Mental mirroring as the origin of attributions
(Word count: 10,167)

Daniel A. Weiskopf

Department of Philosophy

University of South Florida

4202 East Fowler Ave., FAO 226

Tampa, FL 33620

Phone: 813–974–6915

Fax: 813–974–5914

Email: weiskopf@luna.cas.usf.edu

Abstract: A 'Radical Simulationist' account of how folk psychology functions has been developed by Robert Gordon. I argue that Radical Simulationism is false. In its simplest form it is not sufficient to explain our attribution of mental states to subjects whose desires and preferences differ from our own. Modifying the theory to capture these attributions invariably generates innumerable other false attributions. Further, the theory predicts that deficits in mentalizing ought to co-occur with certain deficits in imagining perceptually-based scenarios. I present evidence suggesting that this prediction is false, and outline further possible empirical tests of the theory.

Mental mirroring as the origin of attributions

1. Introduction

One of the most remarkable features of our mental lives is the seeming ease with which most psychologically and neurologically normal adults come to attribute mental states to others, and to predict others' behavior on the basis of those attributions. Developmental studies suggest that infants are biased to treat their conspecifics as being moved by different causal influences than the other, unthinking objects in their vicinity, and the emergence of mentalistic talk is well under way by the time the child is four years old (Bartsch & Wellman, 1995; Wellman, 1990).

I will use the term 'folk psychology' to describe our ability to understand ourselves and others as having mental states that are implicated in the explanation of our actions.[1] The exercise of that ability, our folk psychological practice, is a fact about us that demands explanation, as with other cognitive capacities that emerge early and with little if any external training or instruction.

In recent years, two opposing models have been proposed for explaining our folk psychological competence: the Theory Theory and the Simulation Theory. Despite the neat division these labels suggest, as the original models have developed they have become less distinct, and it is increasingly clear that neither pure theory is adequate to

---

[1] This contrasts with the usage of those that identify folk psychology with our naive 'theory of mind'. To say that folk psychology is to be explained by our possession and use of a theory would beg the question at hand. Further, the view I adopt contrasts with that of Goldman (1993), who claims that folk psychology should be thought of denoting our possession of mental state concepts. In my view, it is the ability to deploy those concepts in folk attributions and predictions that is of primary interest.

account for all of the data.[2] Still, the labels are useful insofar as they suggest families of models that have significantly non-overlapping commitments. In this paper, I consider one particular version of Simulation Theory proposed by Robert Gordon, one of the initial proponents of the Simulation Theory, and among its strongest defenders (see also Goldman, 1989 and Heal, 1986). To distinguish his view from others in the simulation camp, Gordon dubs his position 'Radical Simulationism' (Gordon, 1996b). Herein I will argue that Radical Simulationism's account of mental state attribution is insufficient to explain how we ordinarily ascribe attitudes, and leads to empirical predictions that are not borne out by the evidence.

In Section 2, I present Gordon's theory and attempt to make explicit each step in the mechanism of mental state attribution, so that we can more easily see the empirical commitments of the view. In Section 3, I pose what I call the problem of adjustment for Radical Simulationism's explanation of mental state attribution. In Section 4, I argue that a plausible modification of Radical Simulationism to answer the problem of adjustment leads in turn to what I call the problem of inferential promiscuity, and argue that as the account stands the problem is insurmountable. In Section 5, I assess sources of empirical evidence for or against Radical Simulationism, drawing primarily on studies of visual perspective-taking in subjects with impaired mentalizing. In Section 6, I outline a framework for further tests of the Radical Simulationist architecture. I offer some conclusions in Section 7.

---

[2] For central papers in the debate over how to understand folk psychology, see Davies & Stone (1995a, 1995b), Carruthers & Smith (1996).

2. An outline of Radical Simulationism

Folk psychological competence has two components. On the one hand, there is the ability to ascribe to other creatures mental states such as beliefs, desires, emotions, pains, etc. Call this the *interpretive* component of folk psychology. To take some simple examples, I typically suppose that a person is attending to, and therefore has some beliefs about, the part of the world that is currently in front of her opened eyes. My cat bats at her food dish, and on this basis I ascribe to her the desire that her dish be filled. Based on the way my restless students are shifting in their seats, I ascribe to them the wish that class would let out early. In each case, I may draw on different sources of evidence, and have variable confidence in the specific terms in which my attributions are couched, but I typically have some inkling of others' mental states.

Having interpreted a subject's mental states, I can proceed to make predictions about how she will decide to behave. Call this the *predictive* component of folk psychology. If I have ascribed anxiety about his paper grade to my student, I can predict that he will exhibit certain signs, such as gaze avoidance. If I attribute the desire for a bran muffin to my companion, I will expect her to lead me towards the corner cafe in the hope of acquiring one. Although our explanations are often implicitly hedged by the assumption that innumerable other factors will remain equal, they nonetheless succeed in a wide range of cases.

Radical Simulationism is a theory about how we are able to interpret others and predict their behavior. It consists of two parts, corresponding to the distinct components of folk psychology. What I have called interpretation is explained by Gordon as a kind of

internal modeling of another person's mental economy. To interpret another, I must imaginatively create perceptual and interoceptual circumstances that are similar to their own. Having imagined these circumstances, I find that I myself would have certain attitudes if they were the case, and I then treat the attitudes I find myself holding as if they were the other's attitudes. Although this process is itself sometimes referred to as 'simulation', I will follow a hint of Gordon's and call the process by which we attribute mental states to each other 'mirroring' (Gordon, 1996a, p. 165).[3]

I will reserve the term 'simulation' to refer to the mechanisms that implement our ability to engage in prediction on the basis of the attributions generated by mirroring. While discussion of the role of simulation in underwriting prediction has occupied much of the literature on the Simulation Theory (Stich & Nichols, 1992; Stich, Nichols, & Leslie, 1996), I am mainly interested in mirroring as an explanation of folk interpretation. Therefore, I will only briefly mention Radical Simulationism's account of simulation-as-prediction.

Folk prediction is identified with the 'offline' use of the same practical reasoning mechanisms that generate my own behavior. According to a fairly standard model of the causation of behavior, when I act, my beliefs and desires are fed into my internal decision-making apparatus, which does a bit of decision-theoretic calculation and outputs an intention. These intentions are then passed to whatever systems are responsible for

---

[3] A brief terminological note: this use of the term 'mirroring' should not encourage confusion between (i) mirroring as an explanation of our ability to attribute mental states to others and (ii) the mirror neurons discovered by Rizzolatti, Fadiga, Gallese, & Fogassi. (1996). Mirroring in the sense under discussion here is a theory of folk attributive competence, not a theory of the neural implementation of that competence. It is independent of particular neural mechanisms such as those discovered by Rizzolati and colleagues (although some Simulation Theorists have seen the existence of mirror neurons as supporting their view; see Gallese & Goldman, 1998). When I speak of 'mirroring', I will be using the first sense, unless noted otherwise.

moving my body to carry them out.[4] Simulation involves decoupling the decision-making system from its typical inputs and outputs. Whereas it usually accepts attitudes from my own mental economy, in simulation it accepts a set of artifactual attitudes constructed with the goal of modeling my interlocutor's mind, and its outputs are treated as predictions of her future state rather than as instructions for my own motor systems.

This exposition illustrates some principles of Radical Simulationism. It explains our facility at prediction by appeal to the fact that we can (at least sometimes) count on others who share our mental states behaving in essentially the same way as we do, in virtue of our shared cognitive architecture. As a version of the Simulation Theory, it claims that we do not need a theory of the causal interaction of inner states in order to say under what circumstances a creature will have certain attitudes, nor what actions it will undertake as a result. At least, we do not explicitly operate with such a theory in making predictions. We are architecturally similar to our subjects, and we form plans of action all the time. Our predictions will be accurate given, first, that our practical reasoning mechanisms are approximately the same as those of the subject we are predicting; and, second, that the attitudes that we feed to it are an accurate enough reflection of the subject's own. An account of mirroring, or theory-free attribution, is intended to fulfill the second condition.

