Neuron 50, 531–534, May 18, 2006 ©2006 Elsevier Inc. DOI 10.1016/j.neuron.2006.05.001

The Neural Basis of Mentalizing

Minireview

Chris D. Frith[1,*] and Uta Frith[2]
[1] Wellcome Department of Imaging Neuroscience
[2] Institute of Cognitive Neuroscience
University College London
London WC1N 3BG
United Kingdom

Mentalizing refers to our ability to read the mental states of other agents and engages many neural processes. The brain's mirror system allows us to share the emotions of others. Through perspective taking, we can infer what a person currently believes about the world given their point of view. Finally, the human brain has the unique ability to represent the mental states of the self and the other and the relationship between these mental states, making possible the communication of ideas.

What Is Mentalizing?

The term mentalizing was coined to refer to the process by which we make inferences about mental states. Much of the time these inferences are made automatically, without any thought or deliberation. It is important for us to be able to read the minds of others because it is their mental states that determine their actions. This assumption that behavior is caused by mental states has been called "the intentional stance" (Dennett, 1987) or "having a theory of mind" (Premack and Woodruff, 1978). There are many different types of mental states that can affect the way we interact with others. There are long-term dispositions: one person may be trustworthy and reliable while another is hopelessly volatile. There are short-term emotional states like happiness and anger. There are desires like thirst and their associated goal-directed intentions (e.g., fetching a bottle of wine from the fridge). There are the beliefs that we have about the world. These beliefs determine our behavior even when they are false (someone has secretly removed the wine from the fridge) or not shared by others (English wine can be very pleasant). Finally, we shall consider the role of communicative intent.

How Do We Mentalize?

Many cues in different modalities can trigger the process of mentalizing as long as they originate from an agent. Agency can be perceived in other animals and even in moving objects (Heider and Simmel, 1944), but the agents we are most interested in are our conspecifics. Their faces, in particular, are an important source of information about their inner states. For example, there is agreement about what a trustworthy person looks like even though this is an example of a prejudice with little basis in reality. Emotions, on the other hand, can be validly read from facial expressions, from voices, and from whole-body movements (Adolphs, 2002). Desires, goals, and intentions can be read from eye gaze direction and body movements (Langton et al., 2000). Beliefs are computed by recognizing that knowledge depends on experience, so that someone may not know what we know because they have not seen what we have seen (Wimmer et al., 1988). Note that this example involves perspective taking, a vital aspect of successful mentalizing. Communicative intentions are perceived when someone calls our name or makes eye contact (Sperber and Wilson, 1995).

The Brain's Mirror System

Simulation theory proposes that we can understand the mental states of others on the basis of our own mental states (Gallese and Goldman, 1998). Commonalities between the self and the other have been observed in a number of brain imaging studies. There is a "mirror" system in the brain such that the same areas are activated when we observe another person experiencing an emotion as when we experience the same emotion ourselves, as if by contagion. Through such a mechanism we can experience the emotional states of another person. The brain's mirror system is engaged by actions as well as emotions (Rizzolatti and Craighero, 2004) to the extent that we automatically imitate the movements of others (Chartrand and Bargh, 1999) even when this interferes with our own actions (Kilner et al., 2003).

However, experiencing the same emotion as another is not sufficient to infer the cause of that emotion and hence is only a first step for mentalizing. Likewise, covertly performing the same action as another is not sufficient to infer the goals and intentions behind that action. Furthermore, as Mitchell et al. (2006) (this issue of *Neuron*) point out, while the mirror system is ideally suited for tracking the continually changing states of emotion and intention of the other, it can tell us nothing about the stable attitudes and predilections of the other, which we also perceive as important determinants of behavior, and hence of our "theory of mind."

There are two problems to be solved. First, how do we infer the causes of the emotions and actions of the other? Second, on the basis of what we know about the other, how do we predict what he or she will do next? Some answers to these questions come from studies of the neural correlates of mentalizing.

The Neural Correlates of Mentalizing

Over the last ten years, many imaging studies have been conducted on the neural basis of mentalizing. A wide range of different paradigms have been used in which, for example, participants read stories, watch moving shapes, or play interactive games. The common feature in all these paradigms is that the participants have to think about the mental states of another person. The results have been remarkably consistent, implicating a set of regions that include pSTS/TPJ, the temporal poles and the medial prefrontal cortex (Frith and Frith, 2003). The precise roles of these different regions are now beginning to emerge.

