Taking Solipsism Seriously:
Nonhuman Animals and Meta-cognitive Theories of Consciousness[1]

> What else is it that should trace the insuperable line? Is it the faculty of reason, or, perhaps, the faculty of discourse? But a full-grown horse or dog is beyond comparison a more rational as well as a more conversable animal, than an infant of a day, or a week, or even a month, old. But suppose the case were otherwise, what would it avail? the question is not, Can they *reason*? nor, Can they *talk*? but, Can they *suffer*?
>
> - Jeremy Bentham

Bentham's question is meant to be rhetorical, but is not always so taken. Infamously, Cartesians have held that nonhuman animals do not suffer, or have any conscious experiences. More recently, it has been argued that, perhaps with the exception of the great apes, nonhuman animals do not suffer "in the sense that makes their experiences an appropriate object of concern."[2] Unlike its Cartesian predecessors, this more recent argument does not rely upon the assumption of mind-body dualism.

Although the bold conclusion this argument contradicts common sense, it would be a mistake to dismiss it out of hand. In fact, each of the three main components of the argument has a considerable prima facie plausibility, and is embraced by a wide variety of philosophers.[3] The first element of the argument is the Higher-Order-Representation (HOR) theory of consciousness, which comes in two varieties - the Higher-Order-Perception (HOP) sort and the Higher-Order-Thought (HOT) version. On each of these theories, for a mental state to be conscious, in one important sense of 'conscious', is for the agent to have a higher-order mental representation of that state. The perception-version of the theory holds that these higher-order representations are, in some sense, perception-like, whereas the thought-version holds that they are thoughts. Defending a more precise characterization of the difference between these versions of the HOR theory is somewhat difficult and controversial.[4] In any case, HOR theories have a good deal to recommend them (as I shall briefly argue below), and have been given an impressive and

extensive defense by such philosophers as David Armstrong, Paul Churchland, William James, John Locke, William Lycan, and David Rosenthal. Norton Nelkin has also argued for a view that is arguably a HOR theory, and Daniel Dennett has remarked that the HOT theory is "very close kin" to his own view of consciousness.[5]

The second element of the argument is the idea that the only sort of consciousness that confers a being with direct moral standing is the sort of consciousness picked out by HOR theories. While this piece of the argument has not been given such an extensive defense, it has considerable prima facie plausibility. One way of seeing the plausibility of this claim is to think about the kinds of examples that are used to illustrate the sense of 'consciousness' that HOR theories are meant to analyze. A stock example is the case of someone who is driving an automobile, but whose conscious attention is entirely focused on a conversation she is having with her companion. When she "comes to" and returns her attention to the task of driving, she is alarmed to note that she does not have the faintest idea what she has been doing or seeing for the past several minutes. Nonetheless, there is sense in which she must have been seeing, or she would not have artfully avoided the pothole, stopped at the traffic light, etc. Typically, the HOR theorist argues that the experiences of the road she has while engrossed in conversation are nonconscious experiences in the sense isolated by the HOR theory - they are mental states she was not representing with the appropriate HOR. So if someone never was conscious in the sense picked out by the HOR theorist, it would seem, their entire life would be a string of nonconscious experiences, analogous to the experiences of the road had by the imagined driver before she "comes to." If we subtract those conscious experiences, and leave only the nonconscious ones, we are left with a creature that is unaware of its own mental states, including, presumably, its own pleasures and pains. For many, it is a very small step from this result to the conclusion that an entity without HOR-consciousness lacks direct moral standing, and perhaps reasonably enough; ask yourself whether anything could really *matter* to you if you were never conscious in the HOR theorist's sense.[6]

Descartes, at least, seems to have endorsed this element of the argument, and that, more recently, Peter Carruthers, Joel Feinberg, and Peter Harrison, have defended it.[7] I suspect many other philosophers would accept it, insofar as they accept the HOR theorist's examples as paradigmatic of the sort of consciousness requiring a HOR. For example, many utilitarians might only give direct moral weight to pleasures or pains that were conscious in the sense of 'conscious' the HOR theories are intended to explicate.

Unsurprisingly, the third element of the argument is a deep scepticism about the claim that nonhuman animals ever have the appropriate HORs. Here the main objections have focused on the HOT theory rather than the HOP theory, and I shall in what follows focus on the HOT theory as well. Some philosophers, including Donald Davidson, R.G. Frey, Norman Malcolm, Wilfrid Sellars, Stephen Stich, and perhaps Wittgenstein, are staunchly opposed to the idea that there are thought-balloons over the heads of nonhuman animals at all, much less *meta*-thought-balloons.[8] Other philosophers, and cognitive ethologists, allow that nonhuman animals have some thoughts, but deny that they have the conceptual resources needed for higher-order thoughts, on roughly the following grounds: if they had the conceptual resources for HOTs then they both could and would engage in deceit, but they do not engage in deceit, so they lack the requisite conceptual resources. Jonathan Bennett, Peter Carruthers, Daniel Dennett, Dorothy Cheney and Robert Seyfarth, Richard Byrne and Andrew Whiten, and David Premack all seem to be sympathetic to this general line of argument, though not all of them would endorse the version of it I consider here. [9]

At this point, it should not be too difficult to see why someone who accepted all three of these strands of thought would answer Bentham's rhetorical question with the reply, "No - in the sense relevant to their moral standing they do *not* suffer." Suffering in the sense that is necessary for direct moral standing presupposes consciousness in the HOR theorists' sense, and nonhuman animals do not have such consciousness because they do not have the appropriate HORs. While Carruthers has been uniquely vocal in

putting all three of these pieces together and pressing the argument, it should be clear that a wide range of philosophers would be quite tempted to accept each of its premises, thereby having its conclusion forced upon them. Since many, though not all, of these philosophers would, I suspect, reject, or at least prefer to reject, the conclusion that nonhuman animals lack direct moral standing, the argument represents a real, if largely unnoticed, problem for a wide range of philosophers.

In this paper I grant for the sake of argument the first two pieces of the argument, and focus on the third - the thesis that nonhuman animals lack the appropriate HOTs. Whether they do is ultimately an empirical matter. However, the putative evidence typically given for this thesis would count as good and sufficient evidence for it only if one takes a certain philosophical outlook. In particular, the case for thinking nonhuman animals lack the appropriate HOTs relies on the philosophical assumption that any being that can have thoughts about its own mental states must also be capable of having thoughts about the mental states of others. Discussion of whether nonhuman animals have HOTs invariably started, and ended, with the question with which Jonathan Bennett begins: "What would count as behavioral evidence for us that our animal has a thought about some *other* animal's mind?" (emphasis mine)[10] At one level, this single-minded focus on whether a creature has thoughts about the thoughts of others is understandable; it is probably easier to gather evidence and run experiments to determine whether a creature has thoughts about the mental states of others than it is to determine whether it has thoughts about its own mental states. However, at another level, this narrow focus represents an important philosophical prejudice. In particular, it overlooks the possibility that some nonhuman animals might be *unreflective solipsists*[11] - capable of having thoughts about their own mental states, but incapable of having thoughts about the mental states of others. Insofar as this is a real possibility, it is fallacious to move from evidence that nonhuman animals lack thoughts about the mental states of others to the conclusion that they have no higher-order thoughts whatsoever. Nonetheless, the prejudice

underlying this inference seems to have become a tacit orthodoxy. One of my main aims in this paper is to debunk this prejudice.

