Available online at www.sciencedirect.com
SCIENCE DIRECT®
COGNITION
ELSEVIER
Cognition 94 (2005) 193–213
www.elsevier.com/locate/COGNIT

Simulationist models of face-based emotion recognition

Alvin I. Goldman[a,b,*], Chandra Sekhar Sripada[b]

[a]*Center for Cognitive Science, Rutgers University, P.O. Box 1179, Piscataway, NJ 08855-1179, USA*
[b]*Department of Philosophy, Rutgers University, 26 Nichol Avenue, New Brunswick, NJ 08901-2882, USA*

Received 31 March 2003; revised 10 December 2003; accepted 19 January 2004

Abstract

Recent studies of emotion mindreading reveal that for three emotions, fear, disgust, and anger, deficits in face-based recognition are paired with deficits in the production of the same emotion. What type of mindreading process would explain this pattern of paired deficits? The simulation approach and the theorizing approach are examined to determine their compatibility with the existing evidence. We conclude that the simulation approach offers the best explanation of the data. What computational steps might be used, however, in simulation-style emotion detection? Four alternative models are explored: a generate-and-test model, a reverse simulation model, a variant of the reverse simulation model that employs an "as if" loop, and an unmediated resonance model.

Keywords: Emotion; Fear; Disgust; Anger; Theory of mind; Simulation theory; Theory theory; Facial feedback; Mirror neurons

1. Introduction

Mindreading is the capacity to identify the mental states of others, for example, their beliefs, desires, intentions, goals, experiences, sensations and also emotion states. One approach to mindreading holds that mental-state attributors deploy a naïve psychological theory to infer mental states in others from their behavior, the environment, and/or their other mental states. According to different versions of this "theory-theory" (TT), the naïve psychological theory is either a component of an innate, dedicated module or is acquired

* Corresponding author. Address: Department of Philosophy, Rutgers University, 26 Nichol Avenue, New Brunswick, NJ 08901-2882, USA.
E-mail address: goldman@philosophy.rutgers.edu (A.I. Goldman).

0022-2860/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2004.01.005

by domain-general learning. A second approach holds that people typically execute mindreading by a different sort of process, a simulation process. Roughly, according to simulation theory (ST), an attributor arrives at a mental attribution by simulating, replicating, or reproducing in his own mind the same state as the target's, or by attempting to do so. For example, the attributor would pretend to be in initial states thought to correspond to those of the target, feed these states into parts of his own cognitive equipment (e.g. a decision-making mechanism), which would operate on them to produce an output state that is imputed to the target.

Mindreading has been studied in many disciplines, and both TT and ST have had proponents in each of them. In developmental psychology TT has been endorsed by Gopnik & Meltzoff (1997), Gopnik & Wellman (1992, 1994), Leslie (1994), Perner (1991), Premack & Woodruff (1978), and Wellman (1990), whereas ST has been defended by Harris (1991, 1992). In philosophy ST has been endorsed by Currie & Ravenscroft (2002), Goldman (1989, 1992a, 1992b, 2000), Gordon (1986, 1992, 1996), and Heal (1986, 1996, 1998), whereas TT has been defended as an explicit approach to the execution of mindreading by Fodor (1992), Nichols, Stich, Leslie & Klein (1996), and Stich & Nichols (1992), or as a theory of how the folk conceptualize mental states by Armstrong (1968), Lewis (1972), Sellars (1956), and Shoemaker (1975). Studies of mentalizing in neuroscience (e.g. Fletcher et al., 1995; Frith & Frith, 1999; McCabe, Houser, Ryan, Smith, & Trouard, 2001) typically ignore the TT–ST controversy but work with 'theory of mind' terminology, which is suggestive of TT, and cite the TT-leaning literature. On the other hand, much recent neuroscientific work is quite receptive to simulationist ideas (Blakemore & Decety, 2001; Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003; Chaminade, Meary, Orliaguet, & Decety, 2001; Gallese, 2001, 2003; Gallese & Goldman, 1998; Iacoboni et al., 1999; Jeannerod, 2001), although the majority of this research is addressed less to mindreading per se than to related topics such as simulation of action, imitation, or empathy. In recent years a number of researchers have moved away from pure forms of TT or ST in the direction of some sort of TT/ST hybrid (Adolphs, 2002; Goldman, in preparation; Nichols & Stich, 2003; Perner, 1996), though the exact nature of the hybrid is rather fluid. In light of this continuing controversy, any research that provides substantial evidence for either ST or TT, even in a single subdomain of mindreading, deserves close attention.

In this paper we review a body of neuropsychological research that, we shall argue, supports ST for a certain circumscribed mindreading task. This is the task of attributing emotion states to others based on their facial expressions. This task is different from those usually studied in the mindreading literature, in part because the attributed mental states differ from the usual ones. The vast majority of the literature is devoted to propositional attitudes such as desires and beliefs, almost entirely ignoring emotion states like fear, anger, disgust, or happiness. There is no good reason to exclude these mental states, which are routinely attributed to others in daily life. So it is time to extend research and theory into this subdomain of the mental. At the same time, it cannot be assumed that the style of mindreading in this subdomain is the same as the style that characterizes other subdomains. So we make no attempt to generalize from the type of mental state ascriptions studied here to mindreading *tout court*.

There are at least two reasons why the properties of face-based emotion recognition (FaBER) might not be shared by methods of mindreading in other subdomains. First, the recognition or classification of propositional attitude contents may introduce a level of complexity that goes beyond the task of classifying emotion types. Second, the reading of emotions, especially basic emotions, may have unique survival value, so it is conceivable that specialized programs have evolved for the recognition of emotions, and these specialized programs may not operate in other mindreading tasks. Because of these differences between FaBER and other types of mindreading, it cannot be assumed that the processes characteristic of FaBER can be extrapolated to other types of mindreading.

We begin by reviewing existing findings, some clinical and some experimental, that display a striking pattern of paired deficits between emotion production and face-based recognition (attribution). These findings have not previously been brought together with the explicit intent of examining them in the context of the TT–ST controversy. Next we argue that this pattern readily lends itself to a simulationist explanation, whereas existing data do not fit with a theory-based explanation. Finally, the core project of the paper is to formulate and evaluate four specific models of how normal mindreaders could use simulation to arrive at emotion classifications.

2. Paired deficits in emotion production and face-based recognition

In early studies, Ralph Adolphs and colleagues investigated whether damage to the amygdala affects face-based emotion recognition (Adolphs, 1995; Adolphs, Tranel, Damasio, & Damasio, 1994). These studies were motivated by the well-known role of the amygdala in mediating fear, including its prominent role in fear-conditioning and the storage of fear-related emotion memories (LeDoux, 1993, 2000). One patient studied by Adolphs et al. was SM, a 30-year-old woman with Urbach-Wiethe disease, which resulted in bilateral destruction of her amygdalae with sparing of adjacent hippocampus and other neocortical structures. Consistent with the important role of the amygdala in mediating fear, SM was indeed abnormal in her experience of fear. Antonio Damasio notes that SM "approaches people and situations with a predominantly, indeed excessively, positive attitude".

