The happiness of the fish:

Evidence for a common theory of one's own and others' actions

Rebecca Saxe

Department of Brain and Cognitive Sciences, MIT

To appear in: The Handbook of Imagination and Mental Simulation

Edited by Keith Markman, Bill Klein and Julie Suhr

Contact:

Rebecca Saxe
46-4019, MIT
43 Vassar St
Cambridge MA 02139
(617) 324 2885

Chuangtse and Hueitse had strolled onto the bridge over the Hao, when the former observed, "See how the small fish are darting about! That is the happiness of the fish." "You are not a fish yourself," said Hueitse. "How can you know the happiness of the fish?" "And you not being I," retorted Chuangtse, "how can you know that I do not know?" ~Chuangtse, c. 300 BC

Through introspection, we feel (cf. Nisbett & Wilson, 1977a), that we have direct knowledge of the causes and goals of our own actions; understanding someone else's action seems by contrast like a highly abstract -- if not semi-miraculous -- achievement. Simulation theories offer a demystification of the process: knowledge of others is parasitic on our direct access to ourselves (Gallese, Keysers, & Rizzolatti, 2004; Rizzolatti, Fogassi, & Gallese, 2001). An observer can understand someone else's action using the same cognitive and neural mechanisms that she uses to produce her own – that is, by running her action execution system in a "simulation" mode (Nichols, 2003).

There is a fundamentally different sense in which we use the same mechanisms to understand others and ourselves, though. We explain and predict our own actions just like we understand other people's, by using a theory of how human minds work. So, the central Simulationist claim that we use the "same mechanisms" for understanding our own and other people's actions may be true in at least two unrelated senses. The confusion arises because of two senses in which a person can "understand her own action" (Gopnik, 1993). In one sense, an actor "understands" all of her current, ongoing intentional actions. That is, when the person is acting rationally in pursuit of her own goals, she has responsibility for her actions. However,

there is another sense of "understanding" actions that involves being able to provide verbal reasons for or causes of that action (Malle, 2004). It is in the latter sense, and not the former, that representing and understanding one's own actions depends on a Theory of Mind (Happe, 2003).

In the current chapter, I therefore propose that among mechanisms for understanding human action, the cognitively relevant distinction is not between self and other, but between action execution and perception on the one hand, and action explanation and prediction on the other. Humans possess two distinct cognitive and neural mechanisms for representing actions: the sensorimotor mechanisms for planning, executing, and perceiving goal-directed actions online, and distinct cognitive mechanisms for explaining and predicting actions in terms of a Theory of Mind. Both mechanisms can be applied to others' actions, and to one's own. Evidence for the first mechanism is provided in other chapters of this Handbook (e.g., Beilock & Lyons, ch. X; Decety & Stevens, ch. X). Below I review developmental, social psychological, and neuroscientific evidence for the second mechanism, Theories of Mind that are used to understand both other people's actions, and one's own.

Theories apply to Self and Other: Developmental Evidence

When the actor has a false belief, action predictions based on a conceptual Theory of Mind diverge most obviously from predictions based solely on facts about the local environment (Dennett, 1978). For this reason, many studies of Theory of Mind development require children to make action predictions given a false belief. In one basic design, a child watches while a puppet places a toy in location A. The puppet leaves the scene and the toy is transferred to location B. The puppet returns and the child is asked to predict where the puppet will look for the toy. Three-year-olds predict the puppet will look in location B, where the toy actually is;

older children predict the puppet will look in location A, where the puppet last saw the toy (Wellman, Cross, & Watson, 2001; Wimmer & Perner, 1983).

The striking feature of this developmental pattern is not that five-year-olds pass while three-year-olds fail; performance on most tasks improves with age. What is notable is that the three-year-olds who fail the false belief task are not performing at chance, or confused by the questions. They make systematic predictions, with high confidence (Ruffman, Garnham, Import, & Connolly, 2001).

These results have been described in terms of the development of a concept of belief (Wimmer & Perner, 1983). According to the mature concept, a belief is a constructed representation of the world. It is supposed to be true, and a determinant of the believer's actions, but having a correct belief depends on having current perceptual access and/or reliable sources of knowledge; when these are missing, beliefs can be partially or entirely false, causing predictable mistakes in action. By contrast, the younger children's Theory of Mind doesn't include a complete understanding of access and reliable sources (O'Neill, 1992), and so does not flexibly accommodate error and misrepresentation (Perner, 1991).

