. 2003) seems to lack the resources to explain attributions of sensations or emotions.[2]

2. Levels of Mindreading

The general contours of the simulation approach have been laid down by a number of contributors. In the 1980s philosophers advocated simulation (or "replication") as an alternative to the dominant functionalist, or theory-theory, approach to folk psychology (Gordon 1986; Heal 1986; Goldman 1989). In the 1990s a developmental slant on simulation theory was presented by Harris (1992). Later in the 1990s and 2000s neuroscientific findings steered much of the impetus for simulation theory (Gallese & Goldman 1998; Currie & Ravenscroft 2002; Decety &

[2] A standard taxonomy of approaches to mindreading *other* than simulation theory includes the rationality theory, the child-scientist version of the theory-theory, and the modularity version of the theory-theory. For detailed expositions and critiques of these rivals, see Goldman 2006, chaps. 3-5. Selected problems for some of these rivals are sprinkled throughout this chapter, but length limits preclude detailed treatments.

Greze 2006; Goldman & Sripada 2005; Gallese 2007). The present chapter begins by reviewing the original model that focused on the mindreading of propositional attitudes. It then moves to a later variant directed at the attribution of emotions, sensations, and intentional motion. The second half of the chapter examines more recent findings that could play pivotal roles in the ongoing debate.

Many treatments of theory of mind postulate two or more *levels, components,* or *systems* of mindreading, and we too offer a bi-level approach (cf. Goldman 2006). But not all duplex theories draw the same partitions or have the same rationale. An early two-component architecture similar to one we favor is that of Tager-Flusberg & Sullivan (2000). They distinguish a "social cognitive" component and a "social perceptual" component. Their social-cognitive component features a conceptual understanding of the mind as a representational system, and is highly interactive with other domains such as language. The social-perceptual component is involved in the perception of biological and intentional motion and the recognition of emotion via facial expression. This distinction maps well onto our distinction between "high-level" mindreading of the attitudes versus "low-level," or mirror-based, mindreading of non-propositional states. Essentially the same distinction is carved out neurologically by Waytz and Mitchell (2011). Our two methods of mindreading exemplify many of the contrasts that typify the popular dual-systems, or dual-processes, approach in contemporary cognitive science. Low-level mindreading is comparatively fast, stimulus-driven, and automatic; high-level mindreading is comparatively slow, reflective, and controlled.

Apperly advocates a different two-systems approach (Apperly & Butterfill 2009; Apperly 2011). Both of his systems concern the propositional attitudes, but two systems are posited by analogy with numerical cognition. One system is characterized as efficient but inflexible, the other as flexible but effortful. It is hard to get a firm grip on his systems, however, partly because the account changes significantly between the two publications. The 2009 publication distinguishes between two types of *states* that are mindread -- "registrations" and "beliefs" -- but registrations disappear in the 2011 publication.

3. High-Level Simulational Mindreading

Early formulations of the simulation theory (ST) correspond to what we call high-level mindreading. To pinpoint its most significant features, we contrast it with its perennial foil, theory-theory (TT).   We begin with an example:

> Shaun just left the house and drove away. I ask you where he is going, and you reply: "He didn't say. But I know he wants an espresso and thinks that the best espresso is at Sergio's café. So he probably decided to go to Sergio's."

You have executed a mindreading process, the upshot being the attribution of a certain decision to Shaun. How did you arrive at this? TT would reconstruct your mental process as in Figure 1. [Figures are not available for this pre-publication version.] You start with three beliefs, two specifically about Shaun and one about human psychology in general. All the beliefs are depicted as ovals on the left of Figure 1. You believe of Shaun that he wants an espresso and thinks that Sergio's is the best (nearby) espresso place. Your general belief about human psychology is the "theoretical" proposition that people generally choose actions most likely (by their lights) to satisfy their desires. These three “premise” beliefs are fed into your reasoning mechanism, which outputs the conclusion that Shaun decided to go to Sergio's.

Several things about this simple TT story are noteworthy. First, the mindreader’s states that do all the "work" in the TT story are *belief* states, and the only *processor* used is a theoretical reasoning mechanism. The same would hold of somebody trying to understand and explain the workings of a physical system. Nothing transpires under the theory-oriented account like putting oneself in the target system's shoes. Second, the belief states that do the work are structurally rather complex. They are all *metarepresentational* states, which refer to states of the target that are themselves representational (have *content*). Shaun is portrayed as having a desire and a belief, each of which is a representational state. Third, under TT, mindreading's aptness for success critically depends on the content of the mindreader's naïve psychological theory.   If this theory is ample enough in detail and (approximately) descriptively correct, it may tend to supply fairly accurate mental attributions. But if it is meager or misguided, it will frequently lead the mindreader astray. This will happen especially when the target's mental processes are sophisticated and complex.

One form of TT exploits the adequacy or inadequacy of the mindreader's psychological theory to explain influential patterns of error found in early childhood mindreading, especially errors in (verbal) false-belief tasks. Proponents of this form of TT -- so-called "child scientist" theory-theorists -- contend that children gradually refine and improve their theory of mind during their early years, much as adult scientists refine their theories over time. One such refinement is the replacement of a non-representational theory of mind by a representational theory. The later-developing representational theory allows them to conceptualize the possibility of false belief and thereby improve their performance in false-belief tasks between 3 and 4 years of age.

A second form of TT, the *modularity theory,* denies that children develop a theory of mind by a science-like process. Rather a theory of mind is an innate endowment of one or more dedicated modules (Baron-Cohen 1995; Leslie 1994). How, then, might this sort of theory explain comparatively poor performance by 3-year-olds on false-belief tasks? Leslie et al.

(2005) introduce an additional, non-modular mechanism, the *selection-processor,* to account for this phenomenon. The selection processor selects among candidate belief contents in a target agent's mind by inhibiting the default content -- namely, the content true of the world -- and instead selects an alternative content (which is false of the world). Three-year-olds are weak at this task because their selection processor includes an inhibitory mechanism that hasn't fully matured at three years. Thus, 3-year-olds have a *performance* problem with false-belief tasks, not a *conceptual* problem, as child-scientist theory-theorists claim. Despite this difference, both types of TT hold that mindreading is executed by reliance on a theory of mind, whether an innate theory (embedded in one or more modules) or a gradually developing one.