> S[M] does not experience fear in the same way you or I would in a situation that would normally induce it. At a purely intellectual level she knows what fear is supposed to be, what should cause it, and even what one may do in situations of fear, but little or none of that intellectual baggage, so to speak, is of any use to her in the real world. The fearlessness of her nature, which is the result of the bilateral damage to her amygdalae, has prevented her from learning, throughout her young life, the significance of the unpleasant situations that all of us have lived through. (Damasio, 1999, p. 66)

Other lines of evidence also suggest that SM is abnormal in her experience of fear. For example, in one experiment, SM was presented with a conditioned stimulus repeatedly paired with a startle-inducing unconditioned stimulus, a boat horn delivered at 100 dB.

However, she failed to demonstrate a conditioned autonomic reaction to the conditioned stimulus, indicating she had an abnormality in acquiring or expressing conditioned emotion responses (Bechara et al., 1995).

Adolphs et al. (1994) tested SM against a number of brain-damaged controls in various FaBER tasks. Subjects were presented with photographs or video slides depicting facial expressions and asked to identify the emotion states to which the expressions correspond. SM was abnormal in face-based recognition of the emotion fear; her ratings of fearful faces correlated less with normal ratings than did those of any of 12 brain-damaged control subjects, and fell 2–5 standard deviations below the mean of the controls when the data were converted to a normal distribution. Subsequent studies have both confirmed and qualified these findings regarding co-occurring deficits in the production of fear and the ability to recognize expressions of fear in others. Sprengelmeyer et al. (1999) studied NM, another patient with bilateral amygdala damage. Like SM, NM was abnormal in his experience of fear. He was prone to dangerous activities (such as hunting jaguar in the Amazon river basin and hunting deer in Siberia while dangling from a helicopter!) and tested as abnormal on a self-assessment questionnaire measuring experience of the emotion fear. NM also exhibited a severe and selective impairment in face-based recognition for fear.

Adolphs et al. (1999) conducted an additional study of patients with bilateral amygdala damage with a larger-sized sample (nine patients, including SM). They again found that face-based fear recognition is abnormal among these patients. While deficits were most severe for fear recognition, recognition of other emotions, anger in particular, was also somewhat abnormal. Other neuropsychological studies are also broadly consistent with these findings (see Adolphs, 2002; Lawrence & Calder, 2004 for reviews).

The pattern noted here, that is, a *paired deficit* in the production and face-based recognition of an emotion, is not unique to fear. A similar pattern emerges with respect to at least two other emotions, disgust and anger, to which we now turn.

Paul Rozin and colleagues (Rozin, Haidt, & McCauley, 2000) conceptualized the emotion of disgust as an elaboration of a phylogenetically more primitive distaste response. Many aspects of taste processing are known from animal studies to be localized in the anterior insula region, known as the "gustatory cortex" (Rolls, 1995). Functional neuroimaging studies confirm that the anterior insula plays a similar role in taste processing in humans (Small et al., 2003, 1999).

What neural structures are implicated in the *recognition* of facial expressions of disgust? Sprengelmeyer and colleagues (Sprengelmeyer et al., 1996, 1997), using standard face-based emotion recognition tasks, found that patients with Huntington's disease display selective deficits in face-based recognition of disgust. In light of these findings, Phillips and colleagues undertook an fMRI study to see which brain areas are activated when subjects observe facial expressions of disgust (Phillips et al., 1998, 1997). The most striking finding for perception of facial expressions of disgust was activation in the right insula (adjacent regions such as the amygdala were not activated). They concluded that "appreciation of visual stimuli depicting other's disgust is closely linked to the perception [i.e. experience] of unpleasant tastes and smells" (Phillips et al., 1997, p. 496).

Lesion studies have also found paired deficits in the experience and facial recognition of disgust. Calder and colleagues found this pairing in NK, who suffered insula and basal

ganglia damage (Calder, Keane, Manes, Antoun, & Young, 2000). On a questionnaire NK's overall score for disgust was significantly lower than the scores of controls, whereas his scores for anger and fear did not significantly differ from the controls' mean scores. In tests of his ability to recognize emotions in faces, NK showed significant and selective impairment in disgust recognition. Adolphs, Tranel, and Damasio (2003) similarly found pronounced deficits in the experience and face-based recognition of disgust in a patient with bilateral insular and temporal lobe damage.

Wicker et al. (2003) did an fMRI study of disgust to determine whether the same neural regions are activated in normal subjects both during the experience of disgust and during the observation of the facial expression of disgust. In two "visual" runs, participants passively viewed movies of individuals smelling the contents of a glass (disgusting, pleasant, or neutral) and expressing the facial expressions of the respective emotions. In two "olfactory" runs, the same participants inhaled disgusting or pleasant odorants through a mask on their nose and mouth. The core finding of Wicker et al. was that the left anterior insula and the right anterior cingulate cortex are preferentially activated during the experience of the emotion of disgust evoked by disgusting odorants (compared to activation levels during pleasant and neutral odors), and this same region is preferentially activated during the observation of disgust facial expressions (compared to activation levels during pleasure-expressive and neutral faces). In other words, observation of disgust-expressive faces automatically activates the same neural substrates implicated in the experience of the same emotion.[1]

Anger is a third emotion system for which a paired deficit in emotion production and face-based recognition is found. Social agonistic encounters represent a distinct and phylogenetically recurrent adaptive problem for many animals. Various lines of evidence, reviewed in Lawrence and Calder (2004), suggest that the dopamine system has evolved as a neural subsystem involved in the processing of aggression in social-agonistic encounters in a wide variety of species, and this system plays an important role in mediating the experience of the emotion anger. They note that dopamine levels in rats and a number of other species are elevated in social-agonistic encounters. Conversely, administration of dopamine antagonists, such as the D_2 receptor antagonist *sulpiride*, selectively impairs responses to agonistic encounters. Lawrence and colleagues hypothesized that dopaminergic blockade by the administration of sulpiride would lead to selective disruption of face-based recognition of anger, while sparing the recognition of other emotions (Lawrence, Calder, McGowan, & Grasby 2002). This was indeed found. Following sulpiride administration, subjects were significantly worse at recognizing angry faces, though there were no such impairments in recognizing facial expressions of other emotions.

[1] Previous studies have indicated that the insula, among other structures, is activated during the experience of disgusting odors and tastes (Fulbright et al., 1998; Small et al., 2003). Additionally, previous studies have also established that the insula is preferentially activated during the observation of disgust-expressive faces (Krolak-Salmon et al., 2003; Phillips et al., 1998, 1997; Sprengelmeyer et al., 1998). However, the Wicker et al. (2003) study is the first to demonstrate within a single experiment, using a single mode of investigation and the same pool of subjects, that the same neural substrate subserves both the experience and recognition of disgust.

Based on the studies reviewed above, there is substantial evidence that *deficits in the production of an emotion and deficits in the face-based recognition of that emotion reliably co-occur*. How is this evidence relevant to the question of whether (face-based) emotion mindreading proceeds by tacit theorizing or by simulation?

3. Emotion mindreading by theory versus simulation

Let us further clarify and expound the two basic theoretical positions towards mindreading, which have loomed large in the literature. There are numerous ways of developing the TT idea, but the main idea is that the mindreader selects a mental state for attribution to a target based on *inference* from other information about the target. According to one popular version of TT, such an inference is guided by folk psychological generalizations concerning relationships or transitions between psychological states and/or behavior of the target (Gopnik & Wellman, 1992; Wellman, 1990). But we shall not insist on lawlike generalizations. The fundamental feature of TT is that it is an *information-based* approach. It says that attributors engage in mindreading by deploying folk psychological information. What they don't do, as a means to reading a target's mental state, is (try to) *model* or *instantiate* the very mental process that the target herself undergoes.