A mirror does not need to know a theory of optics to reproduce what is before it. By analogy, neither do we need to know a theory of mental states to reproduce those minds with which we come into contact. How does mirroring employ an interpreter's own emotional, imaginative, and practical resources?

---

[4] This standard picture is well captured by what Nichols & Stich (2000, p. 121) refer to as the 'Basic Architecture Assumption'.

The process can be divided into two major stages: (1) generation of an imaginative scenario, and (2) generation of attributed mental states on the basis of this scenario. During Stage 1, I attempt to imaginatively put myself in another's circumstances. I might begin by shifting my spatial location to coincide with the person I am trying to interpret. Suppose that a venomous snake is rapidly slithering towards my subject. To interpret her mental states, and predict her likely response, I imaginatively shift my spatial location so that it seems as if the snake is slithering towards me. In one of his earliest examples of how mental state attribution might proceed, Gordon considers how he might understand the mind of his waiter:

> As a first step, I shift spatiotemporal perspectives—I am standing over there now, where the waiter is, not sitting here. In some cases, shifting spatiotemporal perspectives might be enough: e.g., for predicting, or explaining, the behavior of a person I see in the path of an oncoming car (Gordon, 1995a, p. 64).

In a later discussion, Gordon claims that when mirroring,

> what you are doing is shifting the locations and vectors of environmental features on your egocentric map—that is, the mental map in which things and events are represented in relation to yourself, here, and now—so as properly to engage your location-specific or vector-specific tendencies to action or emotion. (Gordon, 1995b, p. 108)

In this first stage of mirroring I may either imagine myself to have shifted with respect to the environment, or imagine that the environment itself has changed in some respects, or both.

Having made the appropriate imaginative shifts in my circumstances, in Stage 2 of mirroring I generate a set of attitudes on the basis of those fictional circumstances. Again, this involves exploiting mechanisms that must be in place already. As I move about and experience the world, my beliefs and desires are constantly shifting and being updated as a result of my ongoing perceptions. There must be some mechanism that is responsible for this updating process, although little is understood about it. In mirroring, these processes take the imagined circumstances I have generated as input, rather than taking as their input the normal sensory channels. For example, if I imagine myself faced with an ice cold beer on a hot summer day, and feed this representation to my belief and desire generating mechanisms, I find myself with a desire to drink the beer. If I find myself faced with a venomous snake slithering towards me, I am afraid of the snake and desire to get away from it. These mental states are the result of some inference made from the imagined premises, in combination with the background of my own beliefs, desires, and emotions.

Crucially, the mental states that are generated as the result of the operation of these mechanisms are not fed into my own body of beliefs and desires. I don't come to desire the beer that I am imagining myself faced with. Rather, the mental states that are the output are in turn fed into a temporary memory workspace that segregates them from my own attitudes (Gordon, 1996a, p. 168). This workspace is indexed to a representation of my subject, the person that I am currently attempting to mirror.[5] Having these beliefs and desires in an indexed workspace constitutes attributing them to my subject.

---

[5] Nichols & Stich (2000) argue that it is necessary to maintain a functional segregation between pretended states and beliefs. It should be clear from this exposition that Gordon agrees with them, since imposing this segregation is just what the temporary memory workspace is for.

Finally, we should note that the process of generating the beliefs and desires that I attribute to my subject has access to my own beliefs and desires. These form the 'database' on which the process draws. The attitudes that I attribute are only generated out of the interaction of my own attitudes with the imagined circumstances that I present to my inferential mechanisms. So the processes at work in mirroring should be able both to create and access the contents of a temporary memory workspace and my own belief and desire 'boxes'.[6]

According to Radical Simulationism, then, attributing a mental state to another is a matter of undergoing this kind of imaginative process, drawing on my own emotional and cognitive resources, and arriving at a set of mental states that are kept functionally distinct from my own. The states generated as a result of this process are treated as if they belonged to the subject that I am interpreting.

In attributing mental states to another person, Gordon believes that I begin by filling the workspace with my own attitudes and emotions. However, it will never prove successful to attribute my whole mental economy to someone else. Among other reasons, at any moment the two of us are always located at separate locations in space. Hence she will have a somewhat different outlook on the environment than I have, and vice versa. I may be gazing at the Mondrian print hanging on the wall behind her, while she has a view of the Chicago skyline, to which my back is turned. Whereas I believe that there is a Mondrian on the wall, she may not. So the default strategy of 'total projection' (Gordon, 1995b, p. 102) will typically fail. Initially I might make simple corrections to total projection. So I could use my visuospatial imagination to make it seem to me as if I were

---

[6] As Gordon & Barker (1994, p. 171) put it, in imagining a scenario for the purposes of pretend play, what a person does 'depends not only on the stipulated pretend facts, along with his existing perceptions and beliefs, but also on his existing desires, values, and norms.'

moved forward several feet and rotated 180 degrees. This process might yield the belief

that I can see the Chicago skyline, so this belief is placed into the mirroring workspace.

What makes Radical Simulationism radical is that the process of mirroring does

not require that I have or employ the concept of belief:

> For what I oppose is the claim that simulation requires recognition of our own
>
> mental states as such, along with the corollary that it requires possession of the
>
> concepts of the various mental states simulated in others. (Gordon, 1996b, p. 16)

Call this prohibition on using mental state concepts in simulation the 'Radical

restriction'. According to Gordon, all I need to be able to do is to generate beliefs of my

own in response to the promptings of my imagination. It is the endogenously generated

belief itself that goes into the workspace, not a representation of the belief as such. This

means that we are not operating internally with a description of the subject such as 'Anne

believes that the stars are coming out'. Instead, I treat the beliefs that my imagination

causes in me as belonging to my subject. Mirroring is not determining what *I* would think

in such circumstances, but determining what my companion thinks by transforming part

of myself into a duplicate of her mental states.[7] Gordon calls this 'the crucial difference

between simulating *oneself* in O's situation and simulating *O* in O's situation' (1995c, p.

55). I do not imaginatively generate what my own attitudes would be and then attribute

---

[7] Gordon sometimes refers to mirroring as involving 'pretend' beliefs and desires. But it's a question as to how this distinguishes them from normal beliefs and desires. If they were not closely related to genuine beliefs and desires, it's unclear how they could be fed into my decision making mechanisms, at the stage of generating predictions. They are exceptional in that they are generated and kept in a workspace that is functionally segregated from my own store of beliefs and desires. Hence, they differ from straightforward beliefs and desires in their causes and effects. But as Currie & Ravenscroft (2002, p. 11-19) point out, they must nevertheless share the majority of their internal functional role with beliefs and desires that are not generated in imagination. They must have, for example, the same tendencies to interact to form further mental states, even if those do not necessarily lead to actions.

those attitudes to another. Rather, I generate imagined attitudes that are *directly* ascribed to another.

Radical Simulationism, then, claims that to interpret others, we need only to reproduce to ourselves similar perceptual and interoceptual circumstances to their own, and let our native attitude-generating mechanisms mirror their psychology automatically. Having the machinery could suffice, even if we lacked a theoretical owner's manual.[8]

3. The problem of adjustment

Adjusting my attitudes to match another's presents a special problem for Radical Simulationism. I may differ psychologically from my interlocutor in countless ways. Given the mechanism Radical Simulationism proposes to underlie mirroring, we can predict that mental state attribution will be difficult in proportion to the amount of imaginative stretching necessary in order to match another's states. The problem of adjustment is to find a way in which, using my ability imaginatively to present myself with scenarios, I can generate correct attributions for a subject whose mental states differ significantly from my own.