Could the same tasks be executed in a less informationally demanding manner? Specifically, could they be done with less reliance on refined generalizations about causal connections among mental states? ST takes this tack. It conjectures that mindreaders exploit their own mind as a prototype, or model, of the target's mind. If different minds have the same fundamental processing characteristics, and if the attributor puts her own mind in the same "starting-state" as the target's and lets it be guided by her own cognitive mechanisms, mental mimicry may allow her to determine what the target is going to do. ST embraces this alternative hypothesis, depicted in Figure 2.

A distinctive feature of Figure 2 is the presence of *shaded* shapes, which represent *pretend* states, i.e., products of pretense or imagination. Imagination is assumed to be a faculty that, when you wish to be in a specified mental state M, proceeds to construct an M-*like* state in you. By an "M-like state” we mean a state that is (at least functionally) very similar to a genuine state M, but would normally be produced by cognitive mechanisms other than imagination (e.g., perception, reasoning, emotion-generation). One crucial similarity between a genuine M-state and M-like state is that they produce similar output states when fed into a cognitive mechanism, for example, a choice or decision-making mechanism. Figure 2 depicts the mindreader as constructing a *pretend* desire (intended to "match" Shaun's relevant desire or goal) and a *pretend* belief (intended to match Shaun's relevant belief). These pretend states are fed into a decision-making mechanism, which operates over these inputs and generates an output state: a decision. This output state is also depicted by shading because it is still under the control of the imagination. Notice that Figure 2, unlike Figure 1, makes no reference to a *factual reasoning* mechanism or to a psychological theory of mental processes. The need for theoretical reasoning is replaced in ST by a simulation process, which in this case employs a decision-making mechanism that helps to replicate Shaun's decision-making process. The simulation routine terminates when the decision-making mechanism outputs a decision. This decision is attributed to Shaun (as shown at the far right of the diagram), the attribution being a *genuine* (hence unshaded) belief of the mindreader.

Do any interesting predictions flow from the simulationist model? One prediction is that if the mindreader's imagination performs poorly in constructing the target's starting state(s), the mindreading routine is not likely to succeed (be accurate). A second prediction requires more ground-laying. Mindreading by simulation runs the risk of letting the mindreader’s own mental states get entangled with the pretend ones. A mindreader, after all, will always have her own “genuine” desires, beliefs, and intentions alongside the pretend ones. These genuine desires and thoughts must be segregated from the pretend ones, an activity that may not be trivial. There is a danger that genuine states will interfere with pretend ones, causing confusion and error. To avoid such entanglement, genuine states must be "quarantined" or "inhibited" to avoid confusion with mimicked states of the target. Thus, intensive use of simulation predicts a high incidence of mindreading error -- specifically, egocentric error, reflecting the penetration of the mindreader's own genuine desires, beliefs, and emotions into the interpersonal tracking process.

Would egocentric errors be equally predicted by TT? Since a theorizing mindreader would also have her own thoughts running on a parallel track with those of the target, doesn't she face an equal danger of interference? If so, egocentric mindreading errors will not constitute a discriminating test of the rival theories. We argue that the likelihood of interference is higher under ST than TT. Why? Because customary cognitive acts and processes are more similar to -- hence easier to confuse with -- states of mental mimicry posited by ST than the kinds of cognitive acts and processes posited by TT. Under TT, states deployed during mindreading are exclusively *metarepresentational*, whereas those deployed during simulation are first-order states. Hence, it should be easier for normal thoughts and plans to mistakenly encroach or insinuate themselves into simulated thoughts and plans than under theory-guided mindreading.

Now, the mindreading literature is replete with reports of egocentric errors, or biases (including, but not restricted to, false-belief attribution errors). Much of it goes under the heading of "curse of knowledge," a phrase originally introduced in a study of adults who were forewarned that their targets' knowledge differed from their own, but nonetheless allowed their own knowledge states to seep into attributions to their targets, generating poor task performance (Camerer et al. 1989; Nickerson 1999). The same leakage phenomenon is found in children (Birch & Bloom 2003, 2004). For the reasons indicated, this lends greater support to ST as compared with TT.[3]

[3] The reasoning relies on Bayesian conditionalization. When the likelihood of O given H_1 is greater than the likelihood of O given H_2, observation of O will increase the posterior probability of H_1 more than it increases the posterior probability of H_2. The present contrast between ST and TT is aimed mainly at the child-scientist version of TT. As reported above, Leslie's form of modularity theory shares with ST a reliance on inhibitory mechanisms.

Although ST easily comports with the observed pattern of egocentric biases, shouldn't it predict much more error than is actually observed? Shouldn't simulation lead to rampant error in virtue of the fact that pretend beliefs, desires, and emotions must surely be different from their genuine counterparts? How can imagination-generated states resemble genuine states so closely that similar decisions or new beliefs get outputted when the pretend versus genuine states are inputted into similar cognitive mechanisms? Is there really evidence for a tight enough similarity between pretend and genuine states to support high levels of mindreading accuracy? Yes. Cognitive science and neuroscience is replete with evidence that imagination is powerful enough to produce states that closely match their counterparts. This is most thoroughly researched in the domains of visual and motoric imagery. Neuroscientific studies confirm that visual and motor imagery has substantial neurological correspondence with vision and motor execution respectively (Kosslyn et al. 1997, 1999; Jeannerod 2001). Chronometric studies of motor imagery are particularly striking (Decety et al. 1989; Currie & Ravenscroft 2002: 75 ff). But just how powerful and accurate is imagination in non-perceptual and non-motoric cases? A recent study described in section 5 demonstrates the unexpected power of imagination, which should help deflate skepticism about simulation.

A major strand of the ST-TT debate has hinged on the plausibility of the thesis that a complex skill like mindreading is driven by a theory that unfolds during a child’s early years. Early defenders of (the child-scientist version of) TT claimed to find evidence that children revise their theory of belief between two and four years of age, yielding mature competence only around four (Wellman 1990; Perner 1991; Gopnik & Meltzoff 1997). The details of this claim, however, were blown out of the water when Onishi & Baillargeon (2005) found false-belief competence (in non-verbal tasks) at 15 months of age. But now, in many parts of cognitive science, there is impressive evidence of statistical learning (specifically, Bayesian learning) in many areas of cognition. Does this support a new form of TT as over against the simulation hypothesis?

A study by Baker, Saxe, and Tenenbaum (2009; cf. Baker, Saxe, and Tenenbaum, forthcoming) is a good example of an empirically-based defense of theory-based mindreading supported by Bayesian inference. They propose "a computational framework based on Bayesian inverse planning for modeling human action understanding … which represents an intuitive theory of intentional agents' behavior…. The mental states that caused an agent's behavior are inferred by … Bayesian inference, integrating the likelihood of the observed actions with the prior over mental states" (2009: 329). If the cognitive reality of this framework is indeed empirically sustained, as they claim, doesn't it decisively support TT over ST?

