



HAL
open science

Drawing the boundary between low-level and high-level mindreading

Frédérique de Vignemont

► **To cite this version:**

Frédérique de Vignemont. Drawing the boundary between low-level and high-level mindreading. *Philosophical Studies*, 2009, pp.1-10. ijn_00512317

HAL Id: ijn_00512317

https://hal.science/ijn_00512317

Submitted on 30 Aug 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Drawing the boundary between low-level and high-level mindreading

Frédérique de Vignemont

© Springer Science+Business Media B.V. 2009

Abstract The philosophical world is indebted to Alvin Goldman for a number of reasons, and among them, his defense of the relevance of cognitive science for philosophy of mind. In *Simulating minds*, Goldman discusses with great care and subtlety a wide variety of experimental results related to mindreading from cognitive neuroscience, cognitive psychology, social psychology and developmental psychology. No philosopher has done more to display the resourcefulness of mental simulation. I am sympathetic with much of the general direction of Goldman's theory. I agree with him that mindreading is not a single system based on a single mechanism. And I admire his attempt to bring together the cognitive neuroscientific discovery of mirror system phenomena and the philosophical account of pretense within a unique theoretical framework of mental simulation. To do so, Goldman distinguishes two types of mindreading, respectively, based on low-level and high-level simulation. Yet, I wonder in what sense they are really two distinct processes. Here, I will confine myself largely to spelling out a series of points that take issue with the distinction between low-level and high-level mindreading.

Keywords Mirroring · Pretense · Empathy · Automaticity · Reliability

1 Is there any distinction at all?

For many years, one of the most frequent arguments in favor of the simulation theory offered by Goldman was the pervasiveness of simulating processes in the mind, such as motor and visual imagery, mirror systems, empathy, pretense, etc. The argument was the following: If indeed simulation is everywhere, why not also in mindreading? At that time, one may have worried about a possible abusive use of

F. de Vignemont (✉)
Institut Jean-Nicod, EHESS – ENS – CNRS, 29 rue d'Ulm, 75005 Paris, France
e-mail: fvignemont@isc.cnrs.fr

the notion of simulation, given the diversity of processes Goldman was referring to. However, in *Simulating minds*, Goldman provides a systematic and precise account of the variety of simulation processes. On his view, simulation is ubiquitous, but it is not always based on the same type of mechanism. Conceptual cautiousness is made indispensable by the fact that nowadays scientists use the notion of simulation to refer to any kind of mechanism. Here, Goldman clarifies a number of important issues:

- (1) Simulation is not restricted to mindreading.
- (2) Simulation is not restricted to pretense.
- (3) No single mechanism underlies all instances of mental simulation.

To avoid confusion, he proposes the distinction between two levels of simulation, one of which is associated with higher and more central cognitive processes than the other. In Chap. 6, he presents evidence of ‘unmediated resonance’ (mirroring) for emotions, feelings and intentions. Briefly, it has been shown that performing an action and observing it activate the same internal representations of action (Gallese et al. 1996; Rizzolatti et al. 1995; Grezes and Decety 2001). Shared cortical networks have also been found for empathy. Brain imaging studies have shown overlapping brain activation patterns when subjects feel pain and when they observe others in pain (Singer et al. 2004), when they feel being touched and when they see someone else being touched (Keysers et al. 2004), when they inhale disgusting odorants and when they observe disgust-expressive faces (Wicker et al. 2003). All these phenomena have been counted as evidence of *low-level* mindreading. In chapter 7, Goldman elaborates the notion of ‘Enactment imagination’ (E-imagination), which is supposed to generate pretend states that resemble their intended counterparts. E-imagination is central to *high-level* mindreading: pretend beliefs and desires are fed into a decision-making process in order to understand someone else’s mental states. This corresponds to the classical version of Simulation Theory. However, the use of E-imagination is not restricted to mindreading tasks. It is also involved in visual imagery and motor imagery. To avoid confusion, I distinguish not only low-level and high-level mindreading (i.e. social simulation), but also low-level and high-level simulation (i.e. social or non-social). Both mirroring and E-imagination qualify as instances of simulation; they both can play a role in tasks of third-person mindreading; but they differ in some significant respects. My aim here is to clarify in what respects.