We can analyze the task of carrying out mental mirroring as a sort of inferential game, akin to a process of counterfactual reasoning. The method of mirroring is to make a series of suppositions about the course of my experience that would lead to the appropriate 'conclusion', which is that I occupy a mental state that is type-identical with

---

[8] Currie (1995, p. 26) summarizes this process: 'one way to form a view about your decision is for me to imagine myself in your situation, coming thereby to have 'pretend' beliefs and desires: pretend versions of the beliefs and desires I would have if I were in that situation. These pretend beliefs and desires are then engaged in reasoning and decision making that is run "off-line", disconnected from behavioural outputs'.

my subject's. This goal state is then treated as if it belonged to the subject I am mirroring. I therefore need to find a set of conditions C that can fill in the schema:

> If conditions C were to obtain, then (given the background of my attitudes and emotions) I would have attitude S.

My being caused to have attitude S, the target attitude, as a result of imagining conditions C to obtain constitutes a successful attitude attribution. The Radical restriction discussed earlier further says that conditions C must be specified without the use of mental state concepts. It is permitted, though, for C to describe aspects of the external world, since that is what I am systematically varying in imagination in order to produce the target attitude.

I don't mean to suggest that Gordon would necessarily endorse precisely this way of describing mental mirroring. The representation of the process given above is intended as an analysis of the task of mirroring, not as a psychological description of how people actually carry out that task. We do not, on Gordon's view, explicitly and consciously reason our way from a set of conditions plus a description of our own attitudes to a target mental state. That might suggest that we employ a *theory* of how our mental processes would be influenced by various counterfactual situations and experiences, and Gordon repudiates the use of such psychological theories.

However, Gordon *has* sometimes described his view in terms of reasoning to a target 'conclusion' mental state or action. In his (1995a), for example, he says that 'at [simulation's] heart is a type of reasoning I characterize as *hypothetico-practical*' (p. 64). This is a form of reasoning that involves imagining various transformations of the actual world, generating attitudes on the basis of those transformations, and using those attitudes

to generate predictions of the subject's behavior. The first two stages of this process are what I have been characterizing as mental mirroring. Reasoning is needed here to decide what sorts of imaginative transformations are most likely to lead, ultimately, to the behavior that the subject displays, even if the processes of imagination, attitude generation, etc., are not themselves process of reasoning. The role of reasoning is in choosing the appropriate conditions to imagine.

Although I don't intend to question the Radical restriction here, it is worth considering why one might impose it. One motivation might come from the fact that what we are trying to explain is our capacity to attribute mental states to others. Radical Simulationists claim that attributing a belief to another person is explained by having a belief (or representation playing a sufficiently belief-like role) in a mental workspace indexed to that person. So what needs to be explained is how we can come to generate appropriately indexed imagined beliefs and desires. One way to generate such indexed belief-like representations would be simply to imagine ourselves to have certain mental states—for example, to imagine *my believing that such-and-such*—and then insert those imagined beliefs into the workspace. But what we are trying to explain is precisely how we come to be *able* to generate imagined beliefs and desires. On this account, we would in effect be presupposing one of the capacities that we are trying to explain. Hence, what is imagined to be the case must not include mental state concepts.

Gordon's argument in favor of the Radical restriction proceeds from the fact that 'we have imitative mechanisms that work without such [mental state] concepts' (1995b, p. 113). Infants tend to mimic mother's facial expression when in unfamiliar situations. As a result of this reflexive mimicry, they may tend to take on an emotional state—fear,

anxiety, or calm—that matches mother's. Although they can "catch" these attitudes, says Gordon, they cannot yet attribute these mental states. He concludes: 'Why then should such [attributive] capacities be required where the imitation is limited to the *imagination*?' (1995b, p. 113). In virtue of the ability to imagine undergoing experiences similar to those of the subject of interpretation, we can generate the right attitudes. 'So an adult needn't use the concept of interest, in order to simulate the naturalist's interest in the behavior of wild animals, nor the concept of fear in order to simulate her lack of fear in the proximity of the grizzly' (Gordon, 1995b, p. 114). This argument appears to rest on considerations of conceptual economy: it is unlikely that the child possesses the relevant mentalistic concepts, and it is *unnecessary* to suppose that she does in order to explain her behavior. So it is unnecessary in the case of adult mirroring as well.

One initial difficulty with Gordon's proposal is that there is a large class of beliefs and other attitudes whose causal etiology bears no straightforward relationship to any way that I might imagine the world or my experiences to be. Consider highly theoretical beliefs, which draw on multiple sources of evidence that are far removed from any directly experienced circumstances.[9] For example, take the belief that protons are made up of two up quarks and one down quark. Coming to hold a theoretical belief such as this, as confirmation holists remind us, relies on a vast number of other beliefs that are equally theoretical in character. It is unclear how mirroring might be extended to deal with these cases, but for the time being I will put them aside and focus on what are arguably the best cases for mirroring: attitudes having to do with immediately perceivable circumstances in the here and now, which can be jointly appreciated by both the interpreter and the target. If there is trouble with the best cases, we need not address the worst ones.

---

[9] The problem of theoretical beliefs was suggested to me by an anonymous referee.

With this setup in place, we can pose the problem of adjustment. As a test case, consider the problem of attempting to interpret someone who prefers a style of music you find unpleasant (Gordon, 1995b, p. 109-111).[10] How might interpretation of such a subject take place? Say that my friend Mary prefers listening to the Beatles, while I dislike them and prefer Miles Davis, and I am trying to mirror her enjoyment. How do I mirror her attitude successfully, given my initial state?

For me to construct a model of her, it is clearly not enough to suppose that I am having an auditory sensation as if I were hearing the Beatles (that is, to imagine myself in a situation in which a Beatles album is playing). By hypothesis, this elicits in me the desire to avoid hearing any more such noise, the belief that this is tuneless junk, and so on. These are plainly not the states that she herself would occupy under those conditions. Predictions based on this model would almost certainly leave me unable to explain why she continues to collect their albums, since I don't typically buy albums that I dislike (and neither does she). This failure suggests just putting myself in the same place in the world that she occupies is not sufficient, and that I need to rethink my interpretation.

I might attempt to mirror more fully Mary's experiential situation. For example, she looks contented when she hears the Beatles. 'Contentment' is a mentalistic term, so to continue scrupulously prescinding from mentalistic concepts, let us say she has *this* sort of look, picked out by demonstration. I might imagine wearing a similar expression, which might well cause me to be in a contented state as well. Bodily changes may be part of stepping into her shoes, and Gordon argues that we have an innate tendency to 'catch'

---

[10] Although I will present the problem in terms of mirroring desires and preferences, I think that it arises in connection with other kinds of mental state as well.

the emotions of others by making just such adjustments to the expressive muscles of our faces (Gordon, 1996a, p. 167).

At best, though, imagining myself to be in this enriched situation seems adequate to cause two *distinct* attitudes, namely a sensation of contentment simultaneous with the sensation of hearing the music. But this is not exactly her state, either; she is not just happy and listening to the music, she is happy *with* the music. The caused pleasure I enjoy as the result of mimicking her facial expression is not pleasure at the music, any more than the pleasure I might get from petting my cat while listening to a piece of music I dislike becomes pleasure at the music. It is not the same attitude as pleasure at the music itself. Or at least it is not obvious why the pleasure I might feel as a result of imagining the relevant sorts of bodily changes would automatically take as its object the music itself, particularly given that I don't like that kind of music. Recall that mirrored attitudes are generated in interaction with my own background attitudes. So this form of adjustment doesn't seem sufficient to generate the right attributions, either.