Interestingly, Baker and colleagues themselves concede that their findings do not favor TT over ST. The reason is that mindreaders who use Bayesian methods to ascribe mental states to others may simply be running their Bayesian reasoning capacity *as a simulation* of the target. Thus, as Baker et al. write:

> [T]he models we propose here could be sensibly interpreted under either account.... On a simulation account, goal inference is performed by inverting one's own planning process -- the planning mechanism used in model-based reinforcement learning -- to infer the goals most likely to have generated another agent's observed behavior. (2009: 347)

4. Low-level Simulational Mindreading

The best evidence for low-level simulational mindreading, we submit, is found in research on emotion *mirroring*. First we examine evidence for the existence of mirroring, that is, mental-state contagion. Then we present evidence that mirrored states are used as the causal basis for mindreading.

Both animal and human studies show that the anterior insula is the "gustatory cortex" and primary locus of the primitive distaste response, disgust (Rozin et al., 2000; Phillips et al., 1997). Against this background, Wicker et al. (2003) performed a functional imaging study in which participants first viewed movies of other people smelling the contents of a glass (disgusting, pleasant, or neutral) and displaying congruent facial expressions. After first serving in this observer capacity, the same participants then had their own brains scanned while inhaling disgusting or pleasant odorants through a mask on the nose and mouth. The core finding was that the left anterior insula and the right anterior cingulate cortex (ACC) were preferentially activated both during the inhaling of disgusting odorants and during the observation of facial expressions of disgust. Thus, there is indeed mirroring (or contagion) for disgust.

This study presented no evidence concerning the mindreading of disgust (via observed facial expressions). For evidence that mindreading *is* based on mirrored disgust, however, we turn to neuropsychology. Patient NK, studied by Calder et al. (2000), suffered damage to the insula and basal ganglia. In questionnaire responses NK showed himself to be selectively impaired in experiencing disgust. He was also significantly and selectively impaired in *recognizing* -- that is, *attributing* -- disgust. It is hard to explain why NK would have this *paired deficit* unless the experience of disgust is (normally) causally involved in its attribution. A paired deficit in experience and attribution of disgust seems to be most readily explicable on the assumption that disgust attribution (in observational circumstances) is mediated by its experience. In other words, a normal person uses his intact disgust-experience system to attribute disgust to others. Note that NK was normal with respect to attributing other emotions via observation of facial expressions. Nor was a visual deficit a possible explanation, since NK had the same selective deficit in attributing disgust based on nonverbal *sounds*. Similar findings exist with respect to fear and the amygdala (Goldman & Sripada 2005; Goldman 2006: 115 ff).

TT proponents have not offered any systematic account of these findings. One cannot appeal to damage to a hypothesized theorizing system to account for the disgust-attribution impairment, because the relevant patients performed normally when attributing other emotions based on facial or auditory stimuli. Is there a separate theorizing system for each distinct emotion, and was such a theorizing system coincidentally impaired when disgust experience was impaired? Recalling our criteria of adequacy proposed in section 1, the absence of any story of face-based emotion mindreading is a significant count against TT.

Sensations such as pain are another sub-category of low-level mindreading. The most relevant studies here are by Avenanti et al. (2005, 2006). When a participant experiences pain, motor evoked potentials (MEPs) elicited by transcranial magnetic stimulation (TMS) indicate a marked reduction of corticospinal excitability. Avenanti and colleagues found a similar reduction of excitability when participants merely observed someone else receiving a painful stimulus, for example, a sharp needle being pushed into his hand. Thus, there appeared to be mirroring of pain in the observer. Moreover, when Avenanti and colleagues had participants judge the intensity of pain purportedly felt by a model, judgments of sensory pain seemed to be based on the mirroring of pain experienced by the participant.

The conclusion that mirrored pain is the causal basis of pain attribution is clouded a bit, however, by Danziger et al.'s (2006) findings from twelve patients with congenital insensitivity to pain. Compared to normal controls in pain recognition tasks, these patients did not differ much in their estimates of the painfulness to other people of various verbally described events.

Nor did they differ much from controls in their estimates of degree of pain judged on the basis of facial expression. However, as Carruthers (2011) points out, these individuals with congenital insensitivity to pain may have acquired a different route to pain mindreading than normal people. The findings do not disprove the hypothesis (which Danziger and colleagues embrace) that normal subjects use simulation in reaching their judgments of pain attribution.

Carruthers presses problems for another putative example of low-level simulational mindreading, one concerning face-based mindreading of fear. Adolphs studied a patient, SM, who had a paired-deficit for fear perfectly analogous to the one for disgust displayed by NM. SM, who suffers from bilateral amygdala damage, lacks normal experience of fear and was also selectively impaired in fear attribution. This suggests that the mindreading of fear, like disgust, is ripe for interpretation in simulational terms. However, a later study of SM (Adolphs 2005) showed that she was abnormal in scanning her target's eye areas. When she was directed to scan the eyes thoroughly, she improved on fear attribution. Thus, use of fear experience is apparently not strictly necessary for face-based fear attribution, which ostensibly runs counter to a low-level simulational story for fear attribution. However, we can make a similar hypothesis about this case as Carruthers did for those patients congenitally insensitive to pain. Perhaps SM simply developed (under instructions) a skill for face-based mindreading that differs from the simulation heuristic used by normal subjects.

Moreover, other patients with amygdala damage have been studied, with results that support the ST story. Sprengelmeyer et al. (1999) studied patient NM, who showed selective fear-recognition impairment not only using visual face observation but also using postural and vocal emotional stimuli. These recognition impairments of NM cannot be explained by appeal to inadequate facial scanning, because the targets' eyes were not visible during the bodily posture task, and played no role in the vocal expression task. So it seems that fear impairment due to amygdala damage does indeed provide a causal explanation of NM's recognition impairment, in conformity with the ST account.