There is no doubt that the distinction between low-level and high-level mindreading is important and useful. Yet, surprisingly, Goldman spends little time articulating it. The definitions he gives remain broad, as shown by the following characterization of high-level mindreading:

‘High-level’ mindreading is mindreading with one or more of the following features: (a) it targets mental states of a relatively complex nature, such as propositional attitudes; (b) some components of the mindreading process are subject to voluntary control; and (c) the process has some degree of accessibility to consciousness (p. 147).

The definition of low-level mindreading is also sketchy:

[It] is ‘low-level’ mindreading because it is comparatively simple, primitive, automatic, and largely below the level of consciousness (p. 113).

It is simple because it targets simple mental states such as emotions, feelings and intentions. It is primitive because of its importance in evolution and also because its underlying mechanism is a mirroring process “that is cognitively fairly primitive” (p. 113). Nowhere in the book does Goldman formally specify more his distinction between low and high level mindreading. He spends time describing mirroring and pretense, but as he acknowledges himself, they are just realizations of low and high level mindreading among others. Instead of a rich proposal of architecture of the ‘simulating mind’, we find a small list of differences between the two levels, none of which is, as such, a necessary condition.

There may be several reasons why Goldman leaves his innovative distinction almost unexploited. The book is already rich and full of information coming from various disciplines. He may favor experimental results over theoretical distinctions. Alternatively, it might be that there is nothing more to the distinction than what he describes. It is merely a side-effect of the general cognitive architecture of the mind. Some simulating processes are more automatic or less conscious than others. Far from being a conceptual distinction between two types of simulation, it would be merely a question of degree.

Here, I want to assess in more detail the validity of the distinction by systematically comparing mirroring and pretense on several aspects: their neural basis, the mental states that they target, their availability to consciousness, their underlying mechanism, their reliability, and their fecundity. We will quickly see that the boundary is not easy to draw between low-level and high-level mindreading.

2 Failing to draw the boundary

If low and high level mindreading were two distinct processes, one would expect them to have different neural correlates. And indeed, Goldman argues in favor of the existence of dual brain systems for mindreading: the brain areas that constitute the mirror systems on the one hand, and the so-called ‘mindreading network’ (including the medial prefrontal cortex and the temporo-parietal junction) on the other hand. Low and high level mindreading have thus distinct neural correlates. Therefore, the distinction is empirically valid.

However, if neural correlates are to be taken into account, Goldman faces a difficulty. Among the examples of high-level simulation, he details the case of motor imagery. It is now well documented that motor imagery shares many properties with physical actions at the physiological level (muscle activity), at the kinematic level (similar physical constraints and laws) and at the neural level (shared patterns of brain activation) (Jeannerod 1994). Action execution, action observation, and also action imagination, all three, overlap at the neural level. In other words, here is a case of high-level simulation implemented in the cortical circuit of low-level mindreading. Goldman may reply that the neural correlates do not matter. However, if he did, then he could no longer use the evidence of dual brain systems for mindreading as an

argument in favor of his distinction. He may also reply that motor imagery is not an instance of high-level mindreading, but merely an instance of high-level non-social simulation. However, mirror activity during action observation is often interpreted as a result of the observer mentally imagining herself performing the same movement (Grezes and Decety 2001). If this is true, then motor imagery can also be used in third-person mindreading tasks. Consequently, so-called *low-level* action mirroring involves *high-level* motor imagery.