One further possibility involves the joint use of two mechanisms for sharing mental states that Gordon draws our attention to. The first is the mechanism of emotional contagion discussed above. The second is an attention-tracking mechanism: normal children and adults can follow the direction that a companion is pointing or looking, so that they are also gazing at the same target object. Often this happens automatically; furthermore, in autistic children, the ability to judge where a cartoon figure is gazing appears to be impaired (Baron-Cohen, 1995). Gaze-tracking mechanisms are useful in 'catching' the intentionality of an interlocutor's mental states (Emery, 2000). If I am gazing or pointing at a distal object, and my companion can follow my gaze or point, we

can come to have thoughts that are about the same object. Other cues may also be used in detecting another's attention (Langton, Watt, & Bruce, 2000). Orientation of the head and body both influence the perceived locus of attention. Although evidence suggests that these cues tend to be processed in parallel, head orientation and pointing may at times be even more salient than gaze, even for young children. Perhaps these two sets of mechanisms, one for sharing emotional states and the other for sharing intentional objects, could be combined into a mechanism for sharing the same emotional state directed at the same distal object. If so, this would go a great distance towards overcoming the problem of adjustment.

However, it isn't clear that these two mechanisms can be combined in this way. Mechanisms for the sharing of intentionality might suffice to generate some state in me that shares its intentional object with my companion's state, and mechanisms for emotional contagion might suffice to generate some emotional state in me that closely matches the intrinsic quality of my companion's emotion—that is, both might be states of fear, or pleasure, or disgust. But again, whether these two states combine to form, say, pleasure at the object of our joint attention, depends on how they interact with the vast background of my preexisting attitudes and emotions. And in the case where this standing background inclines me against preferring the object my companion prefers, it isn't obvious how the joint exercise of these two mechanisms will be sufficient to overcome it. I would add that Gordon seems tacitly to endorse this conclusion, since he goes on to propose a method for adjusting the imagined world that I place myself in that he thinks will allow me to mirror my companion's attitudes even when I do not share her preferences and desires.

What is at issue in considering the problem of adjustment is not whether we can always fully appreciate what it is that another finds appealing about some object that we ourselves do not find appealing. If I am a committed jazz snob, and I think that pop music is too simplistic and commercial to be enjoyable, then I probably cannot appreciate from the inside what it is that fans of pop music like about it. That is, I cannot appreciate those musical properties myself, at least not without remaking a large number of my own tastes and judgments, a move that is blocked by the Radical restriction on imagination. But this kind of understanding of the roots of someone else's attitudes is not obviously necessary just to attribute to them a liking that I don't share. And it is just the ability to attribute likings that I don't share that is at issue, since we clearly possess this ability, and mirroring is a putative explanation for it.

4. The problem of inferential promiscuity

Gordon offers a diagnosis of our difficulty with adjustment (Gordon, 1995b, p. 109). He notes that evaluative properties such as being good or being appealing are supervenient on descriptive properties of the appealing object. Call this 'the supervenience of the evaluative on the descriptive': Something is appealing in virtue of having some qualities that make it appealing (to a person or class of people). Our pro attitudes towards something are always maintained under some aspect or other. For example, I desire to hear Miles Davis because of the clear, lonely quality of his playing. His music has the property of being appealing (to me) in virtue of its having other particular qualities that it has. Other music that shared that quality would, other things being equal, also be appealing to me.

This suggests a strategy for mirroring. In order to adjust my attitudes towards Mary's favorite music, I might need to alter the imagined auditory qualities of the musical experience itself, to align them with the qualities that I find appealing in music. This means including among conditions C the proposition that the Beatles' music sounds like that of Miles Davis. (Perhaps they decided to be jazz musicians rather than pop artists.) At least it isn't inconceivable. If it were the case, then I would indeed desire to hear the Beatles. And this is the desire that I am attempting to mirror, so this supposition seems to do the trick.[11]

In fact, however, although this solves the problem of adjustment, a further problem appears in its wake. The problem of adjustment gives way to what I will call the problem of inferential promiscuity. Notice first that there is no end to the number of attitudes that would be caused in me were our condition to obtain. It is true that the target attitude would be present. But so would countless others, including the judgment that the Beatles are a fine jazz combo, or the belief that they are a nice example of post-bebop fusion. I believe that I would hold all of these in a world where the Beatles sounded like Miles Davis. Yet my subject holds none of these attitudes. She judges, for instance, that the Beatles sound like a pop group, that they have British accents, and so on. In supposing the Beatles to sound different than they actually do, I have altered the whole network of judgments that I make about them. The challenge is to find a way of 'screening out' all of the other attitudes that an act of imagination is likely to produce in me, in order to avoid generating such mistaken clusters of ascriptions.

---

[11] Gordon suggests that even more extreme situations might need to be imagined. E.g., he suggests that I might have to imagine that Linda Ronstadt's singing is in fact a philosophy lecture in order to come to enjoy it. I will stick with a comparatively conservative modification for purposes of illustration.

The problem of inferential promiscuity arises because once I suppose that the descriptive properties of an object are altered, I may come to the appropriate evaluative conclusion—I may conclude that under those circumstances I would like that object. Yet I would also conclude that I had many further beliefs about that object that are mistaken. But we do not actually make such attributions to people whose tastes differ from our own. Radical Simulationism needs some supplementation in order to prevent it from overgenerating these erroneous attributions.

Gordon might appeal to what he calls the 'principle of least pretending' as a way to eliminate the unwanted attitudes. This principle is the claim that 'other things being equal, I will prefer the less radical departure from the "real" world—i.e., from what I myself take to be the world' (Gordon, 1995a, p. 65). But this principle cannot be what Gordon needs here, since resisting this sort of imaginative 'stretch' will not succeed where our attitudes actually differ. As we saw, a significant stretch seems necessary to mirror some subjects. Furthermore, the principle is no help in telling me what I *should* imagine in order to arrive at true attributions. Knowing that I will prefer the minimal departure from the default of total projection does not yet tell me how to mirror a psychology different from my own. And even if I were to pick the minimal departure (say one that altered the sound of the Beatles to make them resemble music that I prefer), nothing about this departure would help to block the further unwanted attributions.

It is a general fact about the attitudes that they tend to be caused in bundles of indeterminately large size. Experience causes beliefs about the environment, pro and con attitudes towards various parts of it, and judgments of likeness and dissimilarity, to name only a few states. Changing any aspect of experience, even in imagination, is therefore

likely to have unintended and widespread ramifications. Gordon's strategy to avoid the problem of adjustment is to change the character of the imagined experience. This is a familiar part of our everyday counterfactual reasoning: among the circumstances in which I would admire the Beatles are those are those in which their sound resembles that of the Miles Davis Quintet. Yet in these circumstances I would also have numerous attitudes that fail to mirror those of any Beatles fan I am likely to attempt to interpret.

I would add that the problem that inferential promiscuity is not, *per se*, that the subject does not have the beliefs and desires that I am attributing to her. It might be that, as folk interpreters, we are systematically disposed to misattribute attitudes to others under some conditions. This can lead to predictive failure and other sorts of misunderstanding and miscommunication. But the theory of mental mirroring is an account of folk interpretive practice, with all of its flaws. Where folk interpretations are wrong, it is important that mirroring reproduce those errors as well, just as it should also reproduce the cases in which folk interpretation is correct. If we are modeling human cognitive capacities, we must aim for descriptive accuracy, even where we are normatively impaired.

Here is one possible response to the problem of adjustment. Currie & Ravenscroft (2002, pp. 11-12) draw a distinction between *perception-like* and *propositional* imaginings. Perception-like imaginings are those that have a visual character, or are like visual experience in certain respects. Propositional imaginings, on the other hand, are those that have the character of the propositional attitudes. In propositional states we are able to represent some categories abstractly. For example, if I am asked to *think* of triangles, I may not think of any particular kind of triangle (isosceles, for instance), but

may instead represent triangularity as such, in abstraction from any particular way it might be instantiated. This is distinct from seeing a triangle: every triangle seen must be isosceles, or scalene, etc. If I engage in perception-like imagining, this aspect of visual character ought to be preserved. A perception-like imagining of a triangle should have the content that there is a triangle of a certain determinate kind. Propositional imagining that there is a triangle, though, does not necessarily carry such a commitment. It should be possible to imagine someone holding a triangle while not imagining them holding any particular kind of triangle. As long as what you're imagining them holding doesn't have the properties of, say, a square, this representation will be useful in predicting someone who is holding a particular kind of triangle (since every actual triangle, as opposed to every propositionally imagined triangle, is a triangle of some particular kind or another). Propositional imagining, like belief and other non-perceptual ways of mentally representing the world, can omit certain determinate details in ways that perception-like imagining cannot.