The core idea of ST is that the attributor selects a mental state for attribution after reproducing or "enacting" within herself the very state in question, or a relevantly similar state. In other words, she tries to replicate a target's mental state by undergoing (what she takes to be) the same or a similar mental process to one the target undergoes. If, in her own case, the process yields mental state M as an output, she attributes M to the target. For example, if she wants to attribute a future decision to a target, she might try to replicate the target's decision-making process in her own mind and use the output of this process as the decision to assign to the target. Alternatively, she may test a hypothesized state by simulating it in her own mind and seeing whether its upshots match those of the target. In either scenario, the attributor must recognize her own state as being of type M in order to select M as the state type occupied by the target. This presumably requires some sort of "information" about states of type M, so simulation isn't entirely information-free (as some proponents of simulationism maintain, e.g. Gordon, 1996). However, in contrast to TT, ST says that the relevant information about M is applied to something like a token or facsimile of a mental state in her own mind, not simply to information about the target from which she infers that the target instantiates M. There is, of course, much more to be said about the TT/ST contrast, but these points should suffice for present purposes (for additional details, see Gallese & Goldman, 1998; Goldman, 2000, in preparation).

How would TT and ST be applied to the task of face-based emotion recognition? First, how would TT explain the capacity to attribute emotions from facial expressions? We are not aware of any *specific* TT-based proposal in the literature. However, a general outline of what a TT-based account would look like is not hard to provide. It would propose, at a minimum, that people have a mentally represented body of generalizations for mapping representations of particular facial configurations to names for emotion states. When a target is observed displaying a particular facial expression, the attributor utilizes this body

of information, coupled with ordinary capacities for factual reasoning, to infer and attribute an emotion state to the target. Of course, this account presupposes that there is enough information in the facial expression itself to uniquely select an appropriate corresponding emotion state. This supposition seems plausible, as it has been shown that facial expressions exhibit rich geometric and configural properties sufficient for the purposes of inferring a corresponding emotion state (Calder, Burton, Miller, Young, & Akamatsu, 2001). Thus a TT-based account, elaborated along the lines we have suggested, is one legitimate contender for explaining how face-based emotion recognition occurs.

ST would approach this question in a different way. It would propose that an attributor selects an emotion category to assign to a target by producing an emotion in herself, or running her own emotional "equipment", and seeing which emotion has an appropriate link to the observed facial expression. Exactly how this simulational story would go is a matter to be addressed in detail below. In outline, however, the distinctive characteristic of the simulationist approach is to hypothesize that (normal) attributors execute face-based emotion attribution by means that somehow involve the production of that very emotion (at least in cases of accurate emotion detection).

4. Explaining the emotion recognition data by TT versus ST

The central claim presented in Section 2 was that there is substantial evidence concerning three emotions indicating that deficits in the production (experience) of an emotion and deficits in the face-based recognition of that emotion reliably co-occur. This strongly suggests that the same neural mechanisms subserve both the experience and the recognition of an emotion. In addition, the Wicker et al. (2003) study found that, in normals, the same neural regions were implicated in both the experience of disgust and the observation of disgust-expressive faces. Putting the Wicker et al. data together with the paired-deficit data, we have it that the same neural substrate is implicated in normals when they both experience and observe disgust, and when this same substrate is damaged, subjects fail to experience or recognize disgust at normal levels.

How does this evidence bear on the choice between theorizing and simulating as the explanation of face-based emotion recognition? On the surface, it strongly favors ST. If one (successfully) mindreads via simulation, one undergoes the same, or a relevantly similar, process to the one the target undergoes in using or arriving at the target state. Someone impaired in experiencing a given emotion will be unable to simulate a process that includes that emotion. Thus, ST predicts that someone damaged in experiencing fear, or in the neural substrate of fear experience, would have trouble mindreading fear. Hence, the phenomenon observed in SM—a paired deficit in fear experience and recognition—is straightforwardly predictable on the simulationist hypothesis. Similarly for the other paired deficits. By contrast, there is no reason to expect a paired deficit under TT. Why should conceptual representations of fear occur in the same region that underlies fear experience? That is, why should the processing of theoretical information *about* fear occur in the same region (or one of the regions) as the region subserving fear itself? TT predicts no such finding. Perhaps TT could be supplemented with auxiliary assumptions to make it consistent with the finding. We will discuss possible auxiliary assumptions the theory

theorist might deploy in a moment, but here we merely note that in the absence of such assumptions, TT would not lead one to expect the identified paired deficits. Thus, there is a prima facie case for ST over TT as an explanation of the phenomena.

Skeptics might reply that paired deficits could be due to a merely contingent relationship, for example the contingent co-localization of emotion experience and face-based emotion recognition (or classification) capacities, rather than the functional co-dependence that follows from a simulationist account. For example, lesions to the fusiform gyrus of the right occipital cortex produce both prosopagnosia, impaired ability to recognize faces (see Kanwisher, 2000), and achromatopsia, impaired perception of color (see Bartels & Zeki, 2000). But these two deficits have no interesting functional relationship to one another. It just so happens that the impaired capacities are at least partially co-localized in the fusiform gyrus, leading to the paired deficit. Isn't it possible that such a purely happenstantial story also applies to the paired deficits found in fear, disgust, and anger recognition?

In theory this is possible. But the fact that paired deficits in emotion production and emotion recognition occur for three distinct emotions makes the pairings seem far from contingent or accidental. They seem to reflect a *systematic* relationship between emotion experience and FaBER. So while a theory-based account can appeal to various auxiliary assumptions or hypotheses to account for the paired deficit data, for example the contingent co-localization hypothesis, these assumptions and hypotheses appear ad hoc relative to the simulationist explanation, which predicts the paired deficits in a more principled way.

Let us be more specific about how a TT explanation of the paired deficits might go. Three types of declarative knowledge might be used in the normal execution of FaBER tasks according to TT: (1) visually obtained knowledge of the facial configuration of the target, (2) semantic knowledge concerning these configurations, in particular knowledge that facial configuration C is paired with emotion label 'E', and (3) general knowledge concerning a given emotion, i.e. its typical elicitors or behavioral effects. In order to account for a paired deficit in one emotion, TT must say that one or more of these types of knowledge concerning the emotion in question is selectively damaged, while similar types of knowledge are preserved for other emotions. As it happens, there is specific evidence from the paired-deficit literature that paired-deficit patients do *not* suffer from reduced knowledge of types (1) or (3). A deficit in knowledge of type (2) is not specifically contravened by the evidence, but this proposal suffers from other problems.

The evidence about knowledge of types (1) and (3) is as follows. First, the paired deficit studies present evidence that subjects have no difficulty with perceptual processing of faces. In most of these studies, subjects performed normally on measures designed to identify any such impairments. The most common measure used was the Benton Face Matching Task, in which different views of unfamiliar faces must be categorized as belonging to the same face (Benton, Hamsher, Varney, & Spreen, 1983). Additionally, subjects were often found to be able to recognize other high-level properties of faces including age, gender and identity. SM's ability to recognize facial identity, for example, was fully preserved.