Even if one grants that mental attribution in these cases is caused by the mirroring of others' emotions and sensations, one might balk at the idea that this qualifies as *simulation*-based attribution. Why does it so qualify? Here is our answer. Consider the diagram in Figure 3, where an unshaded shape (on the left side of the diagram, depicting mental states of the target) represents an actual occurrence of disgust and a shaded shape (on the right side of the diagram, depicting states of the observer) represents an observation-induced disgust experience. Just as the Figure 2 mindreader imputes a specific decision to her target because she herself "makes" that very decision, so the Figure 3 observer undergoes a mirrored experience of disgust, classifies it *as* an instance of disgust, and projects -- i.e., imputes -- it to the target. Such a projection of a self-experienced state is a signature of simulational mindreading. Thus, it seems reasonable to regard this as a species of simulation and simulation-based mindreading, even though it is distinguishable from high-level simulational mindreading in some respects (for example, by not being a product of imagination). Both cases involve a process of (genuine or attempted) mental matching between attributor states and target states. The mirror-produced mindreading may be called a *single-step* simulation process because simulation takes place more or less directly, whereas the decision case is a *multi-step* simulation process.

Our discussion thus far reviews older evidence pertaining to high- and low-level simulational mindreading. In the remainder of the chapter we adduce more recent lines of evidence and assess their bearing on simulation theory and its competitors.

5. The Power of Imagination

As noted in section 3, ST implies that mindreaders' success in high-level mindreading depends on their ability to enact starting states that sufficiently match those of the target. It follows that if simulational mindreading is to succeed, imagination must be a highly precise mechanism, capable not only of generating suitable pretend states but of firmly holding their progeny in mind while a multi-step simulational exercise unfolds. What is called for is no minor feat. Is the human imagination powerful enough to meet the challenge?

What exactly does it mean to say that we imagine things from another person's perspective? In what sense are mindreaders capable of imagining how a target thinks or feels? The sense of imagination we have in mind is a kind of *enactment imagination*, or *E-imagination* (Goldman, 2006: chap. 7). To E-imagine a state is to recreate the *feeling* of a state, or conjure up what it is like to experience that state—in a sense, to *enact* that very state. To E-imagine feeling embarrassed involves using one's imagination to create inside oneself a pretend state that phenomenally feels somewhat like embarrassment. This enactment sense of imagination should be distinguished from another everyday notion of imagination that consists in imagining *that* such-and-such is the case (as if someone asked you to imagine *that* you are embarrassed.). This ordinary sense of "imagine" means something like *suppose* or *assume that* you are embarrassed—which merely requires you to think about or consider a hypothetical situation of embarrassment. It does not require you to conjure up in yourself something resembling the feeling of embarrassment.[4]

Inspired by evidence of similarities between perception and mental imagery, researchers recently conducted a study to test the effect of imagined eating on actual subsequent eating

[4] We intend our use of the word 'imagination' to be understood in the E-imagination sense.

(Morewedge et al., 2010). The results indicate a striking similarity between the states that result from actual eating and those that result from merely imagined eating—namely, both activities result in habituation to the presented stimulus. A series of experiments were conducted to ensure that it was in fact the act of imagining eating that led to a decrease in consumption, and that this decrease in consumption was indeed an effect of habituation. The important experiments and results for our purposes are summarized below.

The first experiment was designed to test whether repeatedly imagining consuming a particular food would influence subsequent consumption of that food. Participants were divided into three groups, each of which imagined performing 33 repetitive actions. The control group imagined putting 33 quarters into a laundry machine (an action similar to putting M&M's in one's mouth); the second group imagined putting 30 quarters into a laundry machine and then eating 3 M&M's; and the third group imagined putting 3 quarters into a laundry machine and then eating 30 M&M's. All participants then ate freely from a bowl of M&M's until they indicated that they were finished eating. How much each participant ate from the bowl was measured and compared. The results showed that participants who imagined eating 30 M&M's subsequently ate significantly *fewer* M&M's than participants in the other groups.

Another experiment tested whether the decrease in consumption was due to habituation or if it was a priming effect resulting from repeated exposure to the stimulus. This time participants either imagined eating 3 or 30 M&M's or they imagined putting 3 or 30 M&M's into a bowl; then, as before, participants ate freely from a bowl of M&M's. Again, results revealed that subjects who imagined eating 30 M&M's ate significantly less than those who only imagined eating 3; but results also showed that subjects who imagined putting 30 M&M's in a bowl ended up eating significantly *more* than subjects who imagined putting only 3 in a bowl.

This experiment strongly suggests that not only is priming *not* the cause of the decrease in consumption, but may even have the opposite effect of *increasing* subsequent food intake. A further experiment was designed to determine if imaginary eating was habituating people to particular food (causing them to eat less of it), or if it was an overall primed feeling of "fullness" responsible for the decrease in food intake. Here participants imagined eating either 3 or 30 M&M's or cubes of cheese, and then ate freely from a bowl of cheese cubes. The participants who imagined eating 30 cubes ate significantly less than those who imagined eating 3; but participants who imagined eating 30 M&M's did not differ in subsequent cheese consumption from those who imagined eating 3 M&M's. Thus, it seems that the effect of imaginary eating is *stimulus specific*—providing additional evidence that the reduction in food intake is a result of habituation, and not of priming.

The results of this study are remarkable. Merely imagining eating can impact how much we actually eat. But how is a study concerning food consumption relevant to mindreading? We argue that this study is easily interpreted as a demonstration of the power of imagination, and to that extent supports our version of ST. In order for the case to be convincing, there are two things that need to be established: (1) the study's use of imaginary eating counts as an instance of imagination in our *enactment* sense of the word, and (2) the states generated by the imagination really do appropriately resemble their actual counterparts. Our first task is relatively straightforward given the study's experimental design. Participants in the first experiment were asked to repeatedly imagine themselves eating units of food one at a time, not merely to imagine *that* they had eaten a certain amount of food. Furthermore, results indicated that merely thinking about a particular food repeatedly was not enough. For the habituation effect to occur a person had to actually imagine undergoing the experience of eating a particular food. But this is just how we have characterized an act of enactment imagination: as an attempt to re-enact or re-experience a particular feeling or state.

What about our *second* task? Have Morewedge and colleagues shown that imagining eating is capable of producing accurate pretend states, similar to those that result from actually eating? When a person actually eats a particular food, they gradually habituate. Their desire to eat the food, along with the motivation to obtain it, gradually decreases. If presented with a different food, however, the person's desire and motivation recover. This implies that habituation effects are stimulus specific (Epstein et al., 2003). If this is what happens when we actually eat, then something sufficiently similar to habituation should result when we repeatedly imagine ourselves eating: Additionally, it should be the case that if we imagine eating something only a few times we do *not* habituate. The results of the Morewedge et al. study clearly demonstrate that imagined eating results in habituation, similar to when a person has actually eaten. Furthermore, given that habituation effects are stimulus specific the imagined consumption of a particular food should cause a person to habituate to that food only. As the study demonstrates, this is exactly what happens.