Perhaps this insight provides the key to solving the problems of adjustment and inferential promiscuity. As I have sketched it, Gordon's approach is to adjust the imagined perceptual properties of the distal object until it generates the appropriate desires in me; that is, until it has the property of being desirable to me. Since everything that is good, or desirable, is so in virtue of some subjacent properties it has, adjusting those subjacent properties seems like the natural move to induce me to desire an object. That is, I have presented mirroring as a process of generating attitudes on the basis of largely perception-like imaginings. But an alternative would be not to alter the subjacent properties (the descriptive properties of a piece of music, for example), but to represent

the object directly as having the relevant supervenient property. That is, imagine not that the acoustic properties of the music have changed, but rather that the music sounds good, without further specifying the properties in virtue of which it sounds good. This is a more abstract state to imagine than the ones we have so far considered. It is plausible to view this as a case of propositional, rather than perception-like, imagining. It involves attributing 'goodness' without specifying the properties in virtue of which goodness is instantiated in this case. On the supposition that the music sounds good, I will want to hear it. This is the target attitude we were trying to mirror, and it does not appear to carry with it the unwanted attitudes generated by changing the descriptive properties of the music. So using our ability to represent abstract states can solve the problem of adjustment without falling prey to the problem of inferential promiscuity.

But this approach faces three immediate difficulties. First, it's not clear that we can imagine something's sounding good despite having properties that we take to be indicative of something that sounds unpleasant. Here we have a clash between what we perceive about the properties of the music, and the content of the propositional imagining. While perception might induce us to dislike the music, we are attempting to imagine that it sounds good. Which attitude will win out? The problem again is that mirroring engages our own dispositions and preferences, and these *prima facie* incline us to take something with these descriptive properties to sound unpleasant.

Second, even if we can imagine the right situation, we do not get the correct ascriptions. What I am imagining in this case is that something sounds good, but there is no guarantee that my conception of good music matches the conception of good music that my subject has. In the case we have been considering, our conceptions diverge by

assumption. So when I imagine that the music she is listening to sounds good, I am applying my own standards of musical aesthetics. The inferences and judgments that I tend to draw from the fact that something is a good piece of music differ from those of my subject, whose judgments are regulated by different standards of goodness. So, for example, where I may judge that a good piece of music is one that exemplifies post-bop modal jazz, she does not. Or I may judge that a good piece of music is one that is similar to *Kind of Blue*. This just raises the problem of inferential promiscuity once more.

Third, it isn't clear that imagining changes to the evaluative properties of distal objects is even permitted, given the Radical restriction. When I propositionally imagine that something is good, or desirable, or disgusting, how is this to be distinguished from imagining that I approve of, desire, or am disgusted by that object? Yet these are all instances of imagining that I have a certain mental state. The response strategy being considered asks that I imagine something to be disgusting, *simpliciter*. This ignores the fact that something's being disgusting is a matter of its affecting a thinking subject in a certain way. Being disgusting supervenes on various non-evaluative properties that I can perceive, and imagine altering. But what is the content of my propositionally imagining that something is disgusting? It's unclear whether there is anything *to* imagining X's being disgusting besides imagining that I have the relevant attitude or disposition to form the attitude of disgust towards X (at least if we are assuming that X's being disgusting is the only change made to X, as we are here).

Radical Simulationism builds folk interpretations out of our dispositions to form attitudes under imagined circumstances. But to change the circumstances themselves is not sufficient, since we carry with us our actual ways of categorizing the world. How to

change these dispositions without appealing to the concept of an inferential disposition, or some other mentalistic concept, is a question that the theory lacks resources to answer.

5. Perspective-taking and impairments of attribution

Feeding imagined inputs into our own beliefs and desires seems unlikely to generate correct attributions. In this section, I will critically assess the empirical evidence suggesting that mentalizing is not underwritten by the kind of architecture Radical Simulationists posit. Two kinds of study are especially relevant here: those involving visual perspective-taking tasks, and those involving imaginative transformation tasks. Both tasks involve using imaginative capacities to modify one's perspective on the actual world, modify some aspect of the perceived world, and draw conclusions about the consequences of these modifications. I will review some studies that raise *prima facie* problems for Radical Simulationists and consider some responses they might offer to the evidence.

Autistic subjects suffer from deficits in attributing mental states to others, yet are capable of passing some visual perspective-taking tasks. Hobson (1984) used a 'hide and seek' paradigm, in which autistic children were shown a model room containing several walls, and asked to place one pipe-cleaner doll so that it could 'hide' from either one or two other dolls that were 'seeking' it. Autistics perform this task as well as cognitively matched normal children. In another task, subjects must arrange a doll either so that it can 'see' a particular side of a cube, or so that two dolls can 'see' the same or different views of the cube simultaneously. Again, autistic children show a high degree of success.[12]

---

[12] Flavell, Everett, Croft, & Flavell (1981) distinguish Level 1 and Level 2 perspective tasks. Level 1 tasks require only that a subject be aware that others may not be able to see an object that they themselves can

Success requires understanding, or at least being able to pretend, that a crude doll figure can 'see', and that what someone sees depends systematically on where they are and what is in front of their eyes. Correctly figuring out the relationship between location and line of sight appears to rely on visuospatial imagination, which suggests that this faculty is not impaired in autism.

Reed & Peterson (1990) directly compared the performance of autistic children on visual and cognitive perspective-taking tasks. The visual perspective-taking tasks were a hide-and-seek task and a 'show me' task in which the child had to turn an object so that the experimenter could see a particular part of the object (e.g., the back side of a toy tow truck). The cognitive perspective-taking tasks were versions of the Sally-Anne false belief task. Ninety-two percent of autistic subjects passed both visual perspective taking tasks, while 23% passed the less demanding cognitive perspective taking task, and only 15% passed the most demanding cognitive task. And although autistic children are capable of taking on another person's visual perspective (based only on gaze) when asked to do so, they do not track gaze when unprompted, even when gaze is combined with overt clues to attention, such as facial expressions of interest (Leekam, Baron-Cohen, Perrett, Milders, & Brown, 1997). Inability to take on another's visual perspective does not seem to be a central deficit in autism.

Autistic children are also capable of following along with pretend transformations and predicting the outcome of these transformations (Kavanaugh & Harris, 1994, Experiment 3). Experimenters first demonstrated a pretend action, such as shaking a

---

see, if something is blocking or occluding the object. Level 2 tasks require the understanding that others may have a different view of an object from their perspective than the subject herself has. Note that Hobson (1984) employs a Level 2 perspective taking task, which is generally agreed to be more difficult than a Level 1 task.

closed can of talcum powder above a toy cat. The children were then asked to point to a drawing that depicted how the cat looked after the action from a set containing a drawing of the cat with powder on it, one of the cat as it was actually seen, and one of the cat having undergone an irrelevant transformation. The majority of autistic children scored better than chance at this task, lending further support to the notion that at least some autistics have intact capacities for visually based imagination, including the capacity to draw inferences from imagined scenarios.

First-person studies of patients with Asperger's syndrome also suggest that in this population, visual imagination and imagery are intact. Hurlburt, Happé, & Frith (1994) asked three men with Asperger's syndrome to periodically (when signaled with a beeper) make diary entries describing the thoughts and ideas they were experiencing at that moment. All subjects reported a great deal of visual imagery, much of which was not merely copied from their ongoing perceptions, but involved novel recombination of elements under endogenous control. Although Asperger's syndrome is associated with less severe deficits in mentalizing than autism, this indicates that rich mental imagery may co-occur even with mild deficits in attribution. Frith & Happé (1999) review additional first-person reporting by individuals with Asperger's syndrome and autism. These subjects also tend to report extensive, vivid, and complex visual imagery (although, interestingly, it is not always clear whether they understood the difference between *seeing* something and having an *image* of it). Finally, Harris & Leevers (2000) note that Temple Grandin, a very intelligent adult woman with autism, reports extensive use of mental imagery, including dynamic use of spatial images, in her work as a designer of equipment for the movement and management of cattle. This ability to form

rich visual images does not appear to have improved her ability to grasp the emotions of others, however.