The traditional false belief task is of course subject to other interpretations. For example, predictions based on false beliefs require children to inhibit the salient true state-of-affairs so developmental trends in prediction may reflect the development of domain-general inhibitory control (Carlson, Moses, & Claxton, 2004; Leslie, 2000; Moses, 2001). The interpretation of these results in terms of the development of a Theory of Mind therefore requires support from other methods; one approach is to focus on action explanation (e.g. why did the puppet look for the toy in location A?) (Bartch, 2007; Hickling & Wellman, 2001) instead of action prediction.

Development of the ability to explain actions in terms of thought and beliefs is correlated with, and precedes, success in action prediction (Amsterlaw, 2006). In general, children who fail to predict future actions based on false beliefs do not explain past actions in terms of false beliefs (Moses, 2001). Instead, they explain actions in terms of desires and other psychological states (Bartch, 2007). For example, (Goodman, 2006) gave children the standard false belief prediction task, but then after the prediction showed children the character looking for the object in the opposite (unpredicted) location. Children were then asked to explain the character's actions. The content of these explanations was theoretically consistent with the child's original answer. The children who predicted that the character would look in the actual location (B) and then saw the character look in the original location (A, the "standard" outcome), explained this action by generating a novel desire (e.g. "well, that's where she wants to look"), and not by appeal to the character's false belief. By contrast, children who predicted that the character would look in the original location (A) and then saw the character look in the actual location (B, the "psychic" outcome), explained this action by generating a novel source of access to the true location (e.g. "I think he heard his sister going over there"). Given these results, it would clearly be misleading to claim that five-year-olds "have" a Theory of Mind whereas three-year-olds do not have one (Bloom & German, 2000). The younger children's theory is coherent, but limited.

If children apply their Theory of Mind to explain and predict their own actions, as well as those of others, then the same limitations should appear. Versions of the false belief task have been developed to explicitly tap children's ability to attribute false beliefs to themselves. On direct tests, three-year-olds do make the same systematic errors about their own past false beliefs as they make about false beliefs of other people (Gopnik, 1993; Gopnik & Astington, 1988). For example (Gopnik & Astington, 1988), children saw a candy box, and then discovered that it was

filled with pencils. They were then asked what they thought was in the box when they first saw it. The youngest children reported that they initially thought the box contained pencils and predicted that other people would think the box contained pencils; success on the first- and third-person versions was correlated.

These results are counter-intuitive. We might expect children to just remember their previous thoughts, and so to have qualitatively different, and better, access to the mental state explanations of their own actions than those of others. But the evidence suggests that they do not.

In an elegant recent study, (Atance & O'Neill, 2004) gave three-year-olds an opportunity to make plans and act based on a false belief. Then, immediately after the action, the children discovered the true state of affairs and were asked to explain their own immediately past action. On one trial, for example, the child was shown a crayon box, and asked to say what she thought was inside. When she said "crayons," the experimenter pointed out a piece of paper and suggested the child retrieve the paper so she could draw with the crayons. Then the box was opened to reveal candles. The child was asked what she initially thought was in the box and, consistent with prior work (Gopnik & Astington, 1988), most children said "candles." Amazingly, though, when asked to explain why they picked up the piece of paper, these children still failed to refer to their prior belief (saying instead, e.g. "Because there was [sic] no crayons," or "There it was on the floor.")

Children's explanations of their own past actions thus show no benefit of direct recall or privileged access. Action understanding does not appear to develop from direct access to the "self" into inferential understanding of "others." The conceptual content is consistent, at a developmental stage, in explanations of both one's own and other people's actions. These results

are best understood in terms of a common theory of one's own and other minds (Happe, 2003). A similar lesson can be drawn from social psychological evidence in adults, to which I now turn.

