CHAPTER ELEVEN

THE THIRD DOGMA OF RATIONALISM

MARK OKRENT

Self Apprehension

John Haugeland has recently suggested that traditional rationalism falls prey to two unfounded 'dogmas', positivism and cognitivism.[1] According to Haugeland, positivism is the metaphysical view that 'reality is exhausted by the facts', while cognivitism is the assumption that 'reason is to be understood in terms of cognitive operations on cognitive states', where a cognitive state is a propositional attitude towards a propositional content and a cognitive operation is a rational inference. Although I am far more sanguine than Haugeland is that, properly qualified and understood, the language of cognitivism, with its talk of beliefs and desires, can be both benign and useful, I agree with his general claim that both of these assumptions, however widespread, are mere unwarranted dogmas. In particular, I think that the doctrine that explicit rational inference is necessary for an agent to act for reasons, engage in purposeful action and possess intentionality is an unwarranted dogma based solely on a misleading view of the nature of purposeful action.

Quine's diagnosis of the first two dogmas of empiricism led to Davidson's diagnosis of a third dogma of empiricism, the dogma that there are conceptual schemes. I have located a third dogma of rationalism that has roughly the same relation to Haugeland's first two dogmas that Davidson's third dogma has to Quine's. The third dogma of rationalism is the view that in order for an agent to have any intentions directed towards the world, that agent must have intentions that are directed towards its own intentions as its own intentions. That is, the third dogma of rationalism is

[1] John Haugeland, "Two Dogmas of Rationalism", unpublished paper, delivered at The Sixth Annual Meeting of The International Society for Phenomenological Studies, Pacific Grove, CA, July 16, 2004.

that the abilities to have second order intentions, to intend one's own intentions as intentions and to apprehend oneself as an intentional agent, are necessary conditions on being an intentional agent at all.

As with any self-respecting dogma, the first clear and precise statement of this dogma appears in Kant, most clearly in Section 16 of the B Deduction. "It must be possible for the 'I think' to accompany all my representations; for otherwise something would be represented in me which could not be thought at all, and that is equivalent to saying that the representation would be impossible, or at least would be nothing to me."[2]

But the view also shows up repeatedly in the work of many prominent twentieth century philosophers, including Davidson, Sellars, Korsgaard, and Brandom, many of whom explicitly mention their debt to Kant. Davidson, for example, frequently expresses the view in the traditional language of cognitivism. To cite one instance, in "Rational Animals" he says "...in order to have any propositional attitude at all, it is necessary to have the concept of a belief, to have a belief about some belief."[3] Given this way of putting the point, one might be tempted to say that for Davidson (and one might as well add, Brandom, Sellars, and Korsgaard) only a being that is capable of self consciousness is capable of true, real, original intentionality. But this way of stating the content of the dogma is accurate only if one is clear that for these thinkers 'self-consciousness' is given a certain definite analysis. What Davidson really suggests in the above quote is that what is necessary for an agent to have any intentional states is that that agent have states that reflexively intend some of their own states *as* intentional states, that the agent have some states that have the content that some other states of theirs are intentional. This is the import of the Kantian 'I think' and I will take this to be the canonical form of the third dogma of rationalism.

In this paper I develop four related themes. In the first part of the paper I offer an interpretation of Kant's statement of the Third Dogma in the B Deduction that is sensitive to its context in Kant's critical philosophy In the second section I argue that the third dogma of rationalism, in its twentieth century guise, is motivated by a certain definite way of understanding the importance, role, and nature of what John Haugeland calls cognitive operations, that is, rational inference. Third, I argue that, perhaps surprisingly, this view of the role and importance of rational inference is motivated by the Kantian analysis of *practical* reason, in

---

[2] Immanuel Kant, *Critique of Pure Reason* (KrV), trans. Norman Kemp Smith (London: Macmillan, 1968), KrV B. 131.

[3] Donald Davidson, "Rational Animals" in *Subjective, Intersubjective, Objective*((Oxford: Oxford University Press, 2001), p. 104.

particular by Kant's analysis of what it is to act for or because of a reason. Fourth, I argue that this analysis of acting for a reason is fundamentally misguided and for that reason the third dogma of rationalism is unwarranted.