Perspective Taking

Through recognizing their facial expression, we might know that someone is afraid. But what are they afraid

*Correspondence: cfirth@fil.ion.ucl.ac.uk

of? A good way to discover the cause of their fear is to check what they are looking at. The region of the brain at the posterior end of the superior temporal sulcus (pSTS) and the adjacent temporo-parietal junction (TPJ) is a prime candidate for this process. First, this region is involved in eye-movement observation and provides information about where someone is looking (Pelphrey et al., 2004). Second, this region is involved in representing the world from different visual perspectives (Aichhorn et al., 2005), probably as a consequence of its role in representing the position of the body in space (Blanke et al., 2005). Knowing where a person is looking and what they can see, given their vantage point, enables us to know what they are looking at and thus identify the cause of their fear. This ability to see the world from another's perspective enables us to realize that other people can have different knowledge from us and may have false beliefs about the world. "He thinks he is safe because he can't see the bear coming up behind him." There is evidence that the temporo-parietal junction has a necessary and more general role in the performance of tasks that depend upon understanding that a person has a false belief about the world (Apperly et al., 2004).

Knowledge of the World

Through experience, we build up a rich store of knowledge about the world that is important for our ability to mentalize. We learn about specific people: what they look like, where they live, whether they are trustworthy, and so on. We also learn about the moment-to-moment changes of the situations in which people find themselves and how their feelings, knowledge, and dispositions affect their behavior. We learn about the kinds of behavior that are appropriate in these different situations. Damage to the temporal poles can impair the ability to use this knowledge (Funnell, 2001). This observation is consistent with the suggestion that the temporal poles are convergence zones where simpler features from different modalities are brought together to define, by their conjunction, unique individuals and situations (Damasio et al., 2004). Through this convergence of information, our understanding of an object can be modified by the context in which it appears (Ganis and Kutas, 2003). These processes instantiated in the temporal poles are important for mentalizing, not only in allowing us to apply general knowledge, but also to use moment-to-moment knowledge about a particular person in a particular place. They specify what the person is likely to be thinking and feeling and the kinds of thoughts and feelings most likely to occur in a particular context, e.g., pride or embarrassment caused by otherwise similar actions. This is even before the other person shows the feelings and performs the actions, which we can pick up via the mirror system.

Anticipating the Future

The final region we shall consider in this review is the medial prefrontal cortex and the adjacent paracingulate cortex. This rather large region has been consistently activated when participants think about mental states, whether these are long-term dispositions and attributes or short-term intentions and beliefs. The precise role of this region remains controversial. Patients with damage to frontal cortex are frequently impaired in the performance of "theory of mind" tasks (Stuss et al., 2001). However, it has not proven easy to locate the critical damage more precisely, and there is one case reported where damage restricted to mPFC did not lead to impairment in the performance of ToM tasks (Bird et al., 2004). Perhaps the processes instantiated in this area are typically elicited when we think about mental states but are not actually necessary for the performance of standard ToM tasks. TMS techniques might provide an appropriate means to explore this question.

Understanding People Like Us

In general, prefrontal cortex is concerned with planning for the future and representing anticipated states of the world (Ingvar, 1985; Shallice, 1988). Thus, in the specific case of mentalizing, mPFC may be concerned with anticipating what a person is going to think and feel and thereby predict what they are going to do. How are such predictions made? It has long been noted that activity is seen in mPFC, not only when thinking about the mental states of others, but also when thinking about mental states of the self (Frith and Frith, 1999). This observation is consistent with a simulationist account that suggests that we can predict what someone else will think and feel by considering what we would think and feel if we were in their situation. The problem for this approach is that it only works well if we are very similar to the person whose behavior we are trying to predict. Mitchell et al. (2006) report an elegant experiment that directly investigates the effect of similarity. Participants were told about two target individuals who were described as having liberal or conservative views. They were then asked to predict the feelings and attitudes of these two targets in various situations (e.g., "would he enjoy having a roommate from a different country"). Subsequently the political attitudes of the participants were also assessed. The results show a different pattern when thinking about a similar or a dissimilar other. Thinking about similar others was associated with activity in ventral mPFC (18, 57, 9—in the region labeled anterior rostral MFC in Amodio and Frith [2006]), while thinking about a dissimilar other was associated with activity in a more dorsal region of mPFC (−9, 45, 42—posterior rostral MFC).