Seeing the Self as an Other: Social Psychological Evidence

Like children, adults explain and predict their past, future and hypothetical actions using the same Theory of Mind that they use for other people. Because adults' theories are so much more robust than children's, though, the traces of theory use are fairly subtle. Nevertheless, by creating experimental situations in which the true explanation of a behavior is ambiguous (Bem, 1967), social psychologists have found systematic evidence that adults reconstruct the best explanation for their own past behavior from current evidence, rather than introspecting and recalling the conscious experience of the event (Nisbett, 1977).

Daryl Bem was the best known advocate, in social psychology, of the view that adults infer the internal reasons for their actions from the externally observable evidence of the actions themselves. Consistent with the current argument, Bem proposed that "an individual's belief and attitude statements and the beliefs and attitudes that an outside observer would attribute to him are often functionally equivalent in that both sets of statements are 'inferences' from the same evidence" (Ben 1965, p 200).

Bem's self-perception theory was initially intended to provide an account of "cognitive dissonance" phenomena, in which a person's self-attribution of beliefs and attitudes is influenced by the person's perception of their own actions. One example is "spreading of alternatives": after they are forced to choose between two good options, people subsequently rate the chosen alternative more favorably and derogate the unchosen option (Festinger 1957). A more counter-intuitive instance is the forced compliance effect. After participants comply with a request to do an action that doesn't fit well with their prior beliefs and attitudes (e.g. describe a boring task as

interesting, or write a speech for a political position they do not hold), and only if there was apparently little external pressure compelling the action (e.g. they were apparently given a choice whether to act, and very little monetary reward for acting), participants change their reported attitude, claiming that they found the task less boring, or that they agree more with the political position (for a review, see (Olson & J, 2005). Bem explained these results in terms of self perception theory. Upon observing themselves making a choice, or voluntarily acting without a reward, participants made the same inference that an outside observer would make: that they must really have preferred the chosen option, or found the task interesting.

More recently, social psychologists have directly manipulated the observable consequences of the participants' actions. Again, people's subsequent explanations followed the observable evidence, rather than any direct internal access to the causes of their own behaviour. For example, (Johansson, 2005) showed participants two pictures and asked them to pick the more attractive one. The participants were given the chosen photo, and then asked to explain their choice. Through a sleight of hand, the experimenters sometimes gave the participant their actual choice, but sometimes gave the opposite picture. In all, participants detected the swap very rarely. Remarkably, the justifications that participants gave for choosing the swapped photograph (which they did not choose) were largely indistinguishable from justifications they gave for choosing the one that they did choose, in length, confidence, emotionality, detail, or number of embarrassed laughs. Participants' own explanations appeared to operate much the way an external observer would, by finding the property of the outcome that could have justified choosing it rather than by recalling the moment of the choice (seconds earlier) and directly re-experiencing those reasons. Similarly, Evans (1989) asked participants to first solve a problem and then retrospectively explain their reasoning processes. Participants made systematic errors in

the explanations that were predicted by their theories of reasoning, rather than by the actual reasoning they had just completed.

Attributions to past and future selves are not just distorted; they are also qualitatively similar to attributions to other people (Robinson & Clore, 2002; Trope, 2003). For example, other people's actions are ascribed to stable traits, whereas one's own actions are generally seen as variable and situation-dependent (Jones, 1972). A past or future self, however, is just like another person in this respect: past and future selves are characterized in terms of stable traits (both positive and negative) just as much as past or present other people, and significantly more than the present self (Pronin & Ross, 2006).

In sum, social psychological evidence, like the developmental evidence reviewed above, suggests that the mechanisms people use for understanding action are not divided between direct access to the self and inferential understanding of others. Instead, the relevant distinction is that between mechanisms for action execution and theories for action explanation, each of which are applied both to other people and to oneself (see also Batson, ch. X; Epley & Caruso, ch. X; Myers & Hodges, ch. X). Recently, functional neuroimaging has begun to offer a third converging line of evidence for the importance of this distinction.

Brain Regions for Theory of Mind: Neuroscientific Evidence

The current section describes brain imaging evidence for three claims: (1) there are brain regions implicated specifically in explaining actions in terms of mental state causes (Theory of Mind); (2) these brain regions are distinct from those implicated in action execution and action perception; and (3) these same brain regions are used for attributing mental states to one's self. To date, there is much more detailed evidence for the first of these claims than for the second and third; these will be important topics for future research.