An informational deficit of type (3) is also disconfirmed by existing evidence. In the cited studies, deficits in FaBER routinely occur with preservation of subjects' general

declarative knowledge regarding emotions. For example, subjects can readily cite situations in which a person might experience the emotion whose face-based recognition is impaired. Recall that in describing SM, Damasio (1999) noted that "at a purely intellectual level" SM "knows what fear is supposed to be, what should cause it, and even what one may do in situations of fear". Similarly, Calder, Lawrence, and Young (2001) reported that "…patients with disgust recognition impairments are able to provide plausible situations in which a person might feel disgusted and do not show impaired knowledge of the concept of disgust". Finally, in most cases subjects' lesions occurred relatively late in life. So it cannot be plausibly argued that they lacked declarative knowledge about emotions because of deficits in their own emotional experience. They did suffer from experience deficits at the time of examination, but most had ample opportunities in earlier life to undergo relevant experiences and build normal declarative knowledge from those experiences.

So in order to account for the paired deficit data, the theory-theorist is likely to appeal to deficits in information of type (2), information consisting of semantic labels paired with representations of facial configurations. The theory-theorist will need to propose that labeling information of this kind for fear, disgust, and anger depends on the integrity of the amygdala, anterior insula, and dopaminergic system, respectively. Moreover, the theory-theorist must claim that it is possible to damage this labeling information quite selectively, in at least two ways. First, it must be possible to damage this labeling information for one emotion while leaving this information preserved for other emotions. Second, it must be possible to damage the labeling information in a way such that the label is inaccessible for *visual* representations of faces specifically, because, as just reported, impaired subjects have command of the label when verbally discussing general knowledge of the impaired emotion type. While these postulations are certainly possible, in the absence of any independent reason to believe that naming information is stored in this way, such postulations seem quite ad hoc. Thus, the kinds of deficits to which theory-theorists might appeal are either specifically contravened by the evidence or are quite ad hoc.

5. Possible simulationist models

Although the foregoing case for a simulational approach to face-based emotion recognition strikes us as compelling, it leaves open the question of how the simulational process proceeds. Those skeptical of our case for simulation may remain skeptical as long as no plausible, sufficiently detailed story of the simulation process in these cases is forthcoming. We get little help on this question from the existing literature. Articles describing paired deficits often contain conclusions hinting at a simulational explanation, but few pursue any details about the computational mechanisms. Exploring the options for a simulational process is the next task we tackle.

There are several ways a simulation heuristic might be used to attribute a mental state, depending on the nature of the causal link between the evidence events in the target (known to the attributor) and the sought-after state of the target. The causal link might be of two general kinds: (A) the evidence events cause the state, or (B) the state causes the evidence events. When an attributor has knowledge of prior states of the target, for example, specific desires and beliefs, and wants to predict a mental effect of those states,

e.g. a decision, we have an instance of type (A). Here the evidence events cause (or will cause) the sought-after mental state. When an attributor witnesses a target's piece of behavior (including a facial expression), and seeks to identify a mental state responsible for that behavior, we have an instance of type (B). Here the sought-after mental state is what causes the observed evidence. In face-based emotion recognition, the relevant connection is presumably of type (B). The target's emotional state causes her facial expression, and this expression is the evidence used by the attributor to identify the antecedent emotion state. How might simulation be used to exploit this kind of evidence?

5.1. Generate-and-test model

One possibility is a *generate-and-test* heuristic. As shown in Fig. 1, the attributor starts by hypothesizing a certain emotion as the possible cause of the target's facial display and proceeds to "enact" that emotion, that is, produce a facsimile of it in her own system. She lets this facsimile (or pretend) emotion run its typical course, which includes the production of its natural facial expression, or at least a neural instruction to the facial musculature to construct the relevant expression. If the resulting facial expression, or the instruction to construct such an expression, matches the expression observed in the target, then the hypothesized emotion is confirmed and the attributor imputes that emotion to the target. The simulation interpretation of the paired-deficit findings would say that this is the sort of thing that happens in emotion interpreters who are normal with respect to the emotion in question. Someone impaired in the relevant emotion area, however, cannot "enact" that emotion, or produce a facsimile of it. So she cannot generate the relevant face-related downstream activity necessary to recognize the emotion. Hence, a recognition impairment specific to that emotion arises.

Several issues about this model must be addressed. One question concerns the final phase of the postulated process, in which the system tries to "match" a constructed facial

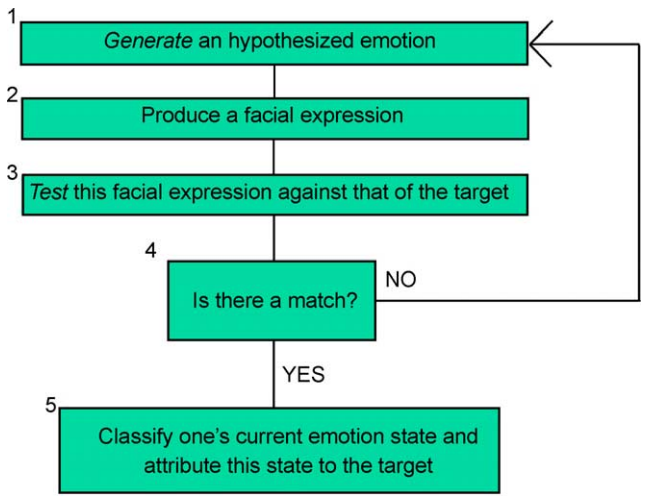

Fig. 1. Generate and test simulation.

expression with the expression observed in the target. The representation of one's own facial expression is presumably a proprioceptive representation, whereas the representation of the target's expression is visual. How can one "match" the other? One possible answer is that the system has acquired an association between proprioceptive and visual representations of the same facial configuration, through some type of learning. Alternatively, there might be an innate cross-modal matching of the sort postulated by Meltzoff and Moore (1997) to account for neonate facial imitation.

Second, there is a problem of how the generation process works. If candidate emotions are generated randomly, say, from the six basic emotions, the observer will have to covertly generate on average three facial expressions before hitting on the right one. This would be too slow to account for actual covert mimicry of displayed facial expressions, which occurs as early as 300 ms after stimulus onset (Dimberg & Thunberg, 1998; Lundquist & Dimberg, 1995). An alternative is to say that "theoretical" information is used to guide the generation process—though it isn't clear what theoretical information it would be. However, this proposal seems to turn the generate-and-test model into more of a theory–simulation hybrid rather than a pure simulationist model. Does this undercut the thrust of our simulationist argument? No. First, the simulational "test" phase of the generate-and-test heuristic is crucial, because without it the model cannot explain the paired deficits data. Second, the timing problems make this first model the least promising of the four we shall offer, and all of the other three are more purely simulationist in character.

5.2. Reverse simulation model

A second possibility, which seems to be implicitly endorsed by a number of theorists in the literature, is a *reverse simulation heuristic*. The idea in reverse simulation is that the attributor engages one of her own mental processes in the reverse direction, so as to attribute to the target a mental state that is temporally prior to the state which serves as the evidence for the attribution. In most cases in which simulation is deployed, reverse simulation is not an option: the standard forward directionality of mental processes precludes the possibility that these processes can be utilized in a reverse direction for the purposes of evidence-posterior interpretation tasks. However, there may be an important exception in the case of FaBER.