Imagined consumption is a clear instance of enactment imagination as well as of the resemblance that can obtain between imagination-induced states and their genuine counterparts. But imaginary eating is not a case of mindreading. Can more be done to make the connection between this research and simulational mindreading clear? After all, if ST is right, mindreaders use their imagination in tasks involving a variety of mental states. So we need to establish that imagination can produce pretend states that closely resemble actual states across a respectable spectrum of cases. We maintain that research on imagined consumption gives us reason to think the imagination has this capacity.

According to Morewedge et al., “Habituation to food occurs too quickly for it to result from digestive feedback, so it is commonly thought to occur as a result of top-down cognitive processes (such as beliefs, memories, or expectations) or pre-ingestive sensory factors (such as texture or smell)” (2010: 1531). This study demonstrates that habituation can occur as a result of imagination alone, without any influence from sensory information. This is significant because habituation is a very general phenomenon, specific neither to food nor to eating. Research indicates that we habituate to a wide range of complex emotions, attitudes, feelings, and moods: from states of happiness and love to states of fear and anxiety (Solomon 1980). This study confirms that the power of imagination could be very general indeed. [5]

6. **Mindreading Acquisition**

Recall that our second criterion of adequacy requires a comprehensive theory of mindreading to give a plausible, empirically sustainable story about how the mindreading capacity is acquired. Past research on the time-course of childhood mentalizing had important implications for theories of how mindreading is acquired, such as the child-scientist approach to theory-theory. This approach claims that mental-state attributions are driven by naive psychological theories that are initiated and gradually revised in the early years. This claim has

[5] Additional confirmatory evidence comes from research on memory distortion involving imagination (Schacter, Guerin and St. Jacques 2011). A study by Mazzoni and Memon (2003) indicated that the strength of subjects’ beliefs that events occurred increased more when they imagined events than when they simply read about them. Nash et al. (2009) showed that imagining that one has performed an act produces about as many false memories of actually having done it as viewing a doctored video that suggests that one did perform the act.

been increasingly undercut, however, by recent research revealing sensitivity to false beliefs even in pre-verbal infants (e.g., Onishi and Baillargeon 2005).

How does the simulation theory fit with such evidence? What is ST's position on the acquisition of the mentalizing capacity? ST does not take a firm stance vis-à-vis nativism. It is prepared to "go with the flow" of evidence. For example, it is prepared to say that the processes or methods of mindreading (or dispositions to use such processes) are part of our native endowment. It might be more skeptical about claims that particular mental-state concepts (belief, desire, pleasure, etc.) are all innate. It is prepared, however, to accommodate the former type of nativism if empirical studies provide warrant for this position. ST's theoretical apparatus does not preclude such strands of nativism. Indeed, one might say something stronger from the vantage-point of ST. If imagination is an innate capacity, perhaps young infants automatically compute imaginary states for people around them. Then we would expect the practice of generating imaginary states to be no more cognitively demanding than one's own largely automatic production of mental states. This section will discuss how well such expectations comport with recent evidence in developmental psychology. Our primary focus is a compelling new study conducted by Kovacs et al. (2010) plus the simulational hypothesis we claim to be consistent with this study.[6]

Unlike standard false-belief tasks, this study was designed to investigate mindreading mechanisms *implicitly*—making no reference to others’ beliefs, and not requiring any behavioral predictions based on others’ beliefs. The study had two components: one testing the reaction

[6] Although we focus on the results of the Kovacs et al. study, we find the wording of their study potentially misleading and worry that it may not convey the researchers’ true intentions. Because of this, we will outline the results as reported in the study, but the conclusions drawn will be our own and are not intended to match those of the authors.

time (RT) of adult participants and the other measuring the looking time (LT) of 7-month-old infants. Participants watched a series of short movies involving an animated agent, a ball, and a table with an occluder. At the beginning of each movie the animated agent entered the scene and placed the ball on the table in front of the occluder.[7] The ball then rolled behind the occluder. At this point, depending on the experiment, the ball either stayed in place or rolled off the screen. Then the agent left the scene. The ball's final location and the time the agent left the scene were varied, such that the agent would have a true belief about the ball’s location if he left after the ball reached its final location and a false belief if he left before. The critical variables involved the participant's beliefs about the ball’s presence or absence and the agent's ‘beliefs,’ such that the participant, the agent, both, or neither could believe the ball was behind the occluder (Kovacs 2010: 1831). At the end of the movie, the agent reentered the scene and the occluder was lowered, revealing the ball to be either present or absent. Adult participants were instructed to press a button as soon as they detected the ball. Their RTs and the infants’ LTs were measured in each of the four conditions.

The experimental conditions, for both adults and infants, were compared to a baseline condition where neither the participants nor the agent believed the ball to be behind the occluder.[8] The most important experiment with adult participants was one in which only the agent believed the ball to be behind the occluder. Results indicated that participants’ RTs were faster in this case than in the baseline condition, despite no difference in the participants' beliefs in either condition. Kovacs et al. take this result to demonstrate that the participants not only automatically computed the agents’ beliefs but that these beliefs influenced the participants'

[7] ‘Agent’ always refers to the animated character in the film and ‘participant’ refers to the adult or infant participant in the study.

[8] However, recall that in none of the conditions were the agent’s beliefs relevant to the task.

behavior, despite the agents' beliefs being inconsistent with their own (1832). Additionally, the participants' RTs did not significantly differ when only the agent believed the ball to be behind the occluder and when they themselves believed it to be. Kovacs et al. further conclude: "Thus both types of belief representations speeded up the participants' RTs to similar extents, a result consistent with the view that the agent's beliefs are stored similarly to participants' own representations about the environment" (2010: 1832).

The crucial results with the infant participants similarly involved a comparison between the infants' LTs in two conditions: one in which only the agent believed the ball to be behind the occluder and the other in which neither the infant nor the agent believed the ball to be there. When no ball appeared behind the occluder, the infants looked longer (indicating their 'surprise' by the outcome) in the condition where *only* the agent believed, or expected, the ball to be there. Again, this suggests not only that the infants computed the agent's belief but also that this belief influenced the infants' behavior despite conflicting with their own (genuine) beliefs. It is similarly interesting that with both adults and infants, very similar results obtained even when the agent did not return to the scene and thus was not present when the occlusion was lowered. Infants and adults seemed to compute and maintain the agent's beliefs even when the agent was no longer present.