The first step toward understanding the neural basis of a higher-level cognitive function, like Theory of Mind, is to identify candidate brain regions that may be involved in the operation of that function. Some hypotheses may come from lesion studies (e.g. Broca's and Wernicke's areas) or from animal models (e.g. V1). For higher cognitive functions, though, animal models are not available, so hypotheses about region-function links come from early imaging studies using simple subtraction analyses (e.g. for Theory of Mind, Gallagher et al, 2000). The logic of subtraction analyses is as follows: (1) Assume that to perform a complex, high-level task, participants must use many, interacting cognitive mechanisms (and therefore many brain regions). (2) Most of these mechanisms are used for general aspects of task performance, like perceiving the stimuli and producing the response, but some of the processing corresponds to the cognitive mechanism under investigation – in this case, the representation of a mental state. (3) The goal is therefore to find a second task that demands all of the same general processing, with one key difference – in this case, there is no need to think about mental states.

Following developmental psychology, early neuroimaging investigations of theory of mind used false beliefs conveyed in stories or cartoons (e.g. (Gallagher et al., 2000; Saxe & Kanwisher, 2003). A reliable group of brain regions was implicated in the "false belief" condition (relative to a variety of control conditions, described below), including right and left temporo-parietal junction (right and left TPJ), medial parietal cortex (including posterior cingulate and precuneus), and medial prefrontal cortex (MPFC, not including anterior cingulate cortex). Of these regions, the region in the right TPJ appears to be most selectively recruited for thinking about thoughts relative to controls both for the logical demands and for the social content of the task. The closest control condition for the logical demands of the standard false belief task is the "false sign" task. In the false sign task, participants read (or hear) about a sign

or map that is constructed, and then the reality changes, rendering the sign or map out-of-date. For example, a sign is supposed to point to the current location of an ice-cream van, but the van moves and the ice-cream man forgets to change the sign; or a map is made of the locations of all the toys in a room, and then one toy is moved. As in the false belief task, participants can then be asked about the true state of affair (where the toy really is), or about the content of the representation (where the toy is, in the map). In development, performance on false signs tasks is highly correlated with performance on false belief tasks (Sabbagh, Moses, & Shiverick, 2006).

The false sign task therefore provides an excellent control condition for a subtraction analysis. Any brain region recruited more during false belief tasks than during false sign tasks must play a specific role in thinking about people, or mental states. (Perner, Aichhorn, Kronbichler, Wolfgang, & Laddurner, 2006), had participants read four kinds of vignettes, describing false beliefs, false signs, outdated photographs (Saxe & Kanwisher, 2003; Zaitchik, 1990), and changes in reality over time. The right TPJ showed a significantly higher response for the false beliefs than for false signs, but did not differentiate false signs (which involved the logic of false representations) from temporal change stories (which did not). These results suggest that the role of the right TPJ was specific to social/belief components of the false belief task, rather than the other, more general processing demands of performing the task.

In a recent study from my own lab, we tested an alternative minimal-pair control task. A standard false belief task could be solved without considering beliefs, by using simple rules (Povinelli & Vonk, 2003); Bloom and German 2000) (e.g., when asked where the puppet thinks the object is, point to the object's location when the puppet was last facing it). While still posing complex demands on spatial and temporal memory, these rules would not refer to beliefs, or to any mental or social properties. To investigate the neural mechanisms specific to Theory of

Mind, we (Saxe, Schulz, & Jiang, 2006) therefore induced participants to perform a false belief task by following a non-social stimulus-response rule. The stimuli were short animated films of a girl, and a chocolate bar that moved between two boxes. One set of task instructions (the Algorithm rule) instructed participants to use the girl's facing-direction at the end of the trial (away from the boxes versus toward the boxes) as an arbitrary cue to attend to the chocolate's first, or last, location. The other set of instructions (the Theory of Mind rule) asked participants to identify "where the girl thinks chocolate bar is." For any combination of the girl's position and box location, these two rules generated the same response. When participants used the Algorithm rule, only domain-general brain regions (e.g. intra-parietal sulcus, inferior frontal gyrus) were recruited. On the other hand, the right TPJ was recruited specifically when participants were thinking about the girl's thoughts.