A second, and perhaps less pervasive prejudice that seems to be at work in the literature is a tendency to focus more on whether a creature can have thoughts about thoughts, rather than thoughts about desires, emotions and other conative states. Although this prejudice may not rely on any particular dubious philosophical assumptions, it has led some to conclude too quickly that nonhuman animals do not suffer in a morally salient sense. Interestingly, findings from developmental psychology offer some evidence against both of these prejudices.

If this diagnosis is correct, then it should be of interest both to advocates of the direct moral standing of nonhuman animals *and* for defenders of HOR theories of consciousness. In the former case, the interest of the present argument should be obvious enough. However, as is often the case, one philosopher's modus ponens is another's modus tollens. Where Carruthers and company argue from the plausibility of the HOT theory of consciousness to the denial of conscious mental states to nonhuman animals, other philosophers argue from the thesis that nonhuman animals do have conscious mental states to the denial of the HOT theory. Here, for example, is Fred Dretske:

> There are, however, two objections to HOT theories that are, in my mind, decisive. First, as developmental studies show, children only begin to gain a conception of thought and experience (as ways of representing the world that may or may not be accurate) around their third year…It is hard to see, therefore, how, at this early age, they could have a higher-order thought of the requisite kind…If they are unable to hold higher order beliefs about lower order thoughts and experiences, are we to conclude, therefore, that none of their thoughts and experiences are conscious?…If that is a consequence of a HOT theory, it strikes me as very close to a *reductio* (it would *be* a *reductio* if we really knew, instead of merely having strong intuitions – that their experience was not fundamentally different)…The same should be said about animals.[12]

Dretske and Carruthers argue from the shared assumption that nonhuman animals (and young children) lack higher-order thoughts to very different conclusions. Whereas

Carruthers concludes that nonhuman animals lack direct moral standing, Dretske concludes that HOT theories must not capture the essence of consciousness. If the argument of the present paper is sound, the shared assumption from which they argue – that nonhuman animals (and young children) lack HOTs – is one that has been embraced prematurely. So the argument not only provides a defense of the moral standing of nonhuman animals, it also presents a defense of the HOT theory.[13]

Finally, I offer two caveats. First, for present purposes I put to one side the view those philosophers, like Sellars and Davidson, who argue that nonhuman animals lack any thoughts whatsoever; an examination of that view would go beyond the present scope. I shall throughout be taking for granted the common sense assumption that many nonhuman animals at least have some thoughts, and are hence capable of deploying some concepts. The issue, then, will be whether all their thoughts are focused on the world or whether some are aimed at some of their own mental states.[14] Second, my primary aim is to show the inadequacy of the usual grounds that are given for supposing that nonhuman animals lack HOTs. I shall not, therefore, try to give a positive defense of the thesis that any particular non-human animals do, in fact, have the relevant HOTs. Therefore, my main argument shall not directly assuage the doubts of anyone who approaches the question of whether many nonhuman animals have HOTs with a sort of free-floating scepticism about the conceptual sophistication of many nonhuman animals, even the more intelligent ones. Whether such doubts ultimately can adequately be assuaged is itself largely an empirical matter. The main point of the present argument is that our interpretation of the empirical data should not be biased by an unfounded ex ante philosophical prejudice against the possibility of unreflective solipsism.

I. THE HOT THEORY OF CONSCIOUSNESS

"Brain: n., An apparatus with which we think that we think."

-- Ambrose Bierce (*Cynic's Word Book*, 1906)

Before considering the sorts of arguments given for thinking that nonhuman animals lack higher-order-thoughts, it is worth pausing briefly over the HOT theory of consciousness itself, to see what seems plausible about it. David Rosenthal has provided the classic defense of that theory, so in this section I shall recount some of the main features of his defense. Rosenthal usefully contrasts the HOT theory with what he calls the "Cartesian view of mind." On the Cartesian view, consciousness is a necessary feature of a mental state; it gains its plausibility from the fact that consciousness "is so basic to the way we think about the mind that it can be tempting to suppose that no mental states exist that are not conscious states."[15] The trouble with this picture of the mind is that it does not allow us to explain what makes a mental state conscious in terms of a prior account of mentality, since the Cartesian picture builds consciousness into mentality. This threatens to make the "gulf that seems to separate mind and consciousness from the rest of reality" impossible to bridge.[16] Further, the Cartesian picture precludes our having any genuinely mental states that are not also conscious states, but this is implausible, particularly in the case of beliefs. We often seem to be aware of what someone is thinking when they are initially unaware of so thinking; intuitively, the person is having a non-conscious thought. Similarly, there is David Armstrong's famous case of the long-distance driver who is "on auto-pilot" – who in some sense must be aware of the cars that he deftly maneuvers around, but who sincerely reports not having been conscious of their presence. The HOT theory (much like Armstrong's HOP theory) is meant to provide an explanation of how this is possible – the driver may well have had the relevant beliefs, but they were nonconscious in virtue of there not having been the relevant HOTs about them. In the present context, it is worth noting that *if* what Rosenthal calls the Cartesian picture were

right, there could be no issue of nonhuman animals having non-conscious mental states; if they have any mental states, then on the Cartesian view those states must be conscious.