Alternatively, one may argue that motor imagery is just a case of low-level simulation. However, Goldman insists that it is high-level because it is conscious and under voluntary control. I will come back in the next section to the question of the automaticity of mirroring. One can already notice that consciousness does not seem like a good criterion. Goldman himself repeatedly claims that phenomenology is not a good guide and that a large portion of high-level mindreading is unconscious. On the other hand, low-level mindreading is potentially available to consciousness. It has been indeed found that some people *consciously feel* tactile sensations on their skin when they see someone being touched (Blakemore et al. 2005). Consequently, it is not because motor imagery can be conscious that it is necessarily high-level.

I took the example of motor imagery to illustrate how difficult it is to classify it into either low or high-level simulation. The main reason is that the distinction itself is not clear. One way to save it may be to apply Goldman's first criterion, namely, the type of mental states that is targeted. But again, this is not as straightforward as it might seem. It is true that one does not have access to other people's beliefs and desires through mirroring. However, there are other types of mental states that are common to low and high level simulation. Let us examine the example of empathy. There is more than one way to know what the other feels. One can rely either on mirroring or on E-imagination. In the same way that there may be two levels of simulation, there may be two levels of empathy, namely, mirror empathy and reconstructive empathy (de Vignemont, in press; Goldman, in press). Mirror empathy is induced by the observation of emotional cues that induces emotion mirroring. Reconstructive empathy is induced by the pretense of the emotional situation. I put myself in someone's shoes and imaginatively assume the other person's mental states, which are then operated upon by affective appraisal (Goldman 1993). Despite the fact that one is low-level and the other high-level, they both target the same type of mental states, namely, the emotions of others.

Hereafter, we will focus on empathy. All other things being equal, we will be able to provide a more rigorous comparison between low-level and high-level mindreading. Until now, we have failed to draw a clear boundary between the two levels of mindreading. Let us see if there are more convincing differences at the level of the underlying processes.

3 Automatic versus controlled

Goldman offers a definition of simulation that is wide enough to encompass various mechanisms. Process P is a mental simulation of P' if and only if (i) P and P' are both mental processes, (ii) P resembles P' in some significant respects and (iii) in

duplicating P' , P fulfils one of its functions. I will not discuss here whether it is justified to claim that mirroring plays a role in mindreading, I leave that to others (Jacob, in press). Rather, I would like to discuss in what sense low-level mindreading involves a different mechanism than high-level mindreading. Goldman highlights two main related differences. First, mirroring is automatic, whereas E-imagination can be under voluntary control. The second aspect is a direct consequence of the first. E-imagination can be guided by knowledge, in contrast with mirroring. I will now challenge both claims based on the analysis of empathy.

In cognitive psychology, automatic processing is defined as follows (see Neumann 1984 for review):

- (a) Mode of control: passive consequence of stimulation, on which one has no intentional control
- (b) Mode of operation: immune to interference
- (c) Mode of representation: not necessarily conscious

A good example of automatic processing is the Stroop effect. One cannot help but reading words although one is asked to judge only the color of the ink. Similarly, if mirror empathy were purely automatic, it should not be possible to inhibit it or to interfere with it. However, there is evidence of the contrary. Several studies have now shown that mirror empathy is not systematically elicited each time one perceives a painful event. Imagine that you are playing a money game with people that you just met tonight. One of them is unfair and treats you badly. Just after the game, he hurts himself. Do you feel empathy for him, although you do not like him? Or do you rather feel that he just got what he deserved? This last scenario is actually the design of a brain imaging experiment (Singer et al. 2006). The women participants still showed the activation of the pain network. In contrast, the men participants showed none. Consequently, the relationship that one has with the target modulates the empathic response. This shows at least that mirror empathy is not immune to interference. Another study illustrates that it can be intentionally regulated. Cheng et al. (2007) found no activation of the pain network in medical practitioners while they were observing painful events. It would indeed not be adaptive if they could not regulate their empathy as their distress and anxiety would then prevent their ability to heal. One has thus an intentional control over one's own emotions and empathy. Along the same line, it was also shown that mirror empathy can be guided by semantic knowledge. Participants in a brain imaging study had less empathetic response in pain-related areas when they knew that the pain inflicted to the other was useful (to cure her), than when they knew it did not help the person (Lamm et al. 2007). Thus, mirror empathy can be regulated by top-down factors:

Rather than suggesting that humans respond on the basis of automatically activated stimulus-response linkages, the present findings support the notion that humans regulate their emotions by relying on higher cognitive processes involving knowledge in working memory, long-term memory, and metacognition. Cheng and et al. (2007, p. 1712).