Other studies sample outside of the autistic spectrum population. Congenitally blind children, while delayed in appreciation of false belief, are nevertheless able to pass perceptual perspective-taking tasks. Peterson, Peterson, & Webb (2000) tested 23 blind children, in groups of 6, 8, and 12 years of age. The majority of these children, at all age levels, passed a modified perspective taking task, despite having severe visual impairments. The youngest children, however, displayed difficulty with a variety of false belief tasks. Mentalizing skills were somewhat better in the middle age group, and near-ceiling performance on most tasks was observed by age 12. So mental state attributions appear separable from appreciation of others' perspective in non-autistic subjects as well.

As I have noted, however, these studies are not decisive.[13] Radical Simulationists might explain the existence of intact visual imagination with deficient mentalizing in at least two ways. One possibility is an inability to re-route imagined scenarios into belief and desire formation mechanisms. Autistics might be able to *create* imagined scenarios in Stage 1 of mirroring, but unable to use those scenarios to generate "pretend" mental states in Stage 2. On this account, autistics suffer from a disconnection of their intact visual imagination from their attitude generation mechanisms. A second possibility is that the creation of temporary, indexed memory workspaces might be impaired in autism. This might be the case if there are specific working memory deficits in autism, as some

---

[13] I should stress that it isn't a flaw in the studies themselves that they do not directly address the predictions of Radical Simulationism. Typically, these investigators have been interested in discovering the cognitive profile of autistic (or otherwise developmentally disabled) subjects, an entirely worthwhile goal in itself.

researchers have claimed (Reed, 2002).[14] In either one of these cases, visual imagination might be largely normal, but the products of visual imagination could not be used to generate mental state attributions.[15] The availability of these alternative explanations suggests that these studies are not necessarily strong evidence against Radical Simulationism.

The issue is complicated by the fact that some studies accord with Radical Simulationism's predictions. For instance, autistic subjects have been found to perform poorly on *difficult* visual perspective-taking tasks. Reed (2002) used a series of increasingly complex 'hide and seek' games to examine how well autistic subjects could take account of multiple perspectives at a time. Nearly all normal controls and subjects with mental retardation were able to hide the figure successfully when the task involved taking account of six distinct 'seeking' perspectives; by contrast, only 56% of the autistic group were able to do so. Yirmiya, Sigman, & Zacks (1994) found that even high-functioning autistic children had difficulty in a 'show me' task that required producing one particular perspective on a display containing multiple dissimilar objects. They note that autistic children are capable of passing serial arrangement tasks and simple visual perspective tasks, and suggest that their difficulties with more complex tasks may stem from disorders of executive function.[16] The main point about both studies is that neither

---

[14] However, for doubts about whether there are specific working memory deficits in autism, see Ozonoff & Strayer (2001).

[15] One limitation of these explanations is that they apply mainly to studies involving autistic subjects. A different explanation would be necessary for the delayed mentalizing of congenitally blind children, for example.

[16] Many theorists have claimed that deficits of executive function (EF) are at the heart of the range of disorders displayed in autism, e.g., reduced spontaneous pretend play, repetitive behaviors, and deficient mentalistic and linguistic skills. A full discussion of EF disorders is outside the scope of this paper, since the focus here is not on explaining autism *per se* but on explaining the mechanisms that underlie our folk psychological competence. Evidence favoring an executive dysfunction account of autism can be found in

provides evidence of a general deficit in visuospatial imagination in autism. Rather, the authors propose more widespread, general cognitive deficits (in working memory and executive functioning) underlying the particular problems with complex perspective-taking tasks. Hence it is simply unclear whether these results support Radical Simulationism.

In a highly original study, Scott & Baron-Cohen (1996) provided evidence that autistic children possess *selective* deficits in visual imagination. They are as capable as normal controls at drawing real objects such as houses and men, but are massively impaired at drawing 'impossible' or 'unreal' houses and men. They *are* capable of picking out drawings of such objects, though, which suggests that they can comprehend, but not generate or assemble such representations, despite the fact that their ability to imagine real objects is effectively normal. In addition, they are deficient at producing drawings of unreal objects even when being given explicit instructions by experimenters on what features to add or alter. One explanation is that simple visual memory for observed objects is responsible for their performance with real entities, and impaired visual imagination is responsible for their failures to construct pictures of unreal entities, which all require transforming and recombining elements of real, observed objects.

Leevers & Harris (1998), however, have challenged these findings. They used a picture completion task in which autistic and normal children were given a drawing of a house minus a door and a man minus a head, and asked to complete the figures in either a 'real' or an 'impossible' fashion (e.g., drawing a door on the roof, or two heads on the man). Here there were no significant differences between autistic and normal children.

Harris (1993), Hill (2004), Ozonoff, Pennington, & Rogers (1991), Pennington & Ozonoff (1996), and Russell (1997).

Both groups were more successful at drawing real completions than unreal ones, but autistic children did not show a special deficit for completing unreal figures. Leevers & Harris conjecture that either pragmatic or executive function deficits may be to blame for the failures Scott & Baron-Cohen observed. The verbal instructions might have been confusing to the subjects, or they might have been unable to plan and execute an appropriate response. Given that the precise explanation for these conflicting results is not yet clear, we might suspend judgment as to whether there is a special deficit in imagining unreal entities in autism.

6. Imagination and imagery

A shortcoming of much of the research reviewed here is that it demonstrates only that individuals with mentalizing deficits appear to have intact visual perspective-taking skills. But as I've noted, there are several candidate explanations of this fact. The possible presence of executive dysfunction, working memory disorders, and other impairments in autistic subjects makes it especially difficult to rule out possible alternatives. So these studies show, at best, that intact visual imagination is not *sufficient* for intact mentalizing. The case against Radical Simulationism could be strengthened considerably by demonstration of the pattern complementary to that observed so far: a specific deficit in visual imagination *without* deficits in mentalizing. This is a more significant pattern of evidence to look for, since the existence of mentalizing deficits with intact visual imagination is consistent with the claim that visual imagination is necessary for mirroring.

This suggests a more direct route to test the architecture proposed by Radical Simulationism. Gordon frequently presents the ability to shift spatiotemporal location on an actual scene as a crucial component of mirroring. Further, he suggests that imagining various changes to the perceived properties of objects and events may be necessary to generate appropriate attitudes. These all appear to be, in Currie & Ravenscroft's (2002) sense, perception-like imaginings. If we adopt the hypothesis that such imaginings essentially involve mental imagery, the Radical Simulationist thesis could be tested by looking at subjects who have deficits specific to mental imagery.

The claim that perception-like imagination should be understood as a faculty that essentially involves forming and using mental images enjoys some *prima facie* plausibility. Currie (1995) and Currie & Ravenscroft (2002) argue that mental imagery is the simulation of vision, in the sense that it re-uses much of the same cognitive machinery that visual perception uses.[17] Imagery itself has perception-like characteristics; we spontaneously describe imagery in terms of visual perception, images can sometimes be confused with perceptions, and people may find it difficult to recall whether they saw something or merely imaged it (Currie & Ravenscroft, 2002, pp. 72-3). This similarity in character should be expected if imagery and perception re-use many of the same cognitive systems.