Under conditions of normal operation, the induction of an emotion episode causes a coordinated suite of cognitive and physiological changes, including, at least in the case of the so-called basic emotions, a characteristic facial expression (Ekman, 1992). Interestingly, this causal relationship appears bi-directional. There is substantial evidence that manipulation of the facial musculature, either voluntarily or involuntarily, has a causal effect in generating, at least in attenuated form, the corresponding emotional state and its cognitive and physiological correlates (see the discussion below). Thus, the relationship between emotion states and their facial expressions exhibits a kind of rough one-to-one correspondence in both directions. For this reason, the standard mode of operation in which emotion states causally produce a characteristic facial expression could potentially be utilized in a backwards direction for the purposes of reverse simulation. The underlying idea in "reverse simulation" has also been invoked by others in the mindreading literature.

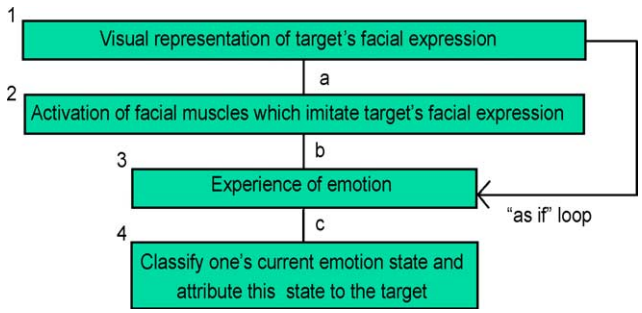

Fig. 2. Reverse simulation.

For example, Blakemore and Decety (2001) propose that the cerebellum houses a "forward model" which maps motor instructions into sensory expectations that would ensue given that an action is performed. They propose that this database can be queried in the reverse direction for the purposes of mindreading.

How would reverse simulation for the purposes of face-based emotion recognition operate? As shown in Fig. 2, a potential attributor who sees an emotion-expressive face starts by mimicking the facial expression she observes, presumably in an attenuated and largely covert manner. As noted above, actual facial exertions appropriate to a certain emotion commonly produce the experience of traces of that very emotion. An experiencer of that mild emotion would then classify her own emotion state, and then, in keeping with the common core of all simulational heuristics, would classify the observed face as being expressive of the same state produced in herself. Of course, all this might happen at a subthreshold level.

The reverse simulation model provides a plausible explanation for the paired deficit studies cited earlier. Someone impaired in the emotion in question would be unable to produce that emotion, or even significant traces thereof, in her own system. The requisite facial exertions would not arouse the appropriate neural activity for emotion production. Hence, such a person would be impaired in recognizing the corresponding emotion in a target. But in addition to explaining the paired deficit results, several independent lines of evidence support the reverse simulation model.

As depicted in Fig. 2, reverse simulation begins with a visual representation of the target's facial expression, which serves to activate facial musculature imitating the expression of the target. That such imitation capacities exist is well established. Meltzoff and Moore (1983) found that infants as young as one-hour-old imitate tongue protrusion, and a range of other facial displays, which they see modeled before them. In addition to the finding that humans *can* imitate the facial expressions of others, there is further evidence that humans *do* in fact spontaneously and rapidly activate facial musculature corresponding to visually presented facial expressions. In a series of studies, Dimberg and colleagues' found that presentation of pictures of facial expressions produces rapid, covert activation of one's own facial musculature, which mimics the presented faces (Dimberg & Thunberg, 1998; Lundquist & Dimberg, 1995). Such muscular activation is often subtle, but is detectable by electromyography and (as noted above) occurs extremely rapidly.

The finding that subjects spontaneously, rapidly, and covertly imitate visually presented facial expressions is consistent with the reverse simulation model (it supports link "a" in Fig. 2), but it is also consistent with a model in which these self-generated facial expressions are the *consequence* of an antecedently generated emotion state. Support for the claim that the muscle movements are primary and in fact give rise to a subsequent emotion state comes from two lines of evidence.

The first line of evidence suggesting that facial movements might occur prior to the emotion experience is that there are reasons to believe that that these facial movements are an instance of a more general "mirroring" phenomenon. An action mirroring system, in which internal action representations activated in the production of an action are also activated when the same action-type is observed in others, is known to exist in monkey and human ventral premotor cortex, and neighboring regions (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Iacoboni et al., 2001; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). Furthermore, the operation of the action mirroring system is known to generate covert activation of distal musculature. In an early experiment that helped establish an action mirroring system in humans, Fadiga and colleagues used transcranial magnetic stimulation (TMS) to enhance distal electromyographic recordings (Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995). They found that observation of actions (e.g. grasping an object, tracing a figure in the air) modeled by a target reliably produced electromyographically detectable activation in the corresponding muscle groups of the observer.

Recent evidence indicates that the action mirroring system may also operate during the observation of facial expressions. An fMRI study by Carr et al. (2003) found that subjects passively observing emotion-expressive faces display neural activation in the premotor cortex and neighboring regions which are normally activated in the production of facial movements, and which are in the region thought to house the action mirroring system. If there is indeed an action mirroring system that operates during the observation of facial expressions, it may help explain the covert activation of facial musculature discussed earlier. In other words, the phenomenon found by Dimberg and colleagues—covert activation of musculature that imitates the muscular activation patterns presented by a target—may be an instance of a mirroring phenomenon that obtains for somatic musculature more generally.[2] And if the facial movements found by Dimberg and colleagues are indeed an instance of a more general mirroring phenomenon, they need not be explained as the product of an antecedent emotion experience.

A second line of evidence suggesting that facial movements might occur prior to the emotion experience is the substantial accumulated data that there is a causal pathway that links manipulations of facial expressions with corresponding changes in emotion states. Subjects made to produce facial expressions voluntarily or involuntarily (for example, by holding a pencil in their mouth or saying "cheese") are found to exhibit cognitive and physiological correlates of emotion experience (Adelman & Zajonc, 1989). A number of theorists endorsing the "facial feedback hypothesis" view this causal pathway by which

[2] It's worth noting that the time after stimulus onset at which the Fadiga group detected covert activation of somatic musculature, 360 ms, is consistent with the time at which Dimberg and his colleagues found that differential activation of facial musculature reached significance, 300–400 ms.

facial expressions produce emotion states to be an important mechanism in mediating the experience of emotion generally (Laird & Bressler, 1992; Tomkins, 1962).

Facial feedback has also been implicated in the interpersonal communication of emotion states. A number of theorists have noted the interesting phenomenon of "primitive emotion contagion" in which motoric mimicry (mimicking facial expressions, prosody, posture, and movements) is the causal basis for convergence in emotion states between interacting individuals (Hatfield, Cacioppo, & Rapson, 1994). These theorists marshal substantial evidence in favor of the existence of primitive emotion contagion, and speculate that this phenomenon may play an important role in facilitating the interpretation of others' mental states, i.e. mindreading. The primitive emotion contagion hypothesis is broadly consistent with simulationist approaches to mindreading, and in particular with the reverse simulation model.

Theorists endorsing a facial-feedback hypothesis (link "b" in Fig. 2) have generally assumed that the mechanism by which self-generated facial expressions produce corresponding emotion states is mediated by proprioceptive sensation of the self-generated facial expression (Tomkins, 1981). Thus link "b" in Fig. 2 of the reverse simulation model appears to require proprioceptive mediation, which in turn implicates proprioceptive centers in the brain, in particular the somatosensory regions of the parietal cortices, in the process of emotion recognition. Adolphs and colleagues studied a large number of patients ($N = 108$ subjects) with cortical lesions, and found a significant association between right parietal lesions and impaired face-based emotion recognition (Adolphs, Damasio, Tranel, Cooper, & Damasio, 2000). A link between somatosensory impairment and face-based emotion recognition is predicted by the reverse simulation model, but is harder to make sense of under the generate-and-test simulation model. Additionally, it is not predicted at all under a theory-based model.