There is evidence of regulation, but when does it occur? Goldman may indeed reply that the fact that mirror empathy has to be inhibited proves that it is automatic.

However, there is no evidence that this regulatory process is inhibitory in nature. It is important here to distinguish between inhibitory and constructive processes. Inhibition is consistent with automatic low-level mirror empathy. But the view that mirror empathy is the output of constructive process is not. At the empirical level, we do not yet know at which stage of empathic processing the top-down modulation occurs. There are two possible routes (de Vignemont and Singer 2006). According to the late appraisal model, the empathic response is directly and automatically activated by the perception of an emotional cue, which can be modulated or inhibited at a later stage. According to the early appraisal model, the emotional cue is evaluated in the context of external and internal information. Whether an empathic response is elicited depends on the outcome of the contextual appraisal process. Current neuroscientific studies on empathy cannot yet distinguish between these two proposed routes. What we know is that mirror empathy is sensitive to a wide range of factors, some being very high-level. There is no reason—conceptual or empirical—to believe that the top-down factors can only be inhibitory. They may as well be part of a constructive mechanism that integrates information about the context, about the person, about your relationship with her, etc. before eliciting the empathic response. This is then incompatible with the view of automatic low-level “unmediated resonance” described by Goldman.

Alternatively, Goldman may agree that mirror empathy can be voluntarily controlled, and yet argue that it cannot be voluntarily triggered, in contrast with reconstructive empathy. At first sight, this argument sounds convincing. I cannot decide to activate my mirror systems. Yet, one may wonder about the mode of producing reconstructive empathy. It cannot be denied that I can decide to engage in perspective-taking, but whether I can decide to feel empathy is quite another matter. What I can decide is to pretend to be in the same emotional situation as the target and to let run my affective appraisal. What I cannot decide is to share the same feeling at the end of my simulating process. The switch from cognitive perspective-taking (i.e. understanding the other’s emotion) to emotional perspective-taking or reconstructive empathy (i.e. sharing the other’s emotion) is not under voluntary control. In this sense, mirror empathy and reconstructive empathy do not differ. When I feel sad with someone, it is not because I want to be sad with her, it just happens to me. And as much as I may wish to share someone’s happiness, it does not necessarily work. There is more to reconstructive empathy than voluntary perspective-taking.

To sum up, on the one hand, mirror empathy can be regulated and guided by knowledge like reconstructive empathy. On the other hand, reconstructive empathy cannot be voluntarily triggered like mirror empathy. We need to look somewhere else to dissociate them.

4 Fecundity and reliability

The distinction between two mechanisms brings us to the question of their respective fecundity and reliability. The fecundity of mindreading is measured by the richness of information that one gets about the other. The reliability of

mindreading is measured by the accuracy of this information. According to Goldman, low-level mindreading is less fecund and more reliable than high-level mindreading. Unfortunately, I will show that even low-level mindreading is not immune to error.

When trying to understand other people's emotions, there are four different parameters to consider: the valence of the emotions (e.g., negative), the type of emotion (e.g., fear), the propositional content (e.g., that I miss my train) and the functional role (e.g., start running). The problem with mirror empathy is that it can be based on very poor input. As a consequence, the output can also be very poor. Seeing a stranger looking worried in the street may make me share his anxiety, but there is no way that it can make me understand why the person looks worried. There is no propositional content associated, just the emotion type. In contrast, the input of reconstructive empathy can be rich, and thus the output too. For instance, upon learning that a friend of mine is pregnant, I pretend that I am pregnant and that I want a child, which leads me to feel happy. My emotional state is about the pregnancy. It has the same focus as my friend's emotion. The causes and reasons of the emotional state are indeed the input of the simulative process. They are constitutive of reconstructive empathy. Furthermore, the output of the simulative process is not exclusively emotional. It also includes the beliefs, desires and intentions that are triggered by the emotional situation. Reconstructive empathy is of particular interest for mindreading because it provides access not only to the fine-grained emotional states, but also to the other mental states that are associated to them.