When compared with the Cartesian picture, the HOT theory has the virtues of allowing us both to explain the possibility and nature of non-conscious thoughts *and* to give an explanation of consciousness in terms of the mental. Since conscious mental states are simply states we are aware of being in, and our being aware of something is intuitively just a matter of having a thought of some sort about it, the HOT theory claims that a mental state's being conscious consists in one's "having a roughly contemporaneous thought that one is in that mental state." Rosenthal originally concluded his characterization of the theory by suggesting that, "since a mental state is conscious if it is accompanied by a suitable higher-order thought, we can explain a mental state's being conscious by hypothesizing that the mental state itself causes the higher-order thought to occur."[17] Rosenthal has since dropped the causal condition, and now requires only that one not be aware of having arrived at the HOT by inference.[18]

Finally, it is worth explaining how an initially seductive objection to the HOT theory is based on a misunderstanding. One might respond to the HOT theory by denying that they have anything like the number of HOTs that the theory seems to entail that they must have, given the seeming ubiquity of their conscious experiences. It simply does not seem to me, the objector might urge, that I have so many metacognitive thoughts – indeed, it seems to me that I hardly ever go in for such abstract thoughts. However, this appeal to phenomenology is perfectly compatible with the HOT theory – one's HOTs themselves would be conscious *only if* one had a tertiary thought about the HOTs in question. The thought at the top of our hierarchy of thoughts and meta-thoughts is itself always non-conscious, in that we are always not conscious of our having that highest level thought.[19] Since we typically do not have such tertiary thoughts, it is no surprise that we are usually unaware of the indeed ubiquitous HOTs that the theory posits to explain our everyday conscious experiences. On the HOT theory as Rosenthal articulates

it, those rare occasions upon which we do have the relevant tertiary thoughts are cases of introspection – cases of a conscious higher-order thought.

II. THE IMPORTANCE OF THE POSSIBILITIES OF UNREFLECTIVE SOLIPSISM AND OF A "SIMPLE DESIRE PSYCHOLOGY."

"Whosoever is delighted in solitude is either a wild beast, or a god."

-- Francis Bacon

Having seen why the HOT theory itself is plausible, I now turn to the question of why one would conclude that nonhuman animals lack the sorts of HOTs required for direct moral standing. A wide variety of philosophers and cognitive ethologists have argued that the apparent inability of nonhuman animals to practice genuine deceit tells against supposing that they have such HOTs, and it is this argument, and its unargued but highly questionable presuppositions, upon which I wish to focus. Since Peter Carruthers has provided the most forceful and explicit presentation of this sort of argument, so I will focus on his version of it. I should note, however, that the points I make against Carruthers, if sound, should have bite for any version of the argument.

Actually, Carruthers considers two arguments for the conclusion that nonhuman animals lack the sort of consciousness necessary for direct moral standing, one of which relies upon the HOT variation of the HOR theory, and one of which relies upon Carruthers' own theory of the mind. At the end of the day, Carruthers rejects this first argument on the grounds that his own theory of the mind is more plausible than the HOT theory. However, as has been argued elsewhere, Carruthers' attacks on the HOT theory, and his proposed alternative theory, face serious objections.[20] Therefore, in my view, the argument Carruthers develops but then sets aside is much more interesting than the one he actually endorses (it is, at any rate, very much of interest to anyone sympathetic to the HOT theory). Moreover, Carruthers does nonetheless make the striking claim that *if* the HOT theory were correct, then we should be very reluctant to attribute any conscious mental states to any nonhuman animals, except perhaps the great apes. Since my aim

here is to refute this *conditional* claim, a real issue remains between Carruthers and myself, in spite of his rejection of the HOT theory and the associated argument.

Carruthers (following Bennett, Dennett, and others) suggests that we should, in searching for nonhuman metacognition, focus on alleged cases of deceit, as genuine deceit is "clearest way in which an animal can manifest second-order beliefs..."[21] Deceit would, one must admit, be excellent evidence of a second-order belief, since to engage in deceit is intentionally to try to implant a false belief in another creature, and intentionally making this attempt would presuppose some understanding of the fact that this other creature can be made to have beliefs with particular contents. Carruthers allows that there is anecdotal evidence that such creatures engage in deceit, but that such anecdotal evidence is "always amenable to more neutral description, precisely because it is merely anecdotal." Here Carruthers gives the case of Donna and her dog Dean. Dean likes to walk and likes to sleep in Donna's chair. One day, Dean brings Donna his leash, but when she gets up to take him for a walk, he jumps into her chair. This might seem like good evidence of deceit, but we could just as easily say that Dean wanted both to walk and to lie in the chair; he aimed to satisfy the first end, but when, fortuitously, the opportunity to satisfy the second arose, he satisfied it instead. Alternatively, as is suggested by Daniel Dennett's remarks about a similar case, we might simply conclude that the dog "is a good behaviorist," and has conditioned Dona to stand up when he brings her his leash. Apparently, this account also does not require supposing that the dog genuinely engages in deceit and hence does not require supposing that she has the concept of a belief.[22] Following Lloyd-Morgan's canon, it is suggested that we should err on the side of more simple attributions.[23] Attributing a HOT will involve the attribution of a host of other thoughts, many of which it might seem ad hoc to attribute. So long as the evidence for deceit is anecdotal, resting on cases like the case of Don and Dean, we should resist the attribution of a HOT in favor of more simple explanations.[24]

The mere fact (if it is a fact) that we have good evidence of deception only in the case of great apes would not justify our assuming that other creatures are incapable of it; we might just not have looked hard enough. Presumably, the suggestion is that at this point we have looked hard enough to be fairly confident. Even granting this supposition, though, the argument looks to be invalid. All that would be shown if it were established that most nonhuman animals did not practice deceit would be that we lack one sort of evidence that they have beliefs about the beliefs of others. The inference that they lack any sort of HOTs whatsoever would not yet be justified. First, the assumption that deceit is in general the "clearest way in which an animal can manifest second-order beliefs" is not obvious. It is one sort of evidence we could have for second-order beliefs, but the claim that it is the "clearest" sort of available evidence stands in need of argument.

Second, even if this would be the clearest sort of possible evidence, we might find other, less clear but still quite persuasive, sorts. For example, even if they never genuinely intend to deceive others, the ability of the members of some species to predict the behavior of conspecifics might speak heavily in favor of their having meta-beliefs. Compare the following: (1) The clearest possible evidence I could have that people walked on the moon would be to have witnessed it firsthand. (2) I have not witnessed it, firsthand. Therefore, (3) I should conclude that people have not walked on the moon.

Third, there is evidence from developmental psychology that suggests we should be very cautious about moving from "no deceit" to "no thoughts about the thoughts of others." Prior to age four or five, children seem incapable of attributing false beliefs to others, or even to themselves. For example, if you show them a candy box and ask what is in it, and they reply "candy" and you then open the box to reveal pencils, then if you go on to ask what they originally thought was in the box, before you opened it just moments ago, they reply "pencils." Not surprisingly, when you ask them what someone else will think is in the box when they first see it, they also reply "pencils" suggesting that they are incapable of attributing false beliefs, either to themselves, or to others. Hence, prior to

age four or five, children seem to be incapable of genuine deception. Nonetheless, they seem perfectly able to attribute beliefs to others; they verbally make claims about what other people believe, and make predictions about what people will do on the basis of those claims. They also tend to say that beliefs, unlike external objects are in some sense "in the head" suggesting that they have at least some primitive understanding of the mental/non-mental distinction. H.M. Wellman and others have suggested that prior to age three, children may have a naive "copy-theory" of the mind, which allows them to attribute representations to themselves and others, but not misrepresentations. Insofar as this is plausible, we have a case of organisms that have thoughts about the thoughts of others but that are incapable of deception.[25]

There is, however, a more charitable reading of the argument that nonhuman animals lack HOTs, although it requires making explicit a suppressed premise appealing to natural selection. The argument has the following form:

(1) Nonhuman animals do not engage in deceit.
(2) If nonhuman animals had higher-order thoughts, then they could engage in deceit.
(3) Given the adaptiveness of deceit, if nonhuman animals could engage in deceit, they would.