At least some data, however, are inconsistent with the reverse simulation model. Hess and Blairy (2001) used a more challenging FaBER task and found that while spontaneous facial mimicry did occur, the occurrence of successful mimicry did not correlate with accuracy in facial recognition, suggesting that facial mimicry may accompany but not actually facilitate recognition. Additionally, a study by Calder and colleagues found that three patients with Mobius syndrome, a congenital syndrome whose most prominent symptom is complete bilateral facial paralysis, performed normally on FaBER tasks (Calder, Keane, Cole, Campbell, & Young, 2000). Keillor and colleagues report a similar finding in which a patient with bilateral facial paralysis performed normally on FaBER tasks (Keillor, Barrett, Crucian, Kortenkamp, & Heilman, 2002). These findings need to be interpreted with caution, however, as given the long-standing nature of their impairments, these subjects' normal performance may reflect the operation of a compensatory strategy.

5.3. Reverse simulation with 'as if' loop

Adolphs et al. (2000) note that while facial feedback along the lines suggested by the reverse simulation model may be utilized in face-based emotion recognition, they speculate that there may be an alternative pathway. In keeping with earlier work (Damasio, 1994), Adolphs and colleagues propose that there may be direct links between a visual representation of a target's facial expression and a somatosensory representation of

"what it would feel like" were the observer to make that expression. It is speculated that these linked visual-somatosensory representations are the basis of an independent pathway (labeled the "as if" pathway in Fig. 2), which bypasses the facial musculature and allows the observer to directly produce an emotion state that corresponds to the facial expression displayed by the target. While details of the "as if" pathway are not fully in place yet, the "as if" model may count as a third simulationist model for face-based emotion recognition, which differs from the two models we've already proposed. In this third model, actual reverse simulation involving the facial musculature does not occur. Instead, the observer sees the facial expression displayed by the target and jumps directly to somatosensory representations of what it would feel like to have made the requisite facial exertions (thus the name "as if"), which in turn serves to bring about the corresponding emotion state as in standard reverse simulation. This "as if" model is superior to the reverse simulation model to the extent that it does not postulate a causal role for facial musculature in the recognition process, and is therefore unthreatened by the results of Calder et al. (2000), Hess and Blairy (2001), and Keillor et al. (2002).

5.4. Unmediated resonance model

A fourth possible model would also accommodate the findings by Calder et al. (2000) and Keillor et al. (2002) of preserved facial recognition in patients with facial paralysis. Unlike the third model, it would not appeal to somatosensory-based feelings normally associated with making a face of the same sort as the one visually presented by the target. This fourth model is what we shall call the *unmediated resonance* model. The idea here is that observation of the target's face "directly", without any mediation of the sorts posited by any of the first three models, triggers (sub-threshold) activation of the same neural substrate associated with the emotion in question. This is the idea behind Gallese's (2001, 2003) "shared manifold hypothesis", and is suggested by Wicker et al. when they speak of an automatic sharing, by the observer, of the displayed emotion (Wicker et al., 2003, p. 661). This would parallel findings of mirror-neuron matching systems found in monkeys and humans, in which internal action representations, normally associated with producing actions, are triggered during the observation of, or listening to, someone else's corresponding actions (Gallese et al., 1996; Kohler et al., 2002; Rizzolatti, Foggasi, & Gallese, 2001).[3] Finally, the fourth model must of course assume that the occurrence, or production, of the relevant emotion in an observer is transmitted to some cognitive center that "recognizes" the experienced emotion, leading to its overt (usually verbal) classification in the experimental set-up *as* that type of emotion. But this assumption would be common to all of the models, not distinctive to the fourth model.

Does this fourth model really fit the pattern of ST? Since the model posits unmediated resonance, it does not fit the usual examples of simulation in which pretend states are

[3] Of course, the role of mirroring in recognizing actions is not wholly clear. For example, Buxbaum, Sirigu, Schwartz and Klatzky (2003) and Halsband et al. (2001) found action production deficits alongside preserved action recognition. However, interpretation of these studies is difficult for several reasons, including most prominently the fact that it is unclear in these studies if the production deficits are due to damage in the region of the premotor cortex and neighboring regions thought to house the action mirroring system in humans.

created and then operated upon by the attributor's own cognitive equipment (e.g. a decision-making mechanism), yielding an output that gets attributed to the target. However, we do not regard the creation of pretend states, or the deployment of cognitive equipment to process such states, as essential to the generic idea of simulation. The general idea of simulation is that the simulating process should be similar, in relevant respects, to the simulated process (Goldman, in preparation). Applied to mindreading, a minimally necessary condition is that the state ascribed to the target is ascribed as a result of the attributor's instantiating, undergoing, or experiencing, that very state. In the case of successful simulation, the experienced state matches that of the target. This minimal condition for simulation is satisfied in the fourth model.

Finally, it should be emphasized that we make no attempt here to choose the "best" of the four simulationist models, or even to express a preference *ordering* among them (apart from the previously indicated doubts about the first model). Additional research is required before there is adequate evidence to select among them. Our sole aim is to show that several simulationist models are available with substantial surface plausibility and consistency with the evidence, which lends further credence to our initial conclusion that *some* sort of simulationist account of FaBER is highly probable.

6. Why simulation?

In the preceding sections, we marshaled evidence that FaBER proceeds by simulation rather than theory-based mechanisms. At any rate, simulation is the fundamental or primitive method of recognizing emotion from faces, although theorizing might also be used, for example, as a compensatory strategy. The evidence is much less clear-cut, however, with regard to distinguishing possible simulationist models, and much further study is warranted. Our hope is that the specific models we have formulated will lend structure to future discussion and investigation. We conclude by briefly addressing three topics. First, we reply to a reasonable worry that may have arisen about simulational models. Second, we speculate about the evolutionary origins of the simulational characteristics of FaBER. Third, we suggest one possible experiment to help select among the competing models.

Three of the simulationist models we proposed involve many disparate systems, including, for example, specific emotion production systems, the facial musculature and somatosensory centers. In contrast, a theory-based mechanism seems to require little more than the mental representation of a number of generalizations linking facial configurations with emotion names. Given desiderata such as simplicity, efficiency, or elegance, it's hard to see why simulational mechanisms would be advantageous relative to theory-based mechanisms for the FaBER task. Our answer to this challenge is straightforward. Simulation might be (somewhat) complex from a functional perspective, but it might be simpler from an evolutionary perspective. Simulation relies upon running the same emotional apparatus (possibly in reverse) that is already used to generate or experience the emotion. As a consequence, simulation routines do not require an organism to be outfitted with entirely new processes in order to confer an ability to recognize emotions in others. Natural selection is a tinkerer, not a planner, and she frequently builds new capacities from

existing ones. For this reason, simulation routines may have been favored in the course of evolution.

Our second remark features a bolder speculation—not original to us—about the evolutionary origins of a simulation mechanism for FaBER. As many writers point out (e.g. Wicker et al., 2003), there are many reasons why it would be adaptive to have mechanisms of emotion contagion. Consider the case of disgust, for example. Disgust is frequently experienced in response to a food item that should not be eaten. If an individual observes a conspecific having such a response to a food item, it would be adaptive for that individual to have the same disgust response vis-à-vis that food item, in order to induce avoidance. Thus, mechanisms of emotion contagion or resonance can be explained in terms of this kind of adaptive advantage. Once in place, the resonance mechanism could be transmuted into a *simulational* method of recognition. By contrast, there are no obvious steps whereby emotion contagion would be transmuted into *theory-based* recognition.