However, the price to pay for this higher fecundity is a decrease of reliability. Mindreading reliability depends on the immediacy of the simulating process. The fewer the intermediary steps, the less likely it is that things could go wrong. According to Goldman, low-level simulation is based on 'unmediated resonance'. The causal link between the target state and the duplicated state is direct, and thus there is almost no chance that the duplicated state does not match the target state. In contrast, high-level simulation involves selecting information to feed the simulating process. One can be mistaken in the selection stage, forgetting to include relevant information or to erase one's own irrelevant mental states. This latter error is what Goldman calls an egocentric bias. He describes several cases, like the curse of knowledge. Even adults have difficulties in inhibiting what they themselves know to predict someone else's behavior. For instance, well-informed subjects had to predict what other less-informed people would forecast for corporate earnings. Despite the fact that they knew that they were less informed, still they predicted that they would say what they themselves forecasted (Camerer et al. 1989). Such egocentric biases happen not only for knowledge, but also for values and feelings.

Goldman's view of low-level and high-level mindreading is accurate so long as mirroring remains unmediated. However, we have seen in the previous section that mirroring is influenced by the context, a context that can enrich mirror processing, as it has been shown for action (Iacoboni et al. 2005). Participants observed someone's hand grasping a cup in two different contexts: before tea or after tea. The intention was thus different: either to drink or to clean. Iacoboni et al. (2005) found different activations of the mirror system depending on the context. By paying attention to the context, participants acquired a deeper understanding of the action.

Similarly, one could expect that emotion mirroring can lead to a richer knowledge of the emotion when contextualized. For example, on the only basis of the color of your cheeks I cannot know why you are blushing, because of happiness or of shame. But if I heard just before someone complimenting you, I know that you feel happy because of the compliments. Low-level mindreading can thus be more fecund if it is based not only on bodily cues, but also on contextual cues. And this is what happens most of the time.

Furthermore, it would be idealistic to believe that low-level mindreading is immune to egocentric biases. It was Diderot who first noticed that a man would never be able to empathize with his wife's pain when she gives birth. If an emotion is not part of your emotional repertoire, then you cannot mirror it. This has been recently shown more experimentally in a study on empathy for pain. Some individuals (often with autism) are alexithymic. They do not feel emotions. No activation of the pain network was found when they saw someone being hurt, contrary to non-alexithymic participants (Silani et al. 2007). Who you are and what you can experience thus affect the way you perceive others. Such people may fail to ascribe pain sensations to others. Low-level mindreading in their case is not immune to egocentric bias.

More generally, we have seen that 'normal' mirroring is modulated by many factors, each factor being a possible source of errors. Alternatively, Goldman may argue that the type of errors that we make is just of omission. We fail to ascribe the others what they actually feel, but we do not ascribe them an emotion that they do not feel. However, we can easily imagine that a masochist would ascribe a sensation of pleasure to the person that he sees in pain or that a person with chronic back pain would ascribe a sensation of pain when he sees someone lifting a heavy object. A last possible way to save the distinction for Goldman would be to point out that in high-level mindreading we can quarantine our own mental states to avoid them contaminating our simulation, whereas in low-level mindreading we have no control over our egocentric biases. This is partly true. Yet, there are some cases where we can actually avoid making errors by neglecting the output (or the lack of output) of the mirroring system. Let us go back to the medical practitioners. They have no pain mirroring, yet they believe that their patients are in pain. Interestingly, Cheng et al.'s brain imaging study showed an activation of the high-level mindreading network. When there is a conflict between low-level and high-level mindreading, the latter can take over the former.