Recently, Kind (2001) has ambitiously argued for an Imagery Model of imagination according to which: (i) imagery plays an essential role in imagination; but (ii) images do not serve to individuate imaginings. Rather, (iii) images are the

---

[17] Currie & Ravenscroft (2002) further argue that mental imagery is simulation in that it is 'cognitively conservative': beyond the input stage, it does not employ information that is not drawn on by vision (or perception generally) itself. The input to mental imagery comes from higher cognitive centers, which employ concepts and other representations that are not part of the normal input to perception. But this does not by itself prevent imagery from simulating perception after the input stage.

representational vehicles of acts of imagination. I will not adopt or defend all three points of the Imagery Model here. I am concerned only with (i), the essentialist claim. We can distinguish several grades of essentialism:

1. Global Imagery Essentialism: Every act of imagination essentially involves mental imagery.

2. Strong Local Imagery Essentialism: Every act of imagination used in mental mirroring essentially involves mental imagery.

3. Weak Local Imagery Essentialism: Some acts of imagination used in mental mirroring essentially involve mental imagery.

Kind (2001) does not employ this taxonomy, but she defends Global Imagery Essentialism, and maintains that mental images are the representational vehicles for episodes of imagination. In light of the distinction between perception-like and propositional imaginings, this may be too strong. Some mental acts that are appropriately thought of as imaginative might lack a specifically sensory character. Here I propose adopting only Local Imagery Essentialism. But it may be that some acts of imagination that are not involved in mirroring are nonimagistic, and possibly not all mirroring involves images. So I will assume only Weak Local Imagery Essentialism, and consider cases where the use of imagery is most obviously presupposed by Gordon.

Imagery can occur in many sense modalities, as well as in motor domains, but visuospatial imagery is especially central to mirroring. Consider putting myself in another's place, as when I try to comprehend why my hiking companion suddenly bolts away after taking a turn in the trail around which I cannot see. This may crucially involve spatial imagery, particularly imagining perceiving the world from a different spatial

location and orientation. To carry out this task I need to form an image of the world from an orientation other than the one that I occupy. It may also involve visual components, such as imagining that some aspects of the world appear differently from my companion's perspective. This may involve elaboration or deletion of elements of the actual perceived scene. Perhaps my companion can see a venomous snake around the next bend, which I would have to form an image of in order to account for her appearance of fear.

There are reported cases in which mental imagery is impaired or deficient, whether congenitally or due to some neurological injury. Stangalino, Semenza, and Mondini (1995) demonstrated the existence of deficits in visual mental imagery that correlate with damage to left posterior cortical regions. Eighty-two subjects with unilateral lesions were tested on a battery of tasks designed to separate imagery deficits from memory and praxic deficits.[18] Nineteen of the original 82 subjects displayed selective impairments of mental imagery. Of these, 15 had left hemispheric lesions, all but one of which were confined to the posterior regions. Interestingly, none of these patients ever complained of imagery loss. This suggests that image generation can be specifically impaired by relatively localized lesions, and may occur without the subject's consciousness of any deficit.[19]

---

[18] The imagery-engaging tasks were question-answering (e.g., 'What color are your city's taxis?'), describing an object from memory, drawing an object without a model, and map drawing. The tasks used to rule out alternative explanations of poor performance on the imagery tasks were visual search, object recognition, matching a name to an object, drawing from a model, and describing an object's functional properties.

[19] Functional imaging studies suggest more bilateral involvement in visual imagery than this lesion study. Mellet, Petit, Mazoyer, Denis, & Tzourio (1998) suggest that left hemisphere involvement in image generation may be limited to forming images of items that are lexicalized, while right hemisphere involvement may be required as images become more complex.

Deficits in spatial mental imagery are also attested. In a well-known study, Bisiach & Luzzatti (1978) showed that patients who suffered damage to right posterior cortical regions, particularly parietal areas, neglect the left side of visual space. In addition, they appear to neglect the left side of imagined space. When asked to imagine standing at one end of a familiar public place and describe what they would see, they describe landmarks and buildings located on the right side of the scene, while failing to describe the left half. Coslett (1997) describes two patients who show a double dissociation between visual and imaginal neglect. Patient M.N., who suffered an ischemic infarction affecting the right frontoparietal areas, performed normally on two tests of visual attention and attention shifting, but performed poorly on an imagine-and-describe task akin to Bisiach & Luzzatti's. This suggests that spatial imagery can be deficient without a patient showing the full neglect syndrome.

Trojano & Grossi (1994) provide an extensive review of subjects with imagery deficits. These deficits are most typically associated with visual recognition and object naming disorders (e.g., associative visual agnosia). However, there is little consistent association between imagery deficits and specific kinds of perceptual disorders. For example, patients with imagery deficits may have visual agnosia, prosopagnosia, color agnosia, or optic aphasia, among other disorders. The trend towards such associations is strong, but not very specific. This trend might be anticipated if mental imagery and visual perception utilize some of the same neural architecture (Farah, 2000a, 2000b; Kosslyn, 1994).[20]

---

[20] However, there is disagreement about whether it is primary visual cortical areas or higher visual areas that are used during mental imagery. See Mellet et al. (1998) and Bartolomeo (2002) for arguments that mainly higher areas, and not primary visual regions are activated during imagery.

If mental mirroring depends on the ability to form and manipulate mental images, we would predict that deficits in mental imagery would be associated with deficits in mental state attribution. Observing such a co-occurrence would be strong evidence for Radical Simulationism. Not finding it, on the other hand, might not be decisive. It would suggest one of the following:

(i)   Attribution does not involve visuospatial imagery, and thus doesn't involve visual imagination in the present sense;

(ii)   Attribution does normally involve imagery, but subjects whose imagery is impaired have learned alternative strategies for interpretation that don't involve imaginative projection; or

(iii)   Imagination should be interpreted in some way other than as involving mental imagery (i.e., no version of Imagery Essentialism is true).

Possibility (ii) is real enough, and should be guarded against, perhaps by using subjects whose imagery deficit is relatively recent, to eliminate possible effects of compensatory learning. If we adopt possibility (iii), though, Radical Simulationists owe an account of just what faculties *are* involved in imaginative projection, if their account is to be complete and empirically testable.

If imagery-impaired subjects do in fact perform worse in visual perspective-taking tasks that involve mental state attribution, this supports the claim that imagery is involved in mental state attribution. This would be a significant validation for Simulation theorists. On the other hand, if patients with demonstrated visuospatial imagery deficits are able to pass tasks that require mentalizing, this would provide evidence that imagery, and by hypothesis visual imagination, is not necessary to folk psychological competence.

7. Conclusions

I have argued for two claims: first, that Gordon's account of mirroring faces problems in explaining how we might come to attribute to our subject mental states that diverge from our own; and, second, that if Gordon's account is correct, on some plausible assumptions we would expect to find deficits in visual imagination and mental imagery co-occurring with deficits in mental state attribution. The existing evidence, mostly from autistic subjects, shows that there can be impaired attribution with intact visual imagination. This is not decisive, however, given that (i) Simulationists don't need to claim that visual imagination is sufficient for mirroring; and (ii) autistic subjects often have other deficits that are perhaps sufficient to explain their performance. I have suggested that perspective-taking tasks involving patients with impairments of mental imagery may strengthen the case against Radical Simulationism. Such empirical tests would help to clarify the cognitive architecture underlying folk attribution.

The objections I have leveled against simulationism center on the requirement that an interpreter actually have the resources to cause herself to enter a mental state that is type-identical to her subject's. That this requirement seems too stringent, particularly given the Radical restriction. *De gustibus non disputandam*, perhaps; but we can surely interpret people with whom we disagree in tastes, preferences, emotions, and beliefs. Models according to which mental state attribution is a kind of theoretical inference rather than an empathetic sharing of attitudes have the advantage of making this explicable, whatever their other shortcomings might be.

Finally, I have only been criticizing the half of Radical Simulationism devoted to explaining interpretation in terms of mental mirroring. I have said nothing about the explanation of behavioral predictions in terms of mental simulation. In my view, that is a separate issue, and one that has been extensively discussed elsewhere in the literature. Mental mirroring has received much less attention. This is regrettable. Mirroring is a necessary precursor to prediction, since predictions require some initial attitudes to work with. This discussion represents one step towards drawing out empirical consequences of a simulation-based account of mental state attribution.