∴ Nonhuman animals do not have higher-order thoughts.

(3) is the argument's suppressed premise – proponents of the argument in the literature have not explicitly invoked it, but it seems to be what is needed for the argument's validity. Moreover, it is not as if the premise has nothing to recommend it. In particular, it gains plausibility from the assumption that natural selection would have favored creatures who engaged in deceit. This is admittedly a rather Panglossian assumption, subject to the usual worries about Just-So Stories, but I am willing to overlook such worries, just as I overlook the quite worries one might have about (1). Instead, I focus my attention on (2), which has gone untouched by the critics of this sort of argument,

henceforth referred to as the "argument from the lack of deceit," and which embodies each of the two prejudices I aim to debunk.

Before examining premise (2), though, it is worth remembering that if this argument is to help show that nonhuman animals lack direct moral standing, then the conclusion that nonhumans lack HOTs must be read in the sense in which HOTs are necessary for conscious experiences on a Rosenthal-style theory. There is a real danger of equivocating here. For there is a continuum of kinds of representational states ranging from the very simple sort of representation one finds in a thermostat all the way up to a very complex, grammatically structured sentence in a Fodorian language of thought that has lots and lots of inferential connections. Intuitively, the HOT theory is perhaps most plausible when we understand the use of 'thought' in that theory as referring to representations that are not so heavily loaded, and are a bit closer to the thermostat end of the spectrum than the Fodorian end of the spectrum. However, the argument from the absence of deceit is most plausible when 'thought' is read as referring to representations more toward the Fodorian end of the continuum. So when we move from the argument from the absence of deceit to the case for no direct moral standing, there is a serious worry about equivocation that must be faced. Having registered this worry, I shall put it to one side, and focus on the argument from the absence of deceit, assuming throughout that no such fallacy is being committed.[26]

In any case, (2) relies on the assumption that a creature could not have HOTs about conative states like desires without also at least being capable of having HOTs about cognitive states like beliefs.[27] For if there were such a creature, it would be a direct counter-example to (2), as the ability to engage in genuine deceit requires the capacity to attribute a false belief to another. We are, however, given no reason whatsoever by the proponents of this argument to think such a creature is not possible. So (2) embodies the second prejudice I aim to debunk - the tendency to ignore HOTs about conative states.

Furthermore, work in developmental psychology strongly suggests that this possibility is not a merely logical one. Children refer to desires considerably earlier than they refer to beliefs. Karen Bartsch's and Henry Wellman's analysis reveals that, "an overwhelming use of desire verbs, often found in conjunction with *no belief verbs at all*, is characteristic before about two and a half years of age. After that time, the amount of belief verb production increases…"[28] One might worry that these early, pre-belief uses of 'desire', 'want' and their cognates were simply disguised imperatives, showing no genuine understanding of the concept of a desire. Bartsch and Wellman argue that this interpretation is not plausible given the ubiquity of *contrastive* uses of such term, in which children explicitly contrast one desire with an informative contrast, e.g., with another desire (perhaps had by another person or by the same person at another time) or with the object or outcome of the desire. Such uses do not seem to be plausible understood as commands or requests. For example, cases in which children mentioned a desire and then explicitly contrasted it with someone else's, as with, "Do you want me to look both ways? I don't wanna look both ways…" were plausibly taken to be relatively clear cases of genuine psychological reference rather than mere requests or commands. Analyzing 10,000 utterances[29] from ten children in which the children used terms like 'want', 'believe', etc., Wellman found that "genuine reference to a character's desire via the term *want* begins quite early and is well established even before the second birthday."[30] Bartsch and Wellman argue that two-year-olds have a "simple desire psychology," and are capable of attributing desires to others, and making predictions on the basis of those attributions, *without* also being able to attribute beliefs to them.

Paul Harris argues that this simple desire psychology develops prior to the richer belief-desire psychology because the former serves a function in planning and practical reasoning that is independent of language, whereas the primary function of the latter is to enhance the child's ability as a conversationalist. On this model, one would not expect the richer belief-desire psychology to arrive until the child begins to engage in

conversation, whereas the simple desire psychology could emerge before then, insofar as the child might well engage in planning and primitve practical reasoning before engaging in conversation. Harris' explanation of the phenomena is speculative and controversial, and nothing here commits me to his account being correct. Rather, I mention Harris' explanation in passing to indicate that the phenomenon Wellman has discovered need not be seen as inexplicable or mysterious – Harris's is no doubt one of many prima facie plausible explanations of the phenomenon. [31]

Philosophical tradition holds that beliefs and desires are a "package-deal" - one can only attribute them together, holistically. I do not aim to challenge that tradition, nor do I think Wellman needs to be understood as doing so. One simply must distinguish the two-year-old's conception of a desire from our more refined, full-fledged conception of a desire. Unlike us, the two-year-old is incapable of seeing that how your desires will lead you to act is a function of how you represent the world; for the two-year-old, it seems, your desires lead you to do what will *in fact* satisfy your desires, *regardless* of whether you are aware that doing so will satisfy them. On this primitive picture of desires, desires are a bit like magnets, pulling the agent toward their object without requiring any intermediate representations of the world "in the agent's head."

However, this magnet metaphor suggests an objection. Perhaps the child's use of desire-talk is meant to attribute an objective property to things in the world, rather than a subjective property to people. So, for example, when a young child says, e.g., that she wants candy and that Billy wants candy, she is attributing a kind of magnetic power *to the candy* and not attributing any sort of property at all to herself or Billy. On this model, young children are a bit like projectivists say we all are in ethical and valuational discourse. In Humean terms, we "stain and guild" the world with our desires without realizing that is what we are doing, and see them as objective properties of the world (though, so far as I know, no projectivist has appealed to this interpretation of developmental psychology to bolster their case for projectivism). The trouble with this

account is the way in which very young children are able to recognize both intersubjective and intrasubjective conflicts of desire.[32] The simplest explanation of this ability would be to attribute an understanding of desires as subjective states of people, rather than as objective states of objects that relate differently to different people. Furthermore, the fact that children's desire language (as opposed, perhaps, to their value language) is a language of attributing states to people and not to objects provides at least some evidence in favor of the more natural, subjectivist interpretation. Finally, the fact that children seem to develop a subjective conception of desires without showing any intervening confusion or signs of a transition from an objective conception to a subjective conception (at any rate, so far as I know, there is no such evidence; the hypothesis is empirically testable) provides further confirmation of the subjectivist interpretation. So findings from developmental psychology speak against the prejudices that have led many to suppose nonhuman animals all lack HOTs. The possibility of a creature's having HOTs about desires without being capable of having HOTs about beliefs is not a mere logical possibility; very young children seem to be actual instantiations of this possibility. Hence, (2) is false, and the argument from the lack of deceit is unsound.