What do these results mean for the study of mindreading? More specifically, how do they fit with what ST says about mindreading? Concerning the question of acquisition, the infant results are of primary interest. How do they fit with theory-driven versus simulation-driven mindreading processes depicted in Figures 1 and 2 respectively? A Figure 1-type story would say that at 7 months of age infants not only compute the beliefs of other agents but that these computations are based on the infants' beliefs *about the beliefs of the agen*t. In other words, TT-type explanations rely on the infants' possession of relatively complex metarepresentational states, plus their possession of some body of psychological laws or generalizations.[9] Thus, TT's approach to mindreading is information rich, and requires a degree of cognitive or informational sophistication that one be may hesitant to attribute to 7-month-old infants.

By contrast, the Figure 2-type story suggests that the same sort of tracking of the agent's thoughts has another, simpler interpretation. ST implies that infants track an agent's perspective in the same way they maintain their own perspective. Just as infants have their own current representations of the environment, they also track other possible states of the environment. This sort of explanation, in contrast to Figure 1-type theories, is an information-poor approach, because it does not attribute to the infants any additional theoretical knowledge or metarepresentational states. To perform perspective computations, ST only requires that infants possess states with *object-level* representational content— information about the way the world seems from the shoes of the agent. This means that infants may track the content of an agent's belief (possible states of the environment) without encoding anything concerning his beliefs or other mental states.

Given what we have said so far, ST is in as good a position as TT to account for the Kovacs study. Might there be reasons to think it may be in a *better* position to explain its findings? We argue that there are. First of all, TT has to say that pre-verbal infants compute metarepresentations. Is it psychologically plausible to impute such cognitively complex mental states to infants? Wouldn't it be preferable, if possible, to account for the infants' behavior without attributing to them such extra computational work or informational baggage? If so, then

[9] Depending on the particular TT-type approach we are discussing, such theories may also require that infants possess other complex theoretical beliefs about human psychology.

the ST explanation is clearly preferable, because it accounts for the evidence without positing the extra complexity of metarepresentational states or a body of psychological generalizations.

Concerning the question of acquisition, there are other reasons to think the Kovacs study supports a simulation story about mindreading. What this study shows, we have claimed, is that 7-month-old infants generate representations of the world that reflect another person's perspective—but to represent the world as it seems to another person just *is* to use one's imagination.[10] Although it seems unlikely that at 7 months infants engage in explicit acts of mindreading (i.e., attribution of mental states to others), they certainly appear to engage in mindreading-*like* activity; furthermore, this mindreading-like activity involves use of their imagination. This means that before they ever engage in a single act of mindreading, infants are already experienced imaginers. By the time they get to the point of *attributing* mental states to other people, they have spent years spontaneously and automatically imagining the world from other people's perspectives.

7. The Neural Basis of Mindreading

Now we apply the third question of adequacy to our version of simulation theory: Is this theory *neurally plausible*, given available empirical evidence? One issue is whether recent evidence from cognitive neuroscience supports (or is consistent with) the claim that simulation is a common method, if not the predominant method, of mindreading. A second issue is whether neuroscience supports our specific version of ST, i.e., a bi-level or duplex version of ST.

[10] This is where our conclusions may come apart from those drawn by Kovacs et al. While the conclusions drawn in the study seem to claim that 7-month-old infants are representing the beliefs of the agent alongside their own beliefs; we claim that the results of this study only demonstrate that infants are generating representations of the world that reflect what the target's or agent's beliefs *would be* (not that they represent them *as* beliefs). Furthermore, the study itself makes no mention of the imagination—rather, the results demonstrate that infants engage in an activity that epitomizes our conception of the use of imagination in simulation.

Because neuroscientific evidence was already adduced in support of the existence of mirroring and the grounds for linking it to ST, we won’t say more about the first issue. We shall concentrate on the second.

Waytz & Mitchell (2011) present the neuroscientific case for a duplex model of simulational mindreading as follows. First, they review the extensive evidence of multiple mirroring phenomena, sometimes referred to as "shared neural representations." These include regions in the inferior frontal cortex and superior parietal lobe (i.e., the parieto-frontal circuit) which are involved in the production and observation of goal-directed motor action.[11] They also include a wide range of regions for the mirroring of pain, touch, disgust, and fear (cf. Rizzolatti & Sinigaglia 2006). Networks in these areas are what we treated under the heading of low-level simulation.

Another set of brain regions has been identified, however, that serves as a substrate for what Buckner & Carroll (2007) call *self-projection.* These regions, known collectively as the "default network" (Raichle et al., 2001), consist of the medial prefrontal cortex, precuneus and posterior cingulate, and lateral parietal cortex. The default network has repeatedly been linked to tasks in which people imagine experiencing fictitious events, consider the possibility of experiencing specified events in the future, or recall their experiences from the past. It has also been reported by studies in which participants contemplated other people's mental states (Frith & Frith, 2003). Thus, the default network is an excellent candidate for the neural substrate of high-

[11] The mirroring theory has characteristically claimed that mirroring is used to understand the actions of others. There is continuing debate, however, over which specific brain networks comprise the action-observation system, and how exactly they function. For example, Kilmer (2011) defends a two-pathway model of action understanding, featuring a dorsal pathway in addition to the initially discovered ventral pathway.

level simulational mindreading. Moreover, this network seems to be quite different (non-identical) from any of the circuits or processes involved in mirroring.

Finally, Waytz & Mitchell point out that there are dissociable functions of mirroring and self-projection. Perceivers mirror only when they see or hear another person's physical actions, observe an emotional expression, or witness a painful situation such as a needle penetrating a hand. But mindreading also occurs when subjects represent targets who are not immediately present, and hence aren't observable. Such targets include fictitious individuals or individuals known only by description, where no observable cues are available. Waytz & Mitchell consider this a demonstration of a dissociation between mirroring and self-projection. They cite a mentalizing study by Zaki et al. (2010) in which participants inferred a target's emotional state under three conditions: during perceptually cued trials, during context-only trials, and when participants had both perceptual and contextual information. Consistent with the proposed division of labor between two systems of mentalizing, they found that perceptual cues tended to elicit stronger activation in mirror-related brain regions (the fronto-parietal circuit) whereas contextual cues engaged the default network.