Finally, we turn to a suggestion for future experimental work prompted by our four models. D. Osherson (personal communication) asks what would happen if subjects were given the task of recognizing facially expressed emotions while doing (unrelated) face exercises, in particular exercises that prevent them from engaging their facial musculature in an emotion-expressive way. Previous studies have found that manipulation of the facial musculature in an emotion-expressive manner induces emotion experiences (see Section 5.2), and also produces interference on various emotion-relevant tasks (see, for example, Laird, Wagener, Halal, & Szegda 1982). But no studies have examined the effects on FaBER of exercises incompatible with making emotion-expressive faces. Would such exercises make FaBER more difficult? Would they induce errors? Both the generate-and-test model and the reverse simulation model seem to predict interference, and hence reduced recognition. On the other hand, neither the 'as-if' loop variant of the reverse simulation model nor the unmediated resonance model makes this prediction, because neither of them posits use of the attributor's own facial musculature in emotion recognition. So the first two models would predict a positive result (reduced recognition) in such a test, and would be undercut by a negative result. But a negative result would not undercut either the third or fourth model. This is one experimental means, then, by which to test the rival models.

Acknowledgements

The authors wish to thank Andrew Lawrence, Vittorio Gallese, Giacomo Rizzolatti, Christian Keysers, Ralph Adolphs, Daniel Osherson, and three anonymous reviewers for valuable comments on earlier drafts of this paper.

References

Adelman, P., & Zajonc, R. (1989). Facial efference and the experience of emotion. *Annual Review of Psychology*, *40*, 249–280.

Adolphs, R. (1995). Fear and the human amygdala. *Journal of Neuroscience*, *15*, 5879–5891.

Adolphs, R. (2002). Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews*, *1*(1), 21–62.

Adolphs, R., Damasio, H., Tranel, D., Cooper, G., & Damasio, A. (2000). A role for the somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *Journal of Neuroscience*, *20*(7), 2683–2690.

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

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the amygdala. *Nature*, *372*, 669–672.

Adolphs, R., Tranel, D., Hamann, S., Young, A. W., Calder, A. J., Phelps, E. A., Anderson, A., Lee, G. P., & Damasio, A. R. (1999). Recognition of facial emotion in nine individuals with bilateral amygdala damage. *Neuropsychologia*, *37*, 1111–1117.

Armstrong, D. M. (1968). *A materialist theory of the mind*. New York: Humanities Press.

Bartels, A., & Zeki, S. (2000). The architecture of the colour centre in the human visual brain: New results and a review. *European Journal of Neuroscience*, *12*, 172–193.

Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., & Damasio, A. R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science*, *269*, 1115–1118.

Benton, A. L., Hamsher, K., Varney, N. R., & Spreen, O. (1983). *Contributions to neuropsychological assessment*. New York: Oxford University Press.

Blakemore, S., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, *2*, 561–567.

Buxbaum, L. J., Sirigu, A., Schwartz, M., & Klatzky, R. (2003). Cognitve representations of hand posture in ideomotor apraxia. *Neuropsychologia*, *41*, 1091–1113.

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

Calder, A. J., Keane, J., Cole, J., Campbell, R., & Young, A. W. (2000). Facial expression recognition by people with Mobius syndrome. *Cognitive Neuropsychology*, *17*(1/2/3), 73–87.

Calder, A. J., Keane, J., Manes, F., Antoun, N., & Young, A. W. (2000). Impaired recognition and experience of disgust following brain injury. *Nature Reviews Neuroscience*, *3*, 1077–1078.

Calder, A. J., Lawrence, A. D., & Young, A. W. (2001). Neuropsychology of fear and loathing. *Nature Reviews Neuroscience*, *2*, 352–363.

Carr, L., Iacoboni, M., Dubeau, M.-C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Science USA*, *100*(9), 5497–5502.

Chaminade, T., Meary, D., Orliaguet, J.-P., & Decety, J. (2001). Is perceptual anticipation a motor simulation? A PET study. *NeuroReport*, *12*(17), 3669–3674.

Currie, G., & Ravenscroft, I. (2002). *Recreative minds*. Oxford: Oxford University Press.

Damasio, A. (1994). *Descartes' error*. New York: Grosset.

Damasio, A. (1999). *The feeling of what happens*. New York: Harcourt Brace.

Dimberg, U., & Thunberg, M. (1998). Rapid facial reactions to emotional facial expressions. *Scandinavian Journal of Psychology*, *39*, 39–45.

Ekman, P. (1992). Are there basic emotions? *Psychological Review*, *99*(3), 550–553.

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*, *73*(6), 2608–2611.

Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S. J., & Frith, C. D. (1995). Other minds in the brain: A functional imaging study of 'theory of mind' in story comprehension. *Cognition*, *57*, 109–128.

Fodor, J. A. (1992). A theory of the child's theory of mind. *Cognition*, *44*, 283–296.

Frith, C. D., & Frith, U. (1999). Interacting minds—a biological basis. *Science*, *286*, 1692–1695.

Fulbright, R. K., Skudlarski, P., Lacadie, C. M., Warrenburg, S., Bowers, A. A., Gore, J. C., & Wexler, B. E. (1998). MR functional imaging of regional brain responses to pleasant and unpleasant odors. *American Journal of Neuroradiology*, *19*(9), 1721–1726.

Gallese, V. (2001). The 'shared manifold' hypothesis: From mirror neurons to empathy. *Journal of Consciousness Studies*, *8*(5–7), 33–50.

Gallese, V. (2003). The manifold nature of interpersonal relations: The quest for a common mechanism. *Philosophical Transactions of the Royal Society of London, B*, *358*, 517–528.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*, 593–609.

Gallese, V., & Goldman, A. I. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, *2*(12), 493–501.

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

Goldman, A. I. (1992a). In defense of the simulation theory. *Mind and Language*, *7*(1–2), 104–119.

Goldman, A. I. (1992b). Empathy, mind and morals. *Proceedings and Addresses of the American Philosophical Association*, *66/3*, 17–41.

Goldman, A. I. (2000). The mentalizing folk. In D. Sperber (Ed.), *Metarepresentations: An interdisciplinary approach* (pp. 171–196). New York: Oxford University Press.

Goldman, A. I (in preparation). *The simulating mind*. New York: Oxford University Press.

Gopnik, A., & Meltzoff, A. (1997). *Words, thoughts, and theories*. MIT Press: Cambridge, MA.

Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really *is* a theory. *Mind and Language*, *7*(1–2), 145–171.

Gopnik, A., & Wellman, H. M. (1994). The theory theory. In L. Hirschfeld, & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257–293). New York: Cambridge University Press.

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

Gordon, R. M. (1992). The simulation theory: Objections and misconceptions. *Mind and Language*, *7*(1–2), 11–34.

Gordon, R. M. (1996). 'Radical' simulationism. In P. Carruthers, & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 11–21). New York: Cambridge University Press.

Halsband, U., Schmitt, J., Weyers, M., Binkofski, F., Grutzner, G., & Freund, H.-J. (2001). Recognition and imitation of pantomimed motor acts after unilateral parietal and premotor lesions: A perspective on apraxia. *Neuropsychologia*, *39*, 200–216.