We have also observed that the right TPJ response is specific to thinking about thoughts relative to thinking about other facts about people. In one study (Saxe & Powell, 2006), we used stories from three conditions, each highlighting a different aspect of reasoning about another person: (1) 'Appearance' stories that described representing socially relevant information about a person that is visible from the outside; (2) 'Bodily Sensations' stories that elicited attribution of subjective states that do not include a representational content, like hunger and tiredness; and (3) 'Thoughts' stories that described the contents of another person's thoughts. The right TPJ showed a significantly greater response to the thoughts stories than to the appearance and bodily-sensations stories, which did not differ from each other or from fixation. In another study (Saxe & Wexler, 2005), we presented two facts about each character in sequence: the character's social background, and his/her belief or desire. The timing of the response in the right TPJ was precisely dependent on the timing of belief information: when the background information was

presented for 6 seconds before a belief was described, the response in the right TPJ was delayed 6 seconds.

Across studies, then, the right TPJ appears to be recruited whenever the participant is required to think about someone else's belief. Belief attribution can be elicited in at least three different ways: (1) explicitly, when the participant reads a verbal sentence that simply states a character's beliefs (e.g., (Saxe & Powell, 2006); (2) by directing participants to consider a character's beliefs in the task instructions (e.g., (Saxe et al., 2006); or (3) by asking participants to predict the actions of a character who has an inferable false belief (e.g., (Sommer et al., 2007). Action prediction based on true beliefs, however, need not involve any consideration of the character's beliefs (Dennett, 1978). This analysis may explain an apparent contradiction in the literature. (Sommer et al., 2007) used a non-verbal action prediction task modeled on the developmental psychologists' false belief task and found more recruitment of right TPJ for false than for true belief trials. Apperly et al. (Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006) recently found that subjects hold on to belief information in this kind of task only while that information is strictly necessary (i.e., when the character is holding a false belief and that belief is relevant for the subject's own task performance). By contrast, using verbal stories and explicit belief statements, (Young, Cushman, Hauser, & Saxe, 2007) found no difference in the response of right TPJ to true versus false beliefs.

Importantly, the Theory of Mind brain regions – including the regions in the right and left TPJ, and in the medial prefrontal and precuneus – are completely distinct, anatomically, from the brain regions implicated in action execution or action perception (Rizzolatti et al., 2001). Many neuroimaging studies, inspired by Simulation theories, have focused on the overlapping activation during action perception and action execution, of ventral premotor cortex, inferior

frontal gyrus and right inferior parietal cortex (e.g., (Grezes, Armony, Rowe, & Passingham, 2003); (Buccino, Binkofski, & Riggio, 2004); (Molnar-Szakacs, Kaplan, Greenfield, & Iacoboni, 2006; Vogt & Thomaschke, 2007). By contrast, the regions implicated in Theory of Mind have no known role in motor planning or action execution. Instead, these regions are among the latest maturing parts of "association" cortex (Gogtay et al., 2004).

Although the two groups of regions are clearly anatomically segregated, their functional properties have not yet been investigated within a common task. Future work should investigate tasks in which action execution and action explanation are invoked to allow for a direct functional dissociation between these mechanisms.

Research into the neural substrates for explaining one's own past actions is also lacking. The key prediction of the current chapter is that the brain regions that are implicated in Theory of Mind for others (1) would not be recruited while participants actually acted or reasoned based on a false belief; but (2) would be recruited when the same participants subsequently explained those actions in terms of false beliefs. The ideal experimental design would thus be an adult version of Atance and O'Neil (Atance & O'Neill, 2004): participants should be induced to believe and act on one idea, then learn the truth and be asked to explain their previous actions. Nothing like this design has yet been used with neuroimaging.

One related study has been conducted, and the results are promising. (Vogeley et al., 2001) had participants read short verbal stories about a protagonist, half of which described actions and thoughts in the second person (e.g. "In the morning, when you leave the hotel, the sky is blue and the sun is shining. So you do not expect it to start raining."). Because these stories describe non-actual events and actions, the participants could not directly experience the narrated thoughts and plans; instead the participants must have interpreted these as the thoughts

and actions of a hypothetical self. These stories should therefore recruit the same brain regions as are implicated in Theory of Mind for others. Just as predicted, these second-person stories elicited a significant response (relative to a scrambled baseline) in the right TPJ. If anything, the stories in the second person elicited a significantly higher response in Theory of Mind regions than did the same stories in the third person (Vogeley et al., 2001).