Even more importantly, though, (2) illustrates how its proponent has fallen prey to another common prejudice. Quite obviously, if there are creatures who can have thoughts about their own mental states but are utterly incapable of having thoughts about the mental states of others, then (2) is false. To support (2), some argument needs to be given for thinking there are no such creatures, but no such argument is forthcoming. The possibility of an unreflectively solipsistic theory of mind is simply ignored. Not just Carruthers, but Bennett, Premack and Woodruff, Seyfarth and Cheney, and Whiten, among others all seem willing to move from "no deceit" to "no HOTs" without so much as commenting upon the possibility of a solipsistic perspective.

Furthermore, we once again have reason to believe that this is not a merely logical possibility. Intuitively, it seems that our own thoughts are more directly and immediately

available to us than the thoughts of others (though few nowadays would maintain that they are incorrigible). This suggests that perhaps we come to know our own thoughts in a rather different way from the way in which we come to know the thoughts of others. Insofar as this is correct, it provides some reason for taking solipsism seriously, as it suggests that having a HOT about another creature's mental states might require more cognitive sophistication than is needed for having a HOT about one's own mental states. However, this intuition may turn out to be mistaken; Gopnik, for example, argues that we are under the illusion that our access to our own thoughts is noninferential because we have become so expert at it. In the same way that Gary Kasparov "just sees" the right move we can now "just see" our own thoughts, though we in fact had to learn how to do so, on Gopnik's account.[33] Gopnik emphasizes that in early childhood we are rather clumsy at taking what Dennett calls the "intentional stance" even though we go on to become quite proficient at it, and forget our initial failings. Strikingly, young children tend to have real problems with attributing false beliefs, as the "pencils in the candy box" case illustrates. On the other hand, as many of Gopnik's commentators point out, Gopnik is, and should be, willing to allow that young children are pretty close to incorrigible about their *present-tense* attributions of mental states to themselves, as compared to their present-tense attributions of mental states to others. This suggests that the ability to attribute mental states to oneself is more basic than attributing such states to others.

Again, evidence from developmental psychology bolsters this suggestion. Patricia Smiley and Janellen Huttenlocher have argued that their data support a model according to which "the child's categories initially cover only internal states of self, then, observable features of others' behavior, and, finally, inferred internal states of others."[34] Strikingly, children begin attributing emotions to themselves some time before they attribute them to others. Also, when they do begin attributing emotions to others, they initially seem to be somewhat less adept at doing so than they are at attributing emotions to themselves,

suggesting that they may be struggling to expand a concept whose extension was initially limited to themselves. Here are Smiley and Huttenlocher:

> Thus, in all the studies where naturally occurring events are the instances covered by word meanings, by about 2 years at least half the children include their own emotional states of at least one sort. At around 2 years, however, probably only a few children include instances concerning other people, and these appear to involve habitual behavioral expressions of emotion – crying, smiling, and stomping around…Thus, children's word meanings at first cover internal states of the self and some observable aspects of others' experience. By six months to a year later, children begin to use words, not just for observable aspects of others' experiences, but apparently for their internal states as well.[35]

Furthermore, Hoffman provides suggestive though anecdotal evidence that when young children observe others crying, or looking sad, they then console *themselves*, as if they could recognize emotions only as in themselves.[36]

The severely autistic may provide another case of unreflective solipsism. Tragically, the severely autistic seem incapable of attributing mental states to others; as Baron-Cohen colorfully puts it, they "fail to develop the capacity to mindread in the normal way."[37] However, those who study autism often characterize it as an "inability to attribute propositional mental states (such as beliefs and knowledge) to *other* people,"[38] implying that perhaps the autistic are able to attribute such mental states to themselves in spite of being unable to attribute them to others. The "mindblindness" hypothesis is supposed to be an account of how an agent's "theory of mind module" is damaged.

While this interpretation of the data would sit well with the possibility of unreflective solipsism, it is not the only interpretation of "mindblindness." For it has also been suggested that the "theory of mind module," if there is one, has as its function "providing intentional markers to the inputs which help fix belief…damage to such a module would not in itself mean that the agent was unable to formulate thoughts about thoughts – and indeed there is evidence that people with autism do not suffer so pervasive a disability as this. But if people with autism were impaired in their capacity to apply intentional markers to perceptual contents, it would not be surprising if they were less

than fully competent at forming beliefs about beliefs."[39] On this interpretation, the mindblind are lacking in certain markers that are attached to their perceptual cues and that would enable them to form belief about the thoughts of others with ease. The capacity to have thoughts about thoughts at all, however, is on this account not simply a function of having such a module, as the belief fixation itself does not happen in the module that attaches the relevant cues. So it would seem possible for someone lacking such a module to suffer from a severe inability to attribute thoughts to others, but still be able to attribute thoughts to themselves with little or no trouble, so long as our self-attributions of mental states does not depend upon this same module. Nor does it seem plausible, if the phenomenology of introspection is given much weight, that self-attributions would depend on this module – it is not as though we must look in the mirror or carefully observe our own behavior to know how we feel at a given time.

At any rate, there is some indication that the autistic are quite proficient at focusing on their own visual representations, suggesting that one's ability to think about one's own mental states can come apart from one's ability to think about the mental states of others.[40] Further, at least one researcher has suggested that the autistic may be *especially* proficient at focusing on their own mental states:

> Deep meditation or prolonged, intense focusing of attention on inner feelings, thoughts or images may produce a state similar to hypnotic analgesia...this human ability voluntarily to direct attention toward inner feelings, thoughts, or images, and to block out all extraneous environmental stimuli, may also explain the autistic child's ability to produce pain anesthesia.[41]

In fact, however, any speculation at this point about the proficiency of the autistic at thinking about their own mental states would at best be premature. For as Uta Frith and Francesca Happe have recently noted, "while the inability to attribute mental states to others has been studied extensively in children with autism, there is scarcely any work on the ability to attribute mental states to self."[42] Furthermore, Frith and Happe's research is

meant to suggest that the autistic "may know as little about their own minds as about the minds of other people."[43] Frith and Happe's account is partly based on the "theory of mind module" theory, though, and this theory is itself still fairly controversial, particularly if we must understand this module as necessarily coming "on-line" only when its fully functioning both respect to self and others and with respect to desires and beliefs (Wellman's research is again quite relevant with respect to the latter).