Harris, P. L. (1991). The work of the imagination. In A. Whiten (Ed.), *Natural theories of mind*. (pp. 283–304). Oxford: Blackwell.

Harris, P. L. (1992). From simulation to folk psychology: The case for development. *Mind and Language*, *7*(1–2), 120–144.

Hatfield, E., Cacioppo, J., & Rapson, R. (1994). *Emotional contagion*. New York: Cambridge University Press.

Heal, J. (1986). Replication and functionalism. In J. Butterfield (Ed.), *Language, mind and logic* (pp. 135–150). Cambridge: Cambridge University Press.

Heal, J. (1996). Simulation and cognitive penetrability. *Mind and Language*, *11*, 44–67.

Heal, J. (1998). Co-cognition and off-line simulation. *Mind and Language*, *13*, 477–498.

Hess, U., & Blairy, S. (2001). Facial mimicry and emotional contagion to dynamic facial expressions and their influence on decoding accuracy. *International Journal of Psychophysiology*, *40*, 129–141.

Iacoboni, M., Koski, L. M., Brass, M., Bekkering, H., Woods, R. P., Dubeau, M., Mazziotta, J. C., & Rizzolatti, G. (2001). Reafferent copies of imitated actions in the right superior temporal cortex. *Proceedings of the National Academy of Science USA*, *98*(24), 13995–13999.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, *286*, 2526–2528.

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

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, *3*(8), 759–763.

Keillor, J. M., Barrett, A. M., Crucian, G. P., Kortenkamp, S., & Heilman, K. M. (2002). Emotional experience and perception in the absence of facial feedback. *Journal of the International Neuropsychological Society*, *8*(1), 130–135.

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science*, *297*, 846–848.

Krolak-Salmon, P., Henaff, M. A., Isnard, J., Tallon-Baudry, C., Geunot, M., Vighetto, A., Bertrand, O., & Mauguiere, F. (2003). An attention modulated response to disgust in human ventral anterior insula. *Annals of Neurology*, *53*, 446–453.

Laird, J. D., & Bressler, C. (1992). The process of emotion experience: A self-perception theory. In M. Clark (Ed.), *Review of personality and social psychology* (*Vol. 13*) (pp. 213–234). *Emotion*, New York: Sage.

Laird, J. D., Wagener, J. J., Halal, M., & Szegda, M. (1982). Remembering what you feel: Effects of emotion and memory. *Journal of Personality and Social Psychology*, *42*, 646–675.

Lawrence, A. D., & Calder, A. J. (2004). Homologizing human emotions. In D. Evans, & P. Cruse (Eds.), *Emotions, evolution and rationality*. New York: Oxford University Press.

Lawrence, A. D., Calder, A. J., McGowan, S. M., & Grasby, P. M. (2002). Selective disruption of the recognition of facial expressions of anger. *NeuroReport*, *13*(6), 881–884.

LeDoux, J. E. (1993). Emotion memory systems in the brain. *Behavioral Brain Research*, *58*, 69–79.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, *23*, 155–184.

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

Lewis, D. K. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, *50*, 249–258.

Lundquist, L., & Dimberg, U. (1995). Facial expressions are contagious. *Journal of Psychophysiology*, *9*, 203–211.

McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Science USA*, *98*, 11832–11835.

Meltzoff, A. N., & Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Development*, *54*, 702–709.

Meltzoff, A. N., & Moore, M. K. (1997). Explaining facial imitation: A theoretical model. *Early Development and Parenting*, *6*, 179–192.

Nichols, S., & Stich, S. (2003). *Mindreading*. Oxford: Oxford University Press.

Nichols, S., Stich, S., Leslie, A., & Klein, D. (1996). Varieties of off-line simulation. In P. Carruthers, & P. Smith (Eds.), *Theories of theories of mind* (pp. 39–74). Cambridge: Cambridge University Press.

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

Perner, J. (1996). Simulation as explicitation of predication-implicit knowledge about the mind: Arguments for a simulation-theory mix. In P. Carruthers, & P. Smith (Eds.), *Theories of theories of mind* (pp. 90–104). Cambridge: Cambridge University Press.

Phillips, M. L., Young, A. W., Scott, S. K., Calder, A. J., Andrew, C., Giampietro, V., Williams, S. C., Bullmore, E. T., Brammer, M., & Gray, J. A. (1998). Neural response to facial and vocal expressions of fear and disgust. *Proceedings of the Royal Society of London, B*, *265*, 1809–1817.

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., & David, S. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature*, *389*, 495–498.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *4*, 515–526.

Rizzolatti, G., Fadiga, L., Gallese, V., & Foggasi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*, 131–141.

Rizzolatti, G., Foggasi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews, Neuroscience*, *2*, 661–670.

Rolls, E. T. (1995). Central taste anatomy and neurophysiology. In R. L. Doty (Ed.), *Handbook of clinical olfaction and gustation* (pp. 549–573). New York: Dekker.

Rozin, P., Haidt, J., & McCauley, C. (2000). Disgust. In M. Lewis, & J. Haviland (Eds.), *Handbook of emotions* (pp. 637–653). New York: The Guilford Press.

Sellars, W. (1956). Empiricism and the philosophy of mind. In H. Feigl, & M. Scriven (Eds.), *The foundations of science and the concepts of psychology and psychoanalysis* (pp. 253–329). *Minnesota studies in the philosophy of science, 1*, Minneapolis, MN: University of Minnesota Press.

Shoemaker, S. (1975). Functionalism and qualia. *Philosophical Studies*, *27*, 291–315.

Small, D. M., Gregory, M., Mak, R., Gitelman, D., Mesulam, M. M., & Parrish, T. (2003). Dissociation of neural representation of intensity and affective valuation in human gustation. *Neuron*, *39*, 701–711.

Small, D. M., Zald, D. H., Jones-Gotman, M., Zatorre, R. J., Pardo, J. V., Frey, S., & Petrides, M. (1999). Brain imaging: Human cortical gustatory areas: A review of functional neuroimaging data. *NeuroReport*, *10*, 7–14.

Sprengelmeyer, R., Rausch, M., Eysel, U. T., & Przuntek, H. (1998). Neural structures associated with recognition of Facial expressions of basic emotions. *Proceedings of the Royal Society of London (Series B: Biology)*, *265*, 1927–1931.

Sprengelmeyer, R., Young, A. W., Calder, A. J., Karnat, A., Lange, H., Homberg, V., Perrett, D. I., & Rowland, D. (1996). Loss of disgust: Perception of faces and emotions in Huntington's disease. *Brain*, *119*, 1647–1665.

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

Sprengelmeyer, R., Young, A. W., Sprengelmeyer, A., Calder, A. J., Rowland, D., Perrett, D., Hornberg, V., & Lange, H. (1997). Recognition of facial expressions: Selective impairment of specific emotions in Huntington's disease. *Cognitive Neuropsychology*, *14*(6), 839–879.

Stich, S., & Nichols, S. (1992). Folk psychology: Simulation or tacit theory? *Mind and Language*, *7*(1–2), 35–71.

Tomkins, S. (1962) (*Vol. 1*). *Affect, imagery, consciousness: The positive affects*, New York: Springer.

Tomkins, S. (1981). The role of facial response in the experience of emotion: A reply to Tourangeau and Ellsworth. *Journal of Personality and Social Psychology*, *45*, 355–357.

Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.

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.