This is strong evidence for segregation of function within the area of medial prefrontal cortex associated with mentalizing. There are hints at segregation in other recent studies (see Figure 1 for an illustration of the segregation in medial frontal cortex suggested by these studies). Hynes et al. (2006) asked participants to make inferences about what other people were thinking (cognitive perspective taking) or what they were feeling (emotional perspective taking). Thinking about people's feelings was associated with activity in medial orbital cortex (18, 63, −7), while perspective taking in general was associated with activity in more dorsal regions (2, 59, 15; −4, 60, 30). Walter et al. (2004) asked participants to make inferences about private intentions (changing a broken light bulb in order to read a book) in contrast to communicative intentions (showing someone a map in order to ask the way). Thinking about communicative intentions activated a more ventral region (−3, 54, 15) than thinking about private intentions. Grèzes et al. (2004) asked participants to infer whether the movements associated with the lifting of a box were intended to be deceptive since the actor might be pretending that the box was heavier than it really was. Movements thought to be deceptive activated anterior

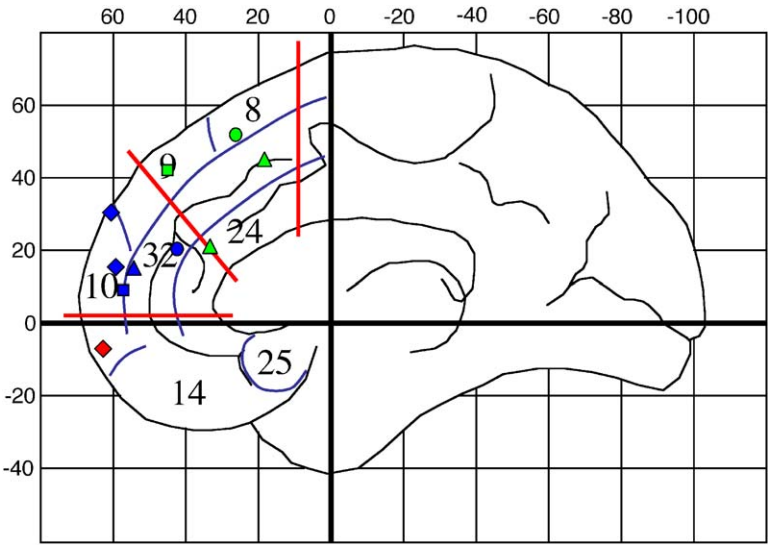

Figure 1. Segregation of Activity Associated with Mentalizing Tasks in Medial Frontal Cortex

This is a view of the medial surface of the brain in Talairach space with numbers on the cortex indicating the approximate location of the relevant Brodmann areas. The red dividing lines are based upon metaanalyses of functional imaging studies reviewed in Amodio and Frith (2006). The colored symbols indicate the location of peak activations from four recent studies of mentalizing: Mitchell et al. (2006)—similar others, blue square; dissimilar others, green square; Hynes et al. (2006)—feelings, red diamond; point of view, blue diamond; Walter et al. (2004)—communicative, blue triangle; private, green triangle; Grèzes et al. (2004)—communicative, blue circle; private, green circle.

rostral MFC (−8, 42, 20). In another experiment the participants observed movements, which sometimes included unexpected adjustments, because the box being picked up was lighter than the actor expected. Observing these unexpected adjustments was associated with activity in posterior rostral MFC (2, 26, 52).

These observations suggest subdivisions of MFC relating to different aspects of mentalizing. Amodio and Frith (2006) suggest that the most ventral region (medial orbital cortex, $z < 2$) is concerned with the monitoring of emotions in self and other, while the most dorsal region (posterior rostral MFC, including the "cognitive" section of ACC) is concerned with the monitoring of actions, again in both self and other. In between these two regions is the region (anterior rostral MFC), which is activated when thinking about people similar to ourselves and is also activated when we perceive that another person intends to communicate with us.

Why should this region be more active when predicting the behavior of people similar to ourselves? One possibility is that this part of the cortex can combine information about emotions and actions, since it is adjacent to the two regions with these specialities. When deciding what to do we are not totally "rational" in our choice of action: our choice is colored by emotions such as anticipated regret or desire for fairness. Inferences about the most rational action in the circumstances apply to everyone whether we are similar to them or not. Such inferences can be made via the action-monitoring system. But we can only take account of the role of emotions in the choice of action in people similar to ourselves who are likely to feel the same emotions.

Communicative Intent

However, while this may be part of the story, it does not explain why this region is specifically activated by tasks involving communicative intent. Communicative intent requires a special form of metacognition (Amodio and Frith, 2006). The point of communication is to alter the mental state of the person we are communicating with, for instance, by imparting new knowledge. To be receptive to this new knowledge, the listener has to be able to perceive the speaker's intention to communicate (Sperber and Wilson, 1995). For successful communication, it is not sufficient for the speaker to represent the mental state of the listener. Both speaker and listener have to recognize that the listener is representing the speaker's mental state and that the signals coming from the speaker, whether words or simply movements, are emitted with the intention of altering the listener's mental state. Perhaps the simplest case for which this kind of understanding is needed is shared attention, when one person gazes at or points to an object so that the object can become the focus of attention for another person also. We would characterize this as a prime example of communicative intent. Saxe (2006) suggests that dorsal MFC may have a specific role in such triadic social interaction, which according to Tomasello et al. (2005) is a uniquely human ability. It is therefore interesting that the region of medial prefrontal cortex implicated in the process has enlarged dramatically in the recent course of evolution (Semendeferi et al., 2001). Returning to the result of Mitchell et al. (2006), when we think about people similar to ourselves, do we automatically activate regions concerned with creating our shared view of the world, poised to communicate with them?