On the other hand, Frith and Happe's hypothesis would, as they note, seem to explain a number of the problems the autistic characteristically have. For instance, a lack of self-knowledge might help explain why they adopt another person's contrary opinion without acknowledging that they had changed their mind. Nonetheless, though, on the whole, the jury is still out given the paucity of research on the question, as Frith and Happe themselves note (their own study is itself preliminary and deals with only three subjects). The crucial point for present purposes is that at least some instances of autism may turn out to provide an empirically plausible model of unreflective solipsism; it is not *obvious* that the empirical results will favor the Frith and Happe hypothesis. So to defend the inference from "no deceit" to "no HOTs" on the grounds that there are no empirically plausible instances of unreflective solipsism would, at the very least, be premature.

In light of these considerations, it seems fair to conclude that the argument from the absence of deceit rests on two false assumptions. First, it rests on the assumed impossibility of an agent capable of having HOTs about conative states like desires but incapable of having HOTs about cognitive states like beliefs. Second, it relies on the assumption that there could not be an agent capable of having HOTs about its own mental states but incapable of having HOTs about the mental states of others - that there could

not, in my terms, be an unreflective solipsist. Each of these assumptions, once brought to light, is seen to be unfounded, not just in that these possibilities are genuine conceptual ones that we have as yet been given no evidence against, but also in that we have some reason to think that these possibilities are actually instantiated, at least in young children and the autistic. This last point is important; if the possibilities I am emphasizing were *merely* conceptual ones, then their neglect might be a less serious vice.

My examples have all been examples of human beings capable of speaking, and this has been no coincidence. Even though an agent's sincere utterances in a natural language are not the *only* kind of evidence we can get about her thoughts, they are an especially clear and convincing kind. Since nonhuman animals do not speak, there *may* be no equally clear evidence that any of them are having HOTs about their own desires. The main point, for present purposes, is that the putative absence of deceit provides no compelling positive reason to suppose that nonhuman animals lack HOTs. I do not claim to have proven that they do in fact have such HOTs, only to have shifted the burden of proof back to those who would deny that they do.

A question naturally arises as to where the burden of proof is to be laid in these matters, insofar as practical questions force us not to suspend judgment. On the one hand, Lloyd Morgan's canon enjoins us to settle "on the most killjoy, least romantic hypothesis that will account systematically for the observed and observable behavior."[44] If we accept this suggestion, then we should proceed with the default hypothesis that nonhuman animals lack any HOTs until we are given very good and compelling reason to think otherwise; the burden of proof lies with the romantics. On the other hand, philosophical method, as most clearly exemplified in the Goodmanian and Rawlsian idea

of seeking a "reflective equilibrium," favors giving common sense the benefit of the doubt. To paraphrase J.L. Austin, common sense may not have the last word, but it *is* the first word.[45] Assuming common sense holds that nonhuman animals are capable of suffering in morally salient ways, then philosophical method endorses our holding on to that assumption until we find good and sufficient reason to reject it.

It might seem like we are now at an impasse, with two irreconcilable standards pulling us in different directions. In fact, the two standards are not irreconcilable, so long as one is willing to restrict the scope of each. *Within* scientific practice, putting the burden of proof on the romantics is appropriate; until that burden is met, we cannot claim to know with *scientific* certainty that our view is correct. On the other hand, if we are trying to decide how we should live our lives, and whether we need to give any direct moral weight to the suffering of nonhuman animals, then it would often be perverse to demand scientific certainty. For example, we might not know with complete scientific certainty that smoking increases the risk of lung cancer for the simple reason that we have been unwilling, for good moral reasons, to do the sorts of studies that would be necessary for such certainty. Nonetheless, it would be absurd to suggest that we ought not assume that smoking does increase the risk of lung cancer when making public policy. Likewise, it would be perverse to suggest that we ought not assume that pigs, dogs, and rats are capable of suffering in a way that gives them direct moral significance because this assumption has not been scientifically proven. We need to distinguish scientific certainty from moral certainty, and note that each has its proper role. If we lack scientific certainty, then we should do more research, but this does not imply that in the meantime we should, when making practical and moral decisions, suspend judgment.

III. CONCLUSION

Against what I see as two philosophical prejudices, I have argued that we ought to take seriously both the possibility of a simple desire psychology and of unreflective solipsism. Considerable evidence from developmental psychology and from work on the autistic suggests that each of these possibilities may actually be instantiated in our own species. As I have argued, taking these possibilities seriously might shed new light on the mental lives of nonhuman animals. Considering these possibilities might be of use in the philosophy of mind, as they might help HOT theorists to counter the objection, pressed by Dretske and others, that their view absurdly implies that no nonhuman animals have any conscious mental states. As a moral argument, the present account is meant only to shift the burden of proof to those who suppose, on the basis of the argument from the lack of deceit, that nonhuman animals do not suffer in a way that confers any direct moral significance. While the present discussion may suggest various strategies for determining whether any nonhuman animals actually are unreflective solipsists, I have intentionally not engaged in such speculation here.[46] Instead, my aim has been to avoid unnecessary controversy by drawing examples from language-using members our own species. Any argument that unreflective solipsism or a simple desire psychology is actually instantiated by the members of a particular nonhuman species would inevitably generate considerable controversy that would distract from the more general point that such possibilities must be not be dismissed ex ante, but rather investigated on a case-by-case basis. As far as I can tell, such psychological profiles are perfectly coherent possibilities. There may be subtle and ingenious arguments for the conclusion that unreflective solipsism is incoherent, and I have no way of being sure that I have not overlooked them. Until such

arguments are given, we should take the possibility of unreflective solipsism, as well as the possibility of a simple desire psychology, seriously.

---

[1] Many thanks to Simon Blackburn, Robin Flaig, Dien Ho, William Lycan, Michael Martin, Jay Rosenberg, David Rosenthal, Dan Ryder, J.J.C. Smart, Kim Sterelny, and Ralph Wedgwood for useful discussion of earlier drafts of the present material.

[2]Peter Carruthers, "Brute Experience," *Journal of Philosophy*, LXXXVI, (1989), 258-269, p. 258.

[3]In fact, Carruthers offers two arguments for his thesis, and eventually rejects one of those arguments in favor of the other.

[4]In this vein Guven Guzeldere has argued that the HOP theory, if it is to be plausible, collapses into the HOT theory. See Guven Guzeldere, "What Passes in One's Own Mind," in *Conscious Experience*, ed. Thomas Metzinger, (Schoningh: Imprint Academic, 1995), 335-359.