Conclusions

Brain imaging results thus converge with those from both developmental and social psychology suggesting that a common Theory of Mind is used for explaining both other people's actions and one's own (see also Batson, ch. X; Epley & Caruso, ch. X; Myers & Hodges, ch. X). Theory of Mind is therefore an instance of a "common mechanism" for representing actions by the self and others. Nevertheless, the Theory of Mind is a fundamentally different kind of shared mechanism from those usually envisioned in Simulation theory. First, Theory of Mind is invoked when providing reasons for actions, but not when choosing or executing actions. Second, access to the reasons for one's own actions is not qualitatively privileged. The clearest behavioral consequences of theory use result from the theories' limitations; especially in young children these limitations apply equally to explanations of one's own actions and those of other people. Importantly, though, Theory of Mind is used in action explanation across a wide range of contexts: for true and false beliefs, and when the theory is accurate as well as inaccurate. These "correct" applications of the theory are hard to detect behaviorally, and are therefore an important target for future studies using functional neuroimaging.

References

Amsterlaw, J., Wellman, H.M.,. (2006). Theories of Mind in Transition: A Microgenetic Study of the Develpment of False Belief Understanding. . *Journal of Cognition and Development, 7*(2), 139.

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychol Sci, 17*(10), 841-844.

Atance, C. M., & O'Neill, D. K. (2004). Acting and planning on the basis of a false belief: its effects on 3-year-old children's reasoning about their own false beliefs. *Dev Psychol, 40*(6), 953-964.

Bartch, K., Campbell., M.D., Troseth, G.L. (2007). Why Else Does Jenny Run? Young Children's Extended Psychological Explanations. *Journal of Cognition and Development, 8*(1), 33

.

Bem, D. J. (1967). Self-Perception: An alternative interpretation of cognitive dissonance phenomena. . *Psychological Review 74*(3), 193-200.

Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition, 77*(1), B25-31.

Buccino, G., Binkofski, F., & Riggio, L. (2004). The mirror neuron system and action recognition. *Brain Lang, 89*(2), 370-376.

Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *J Exp Child Psychol, 87*(4), 299-319.

Dennett, D. (1978). Beliefs about beliefs. *Behavioral and Brain Science, 1*, 568-570.

Gallagher, H. L., Happe, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia, 38*(1), 11-21.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends Cogn Sci, 8*(9), 396-403.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., et al. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proc Natl Acad Sci U S A, 101*(21), 8174-8179.

Goodman, N. D., Baker, C.L., Bonawitz, E.B., Mansinghka V.K., Gopnik, A., Wellman, H., Schulz, L.E., Tenenbaum, J.B. (2006). *Intuitive Theories of Mind: A Rational Approach to False Belief*. Paper presented at the Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society, Vancouver, Cananda.

Gopnik, A. (1993). How we know our minds;: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences, 16*(1), 1-14.

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Dev, 59*(1), 26-37.

Grezes, J., Armony, J. L., Rowe, J., & Passingham, R. E. (2003). Activations related to "mirror" and "canonical" neurones in the human brain: an fMRI study. *Neuroimage, 18*(4), 928-937.

Happe, F. (2003). Theory of mind and the self. *Ann N Y Acad Sci, 1001*, 134-144.

Hickling, A. K., & Wellman, H. M. (2001). The emergence of children's causal explanations and theories: evidence from everyday conversation. *Dev Psychol, 37*(5), 668-683.

Johansson, P., Hall, L., Sikström, S., Olsson, A. (2005). Failure to Detect Mismatches Between Intention and Outcome in a Simple Decision Task. *Science, 210*(5745), 116.

Jones, E. E., Nisbett, R.E. (1972). the actor and the observer: Divergent perceptions of the cause of behavior. . In D. E. K. E. E. Jones & R. E. N. H. H. Kelley, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79-94). Morristown, NJ:: General Learning Press.