We are still at the very early stages of understanding how the brain permits us to read the minds of others. Our account of the roles of the various regions implicated in this process is of necessity speculative. However, these speculations generate a number of clear predictions, and the methodology is available for well-designed experiments like that of Mitchell et al. (2006) to test these predictions. Reading the minds of our colleagues in the burgeoning field of social cognitive neuroscience, we predict that over the next few years our understanding of the brain's mentalizing system will increase dramatically.

Selected Reading

Adolphs, R. (2002). Curr. Opin. Neurobiol. *12*, 169–177.

Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., and Ladurner, G. (2005). Neuroimage *30*, 1059–1068. Published online December 6, 2005. 10.1016/j.neuronimage.2005.10.026.

Amodio, D.M., and Frith, C.D. (2006). Nat. Rev. Neurosci. *7*, 268–277.

Apperly, I.A., Samson, D., Chiavarino, C., and Humphreys, G.W. (2004). J. Cogn. Neurosci. *16*, 1773–1784.

Bird, C.M., Castelli, F., Malik, O., Frith, U., and Husain, M. (2004). Brain *127*, 914–928.

Blanke, O., Mohr, C., Michel, C.M., Pascual-Leone, A., Brugger, P., Seeck, M., Landis, T., and Thut, G. (2005). J. Neurosci. *25*, 550–557.

Chartrand, T.L., and Bargh, J.A. (1999). J. Pers. Soc. Psychol. *76*, 893–910.

Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., and Damasio, A. (2004). Cognition *92*, 179–229.

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

Frith, C.D., and Frith, U. (1999). Science *286*, 1692–1695.

Frith, U., and Frith, C.D. (2003). Philos. Trans. R. Soc. Lond. B Biol. Sci. *358*, 459–473.

Funnell, E. (2001). Cogn. Neuropsychol. *18*, 323–341.

Gallese, V., and Goldman, A. (1998). Trends Cogn. Sci. *2*, 493–501.

Ganis, G., and Kutas, M. (2003). Brain Res. Cogn. Brain Res. *16*, 123–144.

Grèzes, J., Frith, C., and Passingham, R.E. (2004). J. Neurosci. *24*, 5500–5505.

Heider, F., and Simmel, M. (1944). Am. J. Psychol. *57*, 243–249.

Hynes, C.A., Baird, A.A., and Grafton, S.T. (2006). Neuropsychologia *44*, 374–383.

Ingvar, D.H. (1985). Hum. Neurobiol. *4*, 127–136.

Kilner, J.M., Paulignan, Y., and Blakemore, S.J. (2003). Curr. Biol. *13*, 522–525.

Langton, S.R., Watt, R.J., and Bruce, I.I. (2000). Trends Cogn. Sci. *4*, 50–59.

Mitchell, J.P., Macrae, C.N., and Banaji, M.R. (2006). Neuron *50*, this issue, 655–663.

Pelphrey, K.A., Viola, R.J., and McCarthy, G. (2004). Psychol. Sci. *15*, 598–603.

Premack, D., and Woodruff, G. (1978). Behav. Brain Sci. *1*, 515–526.

Rizzolatti, G., and Craighero, L. (2004). Annu. Rev. Neurosci. *27*, 169–192.

Saxe, R. (2006). Curr. Opin. Neurobiol. *16*, 235–239.

Semendeferi, K., Armstrong, E., Schleicher, A., Zilles, K., and Van Hoesen, G.W. (2001). Am. J. Phys. Anthropol. *114*, 224–241.

Shallice, T. (1988). From Neuropsychology to Mental Structure (Cambridge: Cambridge University Press).

Sperber, D., and Wilson, D. (1995). Relevance: Communication and Cognition, Second Edition (Oxford: Blackwell).

Stuss, D.T., Gallup, G.G., Jr., and Alexander, M.P. (2001). Brain *124*, 279–286.

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Behav. Brain Sci. *28*, 675–691.

Walter, H., Adenzato, M., Ciaramidaro, A., Enrici, I., Pia, L., and Bara, B.G. (2004). J. Cogn. Neurosci. *16*, 1854–1863.

Wimmer, H., Hogrefe, G.J., and Perner, J. (1988). Child Dev. *59*, 386–396.