[5]David Armstrong, *A Materialist Theory of the Mind* (London: Routledge, 1968), Paul Churchland, *Matter and Consciousness* (Cambridge: MIT Press, 1988), William James, *The Principles of Psychology* (New York: Dover, 1950), John Locke, *An Essay Concerning Human Understanding* (New York: Dover, 1959), William Lycan, *Consciousness* (Cambridge: MA, 1987), David Rosenthal, "Two Concepts of Consciousness," *Philosophical Studies* XCIV, 329-359, Norton Nelkin, *Consciousness and the Origins of Thought* (Cambridge: Cambridge University Press, 1996), Daniel Dennett, *Consciousness Explained* (Boston: Little Brown, 1991) and Dennett, "The Message is: There is no *Me*dium." *Philosophy and Phenomenological Research*, 4, 919-931, p. 928.

[6]One interesting wrinkle here is that it is intuitive to suppose that the best form of happiness consists in being so engrossed in what you are doing that you are not conscious of your own happiness. In fact, I have considerable sympathy for this point. Still, it seems that in the case of pain and things that are bad from a moral point of view, that there is no clear analogue of this point. In any case, I shall put this objection to one side for present purposes. Thanks to Kim Sterelny for bringing this point to my attention.

[7]See Carruthers, *op. cit*., Peter Harrison, "Do Animals Feel Pain?" *Philosophy*, LXVI (1991), 25-40, Joel Feinberg, "The Rights of Animals and Unborn Generations," in *Philosophy and Environmental Crisis*, ed. William Blackstone, (Athens: University of Georgia Press, 1974), 43-68, and Rene Descartes, "The Correspondence," in *The Philosophical Writings of Descartes*, III, ed. John Cottingham, et. al., (Cambridge: Cambridge University Press, 1993), p.148.

[8]See Donald Davidson, "Rational Animals," in ed. LePore, E. and McLaughlin, B., *Actions and Events* (New York: Basil Blackwell, 1985), R.G. Frey, "Why Animals Lack Beliefs and Desires," in ed. Tom Regan and Peter Singer, *Animal Rights and Human Obligations* (Englewood Cliffs: Prentice-Hall, 1989), 39-42, Norman Malcolm, "Thoughtless Brutes," Presidential Address delivered before the Sixty-ninth Annual Eastern Meeting of the American Philosophical Association, Roderick Chisholm and Wilfrid Sellars, "The Chisholm-Sellars Correspondence on Intentionality," *Minnesota Studies in the Philosophy of Science*, vol. II, eds. H. Feigl, et. al. (Minnesota: University of Minnesota Press, 1958), 214-248, Stephen Stich, "Do Animals Have Beliefs?" *Australian Journal of Philosophy*, LVII (1979), 15-28.

[9]Jonathan Bennett, "How to Read Minds in Behavior: A Suggestion From a Philosopher," in ed. Andrew Whiten, *Natural Theories of Mind* (Oxford: Basil Blackwell, 1991), 97-108, Carruthers, *op. cit.*, Daniel Dennett, *The Intentional Stance* (Cambridge: MIT Press, 1987), Dorothy Cheney and Robert Seyfarth, *How Monkeys See the World* (Chicago and London: University of Chicago Press, 1990), Richard Byrne and Andrew Whiten, "Primate Tactical Deception," in Whiten, *op. cit.*, 127-141, and David Premack, "'Does the Chimpanzee Have a Theory of Mind?' revisited," in ed. Richard Byrne and Andrew Whiten, *Machiavellian Intelligence* (Oxford: Clarendon Press, 1988), 160-179.

[10]Bennett, *op. cit.*, p. 103.

[11]Importantly, 'unreflective' is used attributively here - the creatures in question are meant to be unreflective *qua* solipsism. They are incapable of entertaining the thought that other creatures might have any mental states, so they are incapable of considering solipsism and its negation as competing hypothesis.

In another sense, of course, the creatures I have in mind are *quite* reflective - they have thoughts about their own mental states.

[12]Fred Dretske, *Naturalizing the Mind* (MIT Press: Cambridge, MA, 1995), pp. 110-111. Dretske here cites developmental psychologists like Wellman; I present my own interpretation of their work below.

[13]Alex Byrne also presses this sort of worry against the HOR theorists, under the heading of the "dog problem," but concludes that they should simply accept the consequence that many nonhuman animals indeed lack conscious mental states. See Alex Byrne, "Some Like it HOT: Consciousness and Higher-Order Thoughts," *Philosophical Studies*, 86 (1997), pp. 103-129, esp. pp. 112-113.

[14]There is also the position that while having thoughts does not require language, having meta-thoughts does require having language. This position is less extreme than the Davidsonian position, but I am unaware of an argument for it whose premises would not also entail the more radical Davidsonian thesis. I return to this sort of position in a later footnote. For a defense of the view that some nonhuman animals do employ concepts, see Colin Allen and Marc D. Hauser, "Concept Attribution in Nonhuman Animals: Theoretical and Methodological Problems in Ascribing Complex Mental Processes," *Philosophy of Science*, 58 (1991), pp. 221-240. For an explicit defense of the view that thought does not require language (even a "language of thought"), see Ruth Barcan Marcus, "Some Revisionary Proposals about Belief and Believing," *Philosophy and Phenomenological Research*, 50, Supplement. (Autumn, 1990), pp. 133-153.

[15]David Rosenthal, "Two Concepts of Consciousness," *Philosophical Studies*, XLIX (1986), 329-359.

[16]Rosenthal, *op. cit.*, p. 330.

[17]Rosenthal, *op. cit.*, p. 335.

[18] See David Rosenthal, "Consciousness, Content, and Metacognitive Judgments," *Consciousness and Cognition*, IX, 1, January, 2000.

[19] Though we presumably will be conscious of the content of that thought, since it is the content of the HOT, and not the content of the target state, that determines the qualitative character of our experience – indeed, there need not even *be* a target state, as Rosenthal himself is happy to admit. Cases in which we confabulate about our own mental states are a nice case in point of this phenomenon.

[20]See, for example, especially Rocco Gennaro, "Brute Experience and the Higher-Order-Thought Theory of Consciousness," *Philosophical Papers*, XXII (1993), 51-69).

[21]Peter Carruthers, *The Animals Issue* (Cambridge: Cambridge University Press, 1992), p. 137.

[22]See Dan Dennett, "Conditions of Personhood," *The Identities of Persons* (Berkeley: University of California Press, 1976), p. 184.

[23]Interestingly, Morgan himself was willing to attribute second-order thoughts to some nonhuman animals. See Lloyd-Morgan, *The Animal Mind* (London: Edward Arnold, 1930).

[24]Carruthers, *op. cit.*, p. 130.