Leslie, A. (2000). 'Theory of Mind' as a mechanism of selective attention. In M. Gazzaniga (Ed.), *The New Cognitive Neurosciences* (2nd Edition ed., pp. 1235-1247). Cambridge, MA: MIT Press.

Malle, B. F. (2004). *How the mind explains behavior: Folk explanations, meaning, and social interaction.* . Cambridge, MA: MIT Press.

Molnar-Szakacs, I., Kaplan, J., Greenfield, P. M., & Iacoboni, M. (2006). Observing complex action sequences: The role of the fronto-parietal mirror neuron system. *Neuroimage, 33*(3), 923-935.

Moses, L. J. (2001). Executive accounts of theory-of-mind development. *Child Dev, 72*(3), 688-690.

Nichols, S., Stich, S. (2003). *Mindreading: An integrated Account of Pretence, Self-Awareness, and Understanding of Other Minds*: Oxford University Press

.

Nisbett, R. E., Wilson, T.D. (1977). The halo effect: Evidence for unconscious alteration of judgements. *Journal of Personality and Social Psychology, 35*(250-256).

O'Neill, D. K. (1992). Young children's understanding of the role that the sensory experiences play in knowledge acquisition. *Child Development, 63*, 474-491.

Olson, J., & J, S. (2005). The Influence of Behavior on Attitudes. In D. Albarracin, B. Johnson & Z. MP (Eds.), *The Handbook of Attitudes* (pp. 223-272): Routledge.

Perner, J. (1991). Understanding the representational mind. . *Cambridge, MA: MIT Press*.

Perner, J., Aichhorn, M., Kronbichler, M., Wolfgang, S., & Laddurner, G. (2006). Thinking of mental and other representations: The roles of left and right temporo-parietal junction. *Social Neuroscience, 1*(3/4), 235-2258.

Povinelli, D. J., & Vonk, J. (2003). Chimpanzee minds: suspiciously human? *Trends Cogn Sci, 7*(4), 157-160.

Pronin, E., & Ross, L. (2006). Temporal differences in trait self-ascription: when the self is seen as an other. *J Pers Soc Psychol, 90*(2), 197-209.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat Rev Neurosci, 2*(9), 661-670.

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: evidence for an accessibility model of emotional self-report. *Psychol Bull, 128*(6), 934-960.

Ruffman, T., Garnham, W., Import, A., & Connolly, D. (2001). Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. *J Exp Child Psychol, 80*(3), 201-224.

Sabbagh, M. A., Moses, L. J., & Shiverick, S. (2006). Executive functioning and preschoolers' understanding of false beliefs, false photographs, and false signs. *Child Dev, 77*(4), 1034-1049.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *Neuroimage, 19*(4), 1835-1842.

Saxe, R., & Powell, L. J. (2006). It's the thought that counts: specific brain regions for one component of theory of mind. *Psychol Sci, 17*(8), 692-699.

Saxe, R., Schulz, L., & Jiang, Y. (2006). Reading Minds versus Following Rules: Dissociating Theory of Mind and Executive Control in the Brain *Social Neuroscience, 1*(3-4), 284-298.

Saxe, R., & Wexler, A. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia, 43*(10), 1391-1399.

Sommer, M., Dohnel, K., Sodian, B., Meinhardt, J., Thoermer, C., & Hajak, G. (2007). Neural correlates of true and false belief reasoning. *Neuroimage, 35*(3), 1378-1384.

Trope, Y., Lieberman, N. . (2003). Temporal construal. *Psychological Review, 110*, 403-421.

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happe, F., Falkai, P., et al. (2001). Mind reading: neural mechanisms of theory of mind and self-perspective. *Neuroimage, 14*(1 Pt 1), 170-181.

Vogt, S., & Thomaschke, R. (2007). From visuo-motor interactions to imitation learning: behavioural and brain imaging studies. *J Sports Sci, 25*(5), 497-517.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Dev, 72*(3), 655-684.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103-128.

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proc Natl Acad Sci U S A, 104*(20), 8235-8240.

Zaitchik, D. (1990). When representations conflict with reality: the preschooler's problem with false beliefs and "false" photographs. *Cognition, 35*(1), 41-68.