Lombardo et al. (2010) take a somewhat different perspective on the two-systems approach based on their finding of functional connectivity patterns during mentalizing of both self and other. They don't deny the dissociability claim of Waytz & Mitchell , but they argue that the functional connectivity patterns revealed in their studies support a slightly different picture than the one offered here. Indeed, they advance the thesis that some aspects of both high-level and low-level social cognitive processes are "grounded" within a framework of embodied cognition.   We don't believe that there are fundamental differences between their view and ours. At any rate, we find no reason to disagree with a very similar picture presented by Zaki and

Ochsner (2012), who also stress functional connectivity between the two systems during experiences of empathy. As Zaki and Ochsner express it, "naturalistic" (i.e. ecologically valid) situations involve many dynamic social cues (featuring both sensorimotor and contextual information), and such cues unsurprisingly generate dynamical neural interactions among simpler processes (low-level and high-level processes). These more complex processes could not be understood, Zaki and Ochsner acknowledge, without a prior understanding of the simpler processes in isolation, which are coupled during complex social tasks (2012: 678). By our lights, this is a reasonably clear recognition that there *are* simpler processes, which we take to be the low-level and high-level families of processes of our model.[12] It is the existence and distinctness of these "simpler" processes that comprise the core thesis defended in this section.

8. ST's "Mesh" with Evolutionary Theory

We turn finally (and briefly) to the fourth question of section 1: Does our theory mesh with successful theories in other cognitive domains and with plausible accounts of the architecture and evolution of cognition? Do successful theories of other parts of cognition invoke similar explanatory faculties or processes, and does a reasonable account of brain evolution find a natural home for simulationist stories of mindreading?

Begin with ST's account of high-level mindreading, in which imagination occupies a central role. It makes good sense, we submit, to assign a pivotal role to imagination because this faculty has demonstrated its power and versatility in many other domains of cognition. Its robust

[12] One non-trivial point of difference, however, is that Zaki and Ochsner identify the higher-level processes as "mentalizing" processes, implying that lower-level ("shared representation") processes are not involved in mentalizing. By contrast, we claim that the latter also serve as a causal basis of mentalizing.

power and versatility are amply exhibited in such diverse phenomena as visual and motor imagery, the planning of action sequences, and the reduction of food consumption. With respect to low-level mindreading, the discovery of mirror neurons and mirror systems has revolutionized research and thinking about many aspects of low-level cognition (Rizzolatti and Craighero 2004; Gallese et al. 2004). Contemporary social neuroscience is replete with new insights related to mirroring. A primitive kind of mindreading based on mirroring is a good fit with much of this literature.

ST also comports well with current understandings of brain evolution. As Anderson (2010) tells the story, it is very common for neural circuits originally established for one purpose to be exapted -- that is, exploited, recycled, redeployed -- during evolution and put to different uses, without necessarily losing their original functions. Nature has a pervasive strategy of opportunistically exploiting existing neural hardware to solve new problems – or to create new solutions to old problems. Creating whole brain structures *de novo* in order to tackle problems would be expensive. Instead, nature prefers a redeployment strategy.   This idea meshes well with ST's story of low-level mindreading. So, for example, suppose that nature had earlier hit upon the strategy of devising mechanisms by which shared representations are generated in the heads of two interacting individuals. A mental representation (or event) in one individual's brain leads to the generation of a matching representation (or event) in an observer. Once this kind of interpersonal transmission mechanism has evolved, members of the species can secure valuable information by piggy-backing a mental attribution mechanism on top of the shared-representation, or mirroring, mechanism. This is a cheap way to create a reliable mindreading device. It would be unsurprising if something like this evolutionary story were true.

This is what we mean in saying that ST "meshes" well with what is known, or reasonably believed, about brain evolution. According to many philosophers of science, consilience with existing theory is one form of evidence for a new theory. Thus, another chunk of evidential support is added in favor of ST, on top of the more direct kinds of evidence presented in preceding sections.

References

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

Adolphs, R., Tranel, D. and Damasio, A.R. (2003). Dissociable neural systems for recognizing emotions. Brain and Cognition 52: 61-69.

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. Behavioral and Brain Sciences 33: 245-266.

Apperly, I. A. (2011). Mindreaders: The Cognitive Basis of Theory of Mind. Hove: Psychology Press.

Apperly, I. A. and Butterfill, S. A. (2009): Do humans have two systems to track beliefs and belief-like states? Psychological Review 116(4): 953-970.

Avenanti, A. Bueti, D., Galati, G. and Aglioti, S.M. (2005). Transcranial magnetic stimulation highlights the sensorimotor side of empathy for pain. Nature Neuroscience 8: 955-960.

Avenanti, A., Paluello, I.M., Bufalari, J. and Aglioti, S.M. (2006). Stimulus-driven modulation of motor-evoked potentials during observation of others' pain. NeuroImage 32: 316-324.

Baker, C.L., Saxe, R., and Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition* 113: 329-349.

Baker, C.L., Saxe, R., and Tenenbaum, J.B. (forthcoming). Bayesian theory of mind: Modeling joint belief-desire attribution.

Baron-Cohen, S. (1995). Mindblindness: An Essay on Autism and Theory of Mind. Cambridge, MA: MIT Press.

Birch, S. A. J. and Bloom, P. (2003). Children are cursed: an asymmetric bias in mental-state attribution. Psychological Science 14: 283-286.

Birch, S. A. J. and Bloom, P. (2004). Understanding children’s and adults’ limitations in mental state reasoning. Trends in Cognitive Sciences 8: 255-260.

Buckner, R.L. and Carroll, D.C. (2007). Self-projection and the brain. Trends in Cognitive Sciences 11: 49-57.

Calder, A.J., Burton, A.M., Miller, P., Young, A.W. and Akamatsu, S. (2001). A principal component analysis of facial expressions. Vision Research 41, 1179-1208.

Camerer, C., Loewenstein, G. and Weber, M. (1989). The curse of knowledge in economic settings: an experimental analysis. Journal of Political Economy 97: 1232-1254.

Carruthers, P. (2011). Opacity of Mind. Oxford: Oxford University Press.

Csibra, G., Biro, S., Koos, O., and Gergely, G. (2003). One-year old infants use teleological representations of actions productively. *Cognitive Science* 27: 111-133.

Currie, G. and Ravenscroft, I. (2002). Recreative Minds. Oxford: Oxford University Press.

Danziger, N., Faillernot, I., and Peyron, R. (2009). Can we share a pain we never felt? Neural correlates of empathy in patients with congenital insensitivity to pain. Neuron 61: 203-212.