[25]See H.M. Wellman and J.D. Wooley, "From Simple Desires to Ordinary Beliefs," *Cognition*, XXXV (1990), 245-275.

[26]Thanks to Kim Sterelny for emphasizing the importance of this worry.

[27]Rosenthal himself would respond to this argument by questioning premise (2) as well, though along rather different lines from the ones pursued here. On his view, one might have a HOT about a given mental state M but the HOT might nonetheless not involve any mental concepts: "HOTs need not characterize their targets as mental, but only as states." (David Rosenthal, "Consciousness, Content, and Metacognitive Judgment," *Consciousness and Cognition*, IX, 1, January, 2000). In that case, a creature might have HOTs but be utterly incapable of deceit because the creature lacks the concept of a belief.

I eschew this strategy for two main reasons. First, without some specification of what the content of these HOTs might be, the approach seems like trying to have all the benefits of honest toil through theft (to paraphrase Russell). As it stands, the most natural characterization of the content of the relevant HOTs would be as employing mentalistic concepts – if we are to consider alternative contents, then we need an explicit articulation of them to see whether they are really more or less plausibly attributed to non-linguistic creatures. Second, and more importantly, on Rosenthal's own view, it is the character of the HOT, rather than the character of the relevant first-order state, that determines the qualitative character of one's experience. So, for example, if I have a HOT whose content is that I am seeing a blue pen and that HOT is caused by and about a perception of a red pen, it will seem to me that I am seeing a blue pen. In

that case, though, it is hard to see how having a HOT whose content included no mental concepts could ground the relevant qualitative character. Imagine trying to articulate, even in very coarse-grained terms, what one's experiences are like without employing concepts like "painful," "tasty," "loud," "bright," etc.

[28] Karen Bartsch and Henry Wellman, *Children Talk About the Mind* (New York: Oxford University Press, 1995), p. 27.

[29] These utterances were collected from ten different children at very different points in time, and by different investigators, where each investigator had different research goals, none of which included examining talk about mental states (at least, when the data were collected), as Bartsch and Wellman note in *Children Talk About the Mind* (New York: Oxford University Press, 1995), p. 25. This makes it unlikely that the children were in any way specially primed to talk about the mind or exposed to parents or investigators who were especially likely to prompt such talk. That the data were taken from ten different children also makes it somewhat unlikely that the results are simply the result of unusual precocity or language ability of a given child. Finally, Bartsch and Wellman made a point to exclude what, from the context, seemed to be "merely conversational uses" of belief and desire terms. In particular, "a term was not counted as a genuine belief or desire reference if it served only such conversational functions as getting someone's attention (e.g., 'You know what?'), turning over the conversation to someone else (e.g., 'Let's go to the park, what do you think?'), or softening a command or request (e.g., 'I wonder, Mom, can we have spaghetti?' or 'I think its time to watch Sesame Street'). Also excluded were short, unumbellished, or idiomatic phrases, such as 'You know,' 'I think so,' 'Don't know,' and 'I wanna.'" (Bartsch and Wellman, *op. cit.*, pp. 31-32).

[30] See Henry M. Wellman, "From Desires to Beliefs," in *Natural Theories of the Mind*, op. cit., 19-38, p. 33.

[31] Peter Carruthers, for example, has offered the following explanation of why such a simple-desire psychology might have developed, in evolutionary times apart from an ability to attribute beliefs. A simple desire psychology might have conferred an evolutionary advantage because it would allow desire-based predictions with a fairly high success rate. By contrast, a belief psychology with no desire component would have a poor success rate. So, one might conclude, a desire psychology would evolve first, followed by a more complex belief-desire psychology. Paul Harris attributes this position to Carruthers in a footnote to his, "Desires, Beliefs, and Language," in Peter Carruthers and Peter K. Smith, eds., *Theories of Theories of Mind* (Cambridge: Cambridge University Press, 1996), pp. 200-220, p. 219. Carruthers' suggestion is purely speculative, but worth giving serious consideration nonetheless.

[32] See Bartsch and Wellman, *op. cit.*, pp. 85-86.

[33] See Alison Gopnik, "How We Know Our Minds: The Illusion of First-Person Knowledge of Intentionality," *Behavioral and Brain Sciences* XVI (1993), 1-14.

[34] Patricia Smiley and Janellen Huttenlocher, "Young Children's Acquisition of Emotion Concepts," in ed. Carolyn Saarni and Paul Harris, *Children's Understanding of Emotion* (Cambridge: Cambridge University Press, 1989), 27-49, p. 27.

[35] Smiley and Huttenlocher, *op. cit.*, pp. 38-39.

[36] "That is, the cues associated with another person's distress evoke an upset state in him, and he may then seek comfort for himself. Consider a colleague's 11-month-old daughter who, on seeing another child fall and cry, first stared at the victim, appearing as though she were about to cry herself, and then put her thumb in her mouth and buried her head in her mother's lap – her typical response when she has hurt herself and seeks comfort." M. Hoffman, "Developmental Synthesis of Affect and Cognition and its Implications for Altruistic Motivation," *Developmental Psychology*, 11 (1975), 607-622, p. 614.

[37] Simon Baron-Cohen, *Mindblindness*, (Cambridge: MIT Press, 1995), p. 5.

[38] Simon Baron-Cohen, "The Theory of Mind Deficit in Autism: How Specific is It?" in ed. George E. Butterworth, et. al., *Perspectives on the Child's Theory of Mind* (Oxford: Oxford University Press, 1991), 301-314, p. 301.

[39] Greg Currie and Kim Sterelney, "How could social intelligence be modular?" joint paper delivered at the 1999 Australasian Association of Philosophy, July 2-10th, Melbourne, University of Melbourne.

[40] See Temple Grandin's *Thinking In Pictures* (New York: Doubleday, 1995) for some firsthand evidence.

[41] Cheryl Seifert, *Case Studies in Autism* (Lanham: University Press of American, 1990), p. 60.

[42] Uta Frith and Francesca Happe, "Theory of Mind and Self-Consciousness: What is it Like to Be Autistic?" *Mind and Language*, 14 (1999, pp. 1-22, p. 8.
[43] Frith and Happe, p. 7. Many thanks to Michael Martin for bringing this work to my attention.
[44]Dennett, *The Intentional Stance*, p. 246.
[45]See Nelson Goodman, *Fact, Fiction, and Forecast*, fourth edition (Cambridge: Harvard University Press, 1983), John Rawls, *A Theory of Justice* (Cambridge: Harvard University Press, 1971), and J.L. Austin, *Philosophical Papers* (Oxford: Oxford University Press, 1961), p. 185.
[46] I should note in passing, however, that in another paper, in which I analyze the distinction between pain and suffering, I speculate further along these lines. See my "Beastly Suffering," unpublished manuscript.