In summary, Goldman offers a distinction between two levels of mindreading that should help to clarify the field. However, one remains frustrated by the elusiveness of this distinction. He lists a small set of conditions that appear to be sufficient, although not necessary. However, after examination, it turns out that they are neither necessary nor sufficient. More particularly, low-level mindreading cannot be reduced to a purely bottom-up automatic process of low fecundity, but reliable. We have seen that it can be contextualized, and thus that it can be rich in information, but at the cost of egocentric biases.

Should we then give up on the distinction between the two levels of mindreading? More than a dichotomy between two mechanisms with distinct properties, we have seen that low and high-level mindreading constitute a continuum. Mindreading can

be more or less automatic, more or less conscious, more or less fecund, and more or less reliable. One cannot even ground the distinction on the richness of the input and the complexity of the process. Of course, sometimes it is enough to see a smile or to hear the tone of the voice to activate the feeling of happiness or anger, whereas reconstructive empathy involves extracting, selecting, combining and evaluating the relevant information to generate emotional states similar to the person's states. However, if the early appraisal model of mirror empathy were to be shown true, then even mirror empathy would result from a complex constructive process.

To sum up, the low-level and high-level distinction fails to engage with distinctions between types of mindreading at all.

References

- Blakemore, S.-J., Bristow, D., Bird, G., Frith, C., & Ward, J. (2005). Somatosensory activations during the observation of touch and a case of vision-touch synaesthesia. *Brain*, *128*, 1571–1583.
- Camerer, C. F., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy*, *97*, 1232–1254.
- Cheng, Y., Lin, C. P., Liu, H. L., Hsu, Y. Y., Lim, K. E., Hung, D., et al. (2007). Expertise modulates the perception of pain in others. *Current Biology*, *17*(19), 1708–1713.
- de Vignemont, F. (in press). Knowing other people's mental states as if they were one's own. In S. Gallagher & D. Schmicking (Eds.), *Handbook of Phenomenology and Cognitive Science*, Berlin: Springer.
- de Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why? *Trends in Cognitive Science*, *10*(10), 435–441.
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*, 593–609.
- Goldman, A. I. (1993). Consciousness, folk psychology and cognitive science. *Consciousness and Cognition*, *2*, 364–382.
- Goldman, A. I. (in press). Two routes to empathy: insights from cognitive neuroscience. In A. Coplan & P. Goldie (Eds.), *Empathy: philosophical and psychological perspectives*. Oxford: Oxford University Press.
- Grezes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Human Brain Mapping*, *12*(1), 1–19.
- Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology*, *3*(3), 529–535.
- Jacob, P. (in press). The tuning-fork model of human social cognition: A critique. *Consciousness and Cognition*.
- Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain sciences*, *17*(2), 187–245.
- Keysers, C., Wicker, B., Gazzola, V., Anton, J.-L., Fogassi, L., & Gallese, V. (2004). A touching sight: SII/PV activation during the observation of touch. *Neuron*, *42*, 335–346.
- Lamm, C., Batson, C. D., & Decety, J. (2007). The neural substrate of human empathy: Effects of perspective-taking and cognitive appraisal. *Journal of cognitive neuroscience*, *19*(1), 42–58.
- Neumann, O. (1984). Automatic processing: A Review of recent findings and a plea for an old theory. In W. Prinz & A. F. Sanders (Eds.), *Cognition and automatic processing* (pp. 255–293). Berlin: Springer-Verlag.
- Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1995). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*, 131–141.
- Silani, G., Bird, G., Brindley, R., Singer, T., Frith, C., & Frith, U. (2007). Levels of emotional awareness and autism: An fMRI Study. *Social Neuroscience*, *3*(2), 97–112.
- Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R., & Frith, C. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, *303*, 1157–1162.

- Singer, T., Seymour, B., O'Doherty, J., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, *439*, 466–469.
- Wicker, B., Keysers, C., Plailly, J., Royet, J.-P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron*, *40*, 655–664.