Decety, J., Jeannerod, M., and Preblanc, C. (1989). The timing of mentally represented actions. Behavioral and Brain Research 34: 35-42

Decety, J. and Greze, J. (2006). The power of simulation: Imagining one's own and other's behavior. Brain Research 1079: 4-14.

Dennett, D. (1987). The Intentional Stance. Cambridge, MA: MIT Press.

Epstein, L. H., Saad, F.G., Handley, E.A., Roemmich, J.N., Hawk, L. W. and McSweeney, F.K. (2003). Habituation of salivation and motivated responding for food in children. Appetite 41(3): 283-289.

Frith, U. and Frith, C.D. (2003). Development and neurophysiology of mentalizing. Philosophical Transaction of the Royal Society of London. Series B: Biological Sciences, 459: 358.

Gallese, V. (2007). Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition. Philosophical Transactions of Royal Society B: Biology.

Gallese, V. and Goldman, A. (1998). Mirror neurons and the simulation theory of mindreading. Trends in Cognitive Sciences 2: 493-501.

Gallese, V., Keysers, C., and Rizzolatti, G. (2004). A unifying view of the basis of social cognition. Trends in Cognitive Sciences 8: 396-403.

Gergely, G., Nadasdy, Z., Csibra, G. and Biro, S. (1995). Taking the intentional stance at 12 months of age. Cognition 56: 165-193.

Goldman, A.I. (1989). Interpretation psychologized. Mind and Language 4: 161-185.

Goldman, A. I. (2006). Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading. New York: Oxford University Press.

Goldman, A.I. and Sripada, C. (2005). Simulationist models of face-based emotion recognition. Cognition 94: 193-213.

Gordon, R.M. (1986). Folk psychology as simulation. Mind and Language 1: 158-171.

Harris, P. L. (1992). From simulation to folk psychology: The case for development. Mind and Language 7: 120-144.

Heal, J. (1986). Replication and functionalism. In J. Butterfield, ed., Language, Mind, and Logic. Cambridge: Cambridge University Press.

Jeannerod, M. (2001). Neural simulation of action: A unifying mechanism for motor cognition. NeuroImage 14: S103-S109.

Kilmer, J. M. (2011). More than one pathway to action understanding. Trends in Cognitive Sciences 15(8): 352-357.

Kosslyn, S. M., Pascual-Leone, A., Felician, O., and Camposano, S. (1999). The role of area 17 in visual imagery: Convergent evidence from PET from rTMS. Science 284: 167-170.

Kosslyn, S. M., Thompson, W. L. and Alpert, N. M. (1997). Neural systems shared by visual imagery and visual perception: a positron emission tomography study. Neuro-Image 6: 320-334.

Kovacs, A.M., Teglas, E. and Endress, A.D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. Science 330: 1830-1834.

Leslie, A. (1994). Pretending and believing: Issues in the theory of ToMM. Cognition 50: 211-238.

Leslie, A., German, T., and Polizzi, P. (2005). Belief-desire reasoning as a process of selection. Cognitive Psychology 50: 45-85.

Lombardo, M.V., Chakraharti, B., Bullmore, E.T., Wheelwright, S.J., Sadek, S.A., Suckling, J., and Baron-Cohen, S. (2010). Shared neural circuits for mentalizing about the self and others. Journal of Cognitive Neuroscience 22(7): 1623-1633.

Mazzoni, G. and Memon, A. (2003). Imagination can create false autobiographical memories. Psychological Science 14: 186-188.

Morewedge, C.K., Huh, Y.E., and Vosgerau, J. (2010). Thought for food: Imagined consumption reduces actual consumption. Science 330: 1530-1533.

Nash, R.A., Kimberly, A.W. and Lindsay, D.S (2009). Digitally manipulating memory: Effects of doctored videos and imagination in distorting beliefs and memories. Memory and Cognition 37(4): 414-424.

Nickerson, R. S. (1999). How we know – and sometimes misjudge – what others know: imputing one's own knowledge to others. Psychological Bulletin 125: 737-759.

Perner, J. (1991). Understanding the Representational Mind. Cambridge, MA: MIT Press.

Phillips, M. L.,Young, A.W., Senior, C., Brammer, M., Andrew, C., Calder, A.J., Bullmore, E.T., Perrett, D.I., Rowland, D., Williams, S.C.R., Gray, J.A., and David, S. (1997). A specific neural substrate for perceiving facial expressions of disgust. Nature 389: 495-498.

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

Rizzolatti, G. and Craighero, L. (2004). The mirror-neuron system. Annual Review of Neuroscience 27: 169-192.

Rizzolatti, G. and Sinigaglia (2008). Mirrors in the Brain: How Our Minds Share Actions and Emotions. Oxford: Oxford University Press.

Rozin P., Haidt J. and McCauley, C. (2000). Disgust. In M. Lewis and J. Haviland, eds., Handbook of Emotions. New York: The Guilford Press.

Schacter, D.L., Guerin, S.A. and St. Jacques, P.L. (2011). Memory distortion: an adaptive perspective. Trends in Cognitive Sciences 15(10): 467-506.

Solomon, R.L. (1980). The opponent-process theory of acquired motivation: The costs of pleasure and the benefits of pain. American Psychologist 35(8): 691-712.

Sprengelmeyer R., Young, A.W., Schroeder, U., Grossenbacher, P.G., Federlein, J., Buttner, T. and Przuntek, H. (1999). Knowing no fear. Proceedings of the Royal Society, B: Biology 266: 2451-2456.

Tager-Flusberg, H. and Sullivan, K. (2000). A componential view of theory of mind: evidence from Williams Syndrome. Cognition 76: 59-89.

Waytz, A. and Mitchell, J.P. (2011). Two mechanisms for simulating other minds: Dissociations between mirroring and self-projection. Current Directions in Psychological Science 20(3): 197-200.

Wellman, H. (1990). The Child’s Theory of Mind. Cambridge, MA: MIT Press.

Wicker, B., Keysers, C., Plailly, J., Royet, J-P., Gallese, V., and Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. Neuron 40: 655-664.

Zaki, J., Hennigan, K., Weber, J., and Ochsner, K.N. (2010). Social cognitive conflict resolution: Contributions of domain-general and domain-specific neural systems. Journal of Neuroscience 30: 8481-8488.

Zaki, J. and Ochsner, K.N. (2012). The neuroscience of empathy: progress, pitfalls and promise. Nature Neuroscience 15(5): 675-680.