References

Baron-Cohen, S. 1995. *Mindblindness*. Cambridge: MIT Press.

Bartolomeo, P. 2002. The relationship between visual perception and visual mental imagery: A reappraisal of the neuropsychological evidence. *Cortex*, 38, 357–378.

Bartsch, K., & Wellman, H. M. 1995. *Children Talk about the Mind*. New York: Oxford University Press.

Bisiach, E., & Luzzatti, C. 1978. Unilateral neglect of representational space. *Cortex*, 14, 129–133.

Carruthers, P., & Smith, P. (Eds.). 1996. *Theories of Theories of Mind*. Cambridge: Cambridge University Press.

Coslett, H. B. 1997. Neglect in vision and visual imagery: A double dissociation. *Brain*, 120, 1163–1171.

Currie, G. 1995. Visual imagery as the simulation of vision. *Mind and Language*, 10, 25–44.

Currie, G., and Ravenscroft, I. 2002. *Recreative Minds: Imagination in Philosophy and Psychology*. Oxford: Oxford University Press.

Davies, M., & Stone, T. (Eds.). 1995a. *Folk psychology: The Theory of Mind Debate*. Oxford: Blackwell.

Davies, M., & Stone, T. (Eds.). 1995b. *Mental simulation: Evaluations and Applications*. Oxford: Blackwell.

Emery, N. J. 2000. The eyes have it: The neuroethology, function, and evolution of social gaze. *Neuroscience and Biobehavioral Reviews*, 24, 581–604.

Farah, M. J. 2000a. *The Cognitive Neuroscience of Vision*. Oxford: Blackwell.

Farah, M. J. 2000b. The neural bases of mental imagery. In M. S. Gazzaniga (Ed.), *The New Cognitive Neurosciences* (2nd ed., pp. 965–974). Cambridge: MIT Press.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. 1981. Young children's knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. *Developmental Psychology*, 17, 99–103.

Frith, U., & Happé, F. 1999. Theory of mind and self-consciousness: What is it like to be autistic? *Mind and Language*, 14, 1–22.

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

Goldman, A. 1993. The psychology of folk psychology. *Behavioral and Brain Sciences*, 16, 16–28.

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

Gordon, R. M. 1995a. Folk psychology as simulation. In M. Davies & T. Stone (Eds.), *Folk psychology: The Theory of Mind Debate*. Oxford: Blackwell.

Gordon, R. M. 1995b. The simulation theory: Objections and misconceptions. In M. Davies & T. Stone (Eds.), *Folk psychology: The Theory of Mind Debate*. Oxford: Blackwell.

Gordon, R. M. 1995c. Simulation without introspection or inference from me to you. In M. Davies & T. Stone (Eds.), *Mental Simulation: Evaluations and Applications*. Oxford: Blackwell.

Gordon, R. M. 1996a. Sympathy, simulation, and the impartial spectator. In L. May, M. Friedman, & A. Clark (Eds.), *Mind and Morals: Essays on Ethics and Cognitive Science*. Cambridge: MIT Press.

Gordon, R. M. 1996b. 'Radical' simulationism. In P. Carruthers & P. Smith (Eds.), *Theories of Theories of Mind*. Cambridge: Cambridge University Press.

Gordon, R. M., & Barker, J. A. 1994. Autism and the 'theory of mind' debate. In G. Graham & G. L. Stephens (Eds.), *Philosophical Psychopathology* (pp. 163–181). Cambridge: MIT Press.

Harris, P. 1993. Pretending and planning. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding Other Minds: Perspectives from Autism* (pp. 228–246). Oxford: Oxford University Press.

Harris, P. L., & Leevers, H. J. 2000. Pretending, imagery and self-awareness in autism. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience* (pp. 182–202). Oxford: Oxford University Press.

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

Hill, E. L. 2004. Evaluating the theory of executive dysfunction in autism. *Developmental Review*, 24, 189–233.

Hobson, R. P. 1984. Early childhood autism and the question of egocentrism. *Journal of Autism and Developmental Disorders*, 14, 85–104.

Hurlburt, R. T., Happé, F. G. E., & Frith, U. 1994. Sampling the form of inner experience in three adults with Asperger syndrome. *Psychological Medicine*, 24, 385–395.

Kavanaugh, R. D., & Harris, P. L. 1994. Imagining the outcome of pretend transformations: Assessing the competence of normal children and children with autism. *Developmental Psychology*, 30, 847–854.

Kind, A. 2001. Putting the image back in imagination. *Philosophy and Phenomenological Research*, 62, 85–109.

Kosslyn, S. M. 1994. *Image and Brain*. Cambridge: MIT Press.

Langton, S. R. H., Watt, R. J., & Bruce, V. 2000. Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, 4, 50–59.

Leekam, S., Baron-Cohen, S., Perrett, D., Milders, M., & Brown, S. 1997. Eye-direction detection: A dissociation between geometric and joint attention skills in autism. *British Journal of Developmental Psychology*, 15, 77–95.

Leevers, H. J., & Harris, P. L. 1998. Drawing impossible entities: A measure of the imagination in children with autism, children with learning disabilities, and normal 4-year olds. *Journal of Child Psychology, Psychiatry, and Allied Disciplines*, 39, 399–410.

Mellet, E., Petit, L., Mazoyer, B., Denis, M., & Tzourio, N. 1998. Reopening the mental imagery debate: Lessons from functional anatomy. *Neuroimage*, 8, 129–139.

Nichols, S., & Stich, S. 2000. A cognitive theory of pretense. *Cognition*, 74, 115–147.

Ozonoff, S., & Strayer, D. L. 2001. Further evidence of intact working memory in autism. *Journal of Autism and Developmental Disorders*, 31, 257–263.

Ozonoff, S., Pennington, B. F., & Rogers, S. J. 1991. Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology, Psychiatry, and Allied Disciplines*, 32, 1081–1105.

Pennington, B. F., & Ozonoff, S. 1996. Executive functions and developmental psychopathology. *Journal of Child Psychology, Psychiatry, and Allied Disciplines*, 37, 51–87.

Peterson, C. C., Peterson, J. L., & Webb, J. 2000. Factors influencing the development of a theory of mind in blind children. *British Journal of Developmental Psychology*, 18, 431–447.

Reed, T. 2002. Visual perspective taking as a measure of working memory in participants with autism. *Journal of Developmental and Physical Disabilities*, 14, 63–76.

Reed, T., & Peterson, C. 1990. A comparative study of autistic subjects' performance at two levels of visual and cognitive perspective taking. *Journal of Autism and Developmental Disorders*, 20, 555–567.

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

Russell, J. (Ed.). 1997. *Autism as an Executive Disorder*. Oxford: Oxford University Press.

Scott, F. J., & Baron-Cohen, S. 1996. Imagining real and unreal things: Evidence of a dissociation in autism. *Journal of Cognitive Neuroscience*, 8, 371–382.

Stangalino, C., Semenza, C., & Mondini, S. 1995. Generating visual mental images: Deficit after brain damage. *Neuropsychologia*, 33, 1473–1483.

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

Stich, S., Nichols, S., & Leslie, A. 1996. Varieties of offline simulation. In P. Carruthers & P. Smith (Eds.), *Theories of Theories of Mind* (pp. 39–74). Cambridge: Cambridge University Press.

Trojano, L., & Grossi, D. 1994. A critical review of mental imagery deficits. *Brain and Cognition*, 24, 213–243.

Wellman, H. M. 1990. *The Child's Theory of Mind*. Cambridge: MIT Press.

Yirmiya, N., Sigman, M., & Zacks, D. 1994. Perceptual perspective-taking and seriation abilities in high-functioning children with autism. *Development and Psychopathology*, 6, 263–272.