Kant on Judgment and the 'I Think'

For Kant, "It must be possible for the 'I think' to accompany all my representations; for otherwise something would be represented in me which could not be thought at all, and that is equivalent to saying that the representation would be impossible, or at least would be nothing to me."[4] This assertion, while pithy and memorable, is also unfortunately written in such a way that it is easy to misinterpret. Kant clearly is committed by this statement to the view that the possibility of the 'I think' accompanying some representation is necessary for the possibility of that representation being thought as the representation of something. This is what the crucial second clause asserts. It is possible to think of some representation as representing some thing only if it is possible for the 'I think' to accompany this representation. But does this imply that for x to be a representation of some thing it must be possible for the 'I think' to accompany it? That of course depends upon whether or not the possibility of x being thought as a representation of z is essential to x being a representation of z. And this *seems* to be the import of Kant's third clause, where he says that 'something represented in me could not be thought' is equivalent to saying that 'the representation is impossible'. But then he apparently takes this equivalence back in the final, parenthetical, clause. According to this final parenthesis, the assertion "representation x can not be thought by me as representing z, because I can not affix the 'I think' to it", is *not equivalent* to 'x representing z is impossible, because I can not affix the 'I think' to it." Rather, it is equivalent to 'representation x would be nothing to me if I can not affix the 'I think' to it". And this is clearly a different claim than the stronger claim, apparently asserted in the second clause, that no object can be represented without the possibility of the 'I think'. But which of these is Kant's considered opinion on the status and role of the 'I think'?

There is excellent reason to believe that the final parenthetical clause governs the whole and that Kant does not equate x being a representation of z with the possibility of x being 'thought' by me as a representation of z. Indeed, Kant is quite clear, both in the *Critique* and elsewhere, that he believes that it is possible for there to be a representation in y of which y is

[4] Ibid.

not even conscious, let alone capable of thinking. In the division of types of representations in the Dialectic, for example, Kant distinguishes between the genus 'representation' and its' species *perceptio*, or 'representation with consciousness'.[5] More importantly, in the Jasche *Logic* Kant continues the division by distinguishing between two forms of *perceptio*: to be acquainted (*kennen*) with something, "or to represent something in comparison with other things, both as to sameness and as to difference" and being acquainted with something with *consciousness*, or *cognition* (*erkennen*). Both of these, Kant tells us, involve intentions directed towards objects, but animals are only acquainted with objects, they do not cognize them. "Animals are acquainted with objects too, but they do not *cognize* them.[6] It is only in the next division that Kant reaches understanding, "to cognize something through the understanding by means of concepts, or to conceive." So, for Kant in 1800 (the date of the Jasche *Logic*), it is possible for an agent to have a representation of something, be conscious of that representation, and even represent that representation in relation to others in respect to sameness and difference, and thus be acquainted with objects, without that agent using concepts or being conscious *that* they are acquainted with objects. And, since in the Jasche *Logic* Kant uses 'to think' as equivalent with 'to cognize with concepts',[7] it is obvious that when he says in the B Deduction that if it were impossible for the 'I think' to accompany a representation x, then x could not be thought by me, this can't be equivalent to saying that if it were impossible for the 'I think' to accompany x, it would be impossible for x to be a representation of z.

What, then, *is* the 'I think' necessary for? For Kant, it is primarily necessary for two things, both of which are mentioned in the famous quote above: 'thinking' a representation as a representation of an object; and a representation, and the object represented by that representation, being something 'to me'. But how are we to interpret these?

What does Kant mean when he speaks about 'something represented in me which is thought'? One of the keys to interpreting this is given in Kant's division of representations in the Lectures on Logic. He tells us there that animals, who are incapable of having the 'I think' accompany their representations, can be acquainted with objects perceptually, and even represent similarities and differences, but they can't cognize objects. To be acquainted with something is to "represent something with other

[5] Ibid., A320/B376.

[6] Kant, *Lectures on Logic* (VL), ed. J.M. Young (Cambridge: Cambridge University Press, 1992), pp. 569-70.

[7] Cf. ibid., p. 564.

things, both as to sameness and as to difference". Cognition, on the other hand, Kant says, is being acquainted with something with *consciousness*. The acquaintance side of this division is clear enough. When one is acquainted with an object one represents that object as similar to and different from other objects. When my dog Sammie sees other dogs he reacts in similar fashion to all of them but differently in each of those cases than he does when he sees a squirrel. And this gives us reason to believe not only that his representations of the dogs are similar to one another and different from his representations of squirrels, but also that in some sense Sam synthesizes these representations and compares them in regard to their similarities and differences. In Kant's terms, Sammie represents the dogs in comparison with the squirrels in respect to sameness and difference. But what, then, does cognition, which Sam is incapable of, add? Kant says that cognition is acquaintance with consciousness. And at first sight this is odd, because an act in which one is acquainted with an object, such as my dog perceiving the difference between a dog and a squirrel, is already itself a conscious representation for Kant. So what can he mean when he says that cognition is acquaintance with something with consciousness?

In the division of kinds of representation in the Jasche *Logic* Kant says that the division is "in regard to the objective content'. That is, acquaintance is different from cognition, and simple perceptual cognition is different from a conceptual cognitive understanding, in respect to *what is represented* in these various types of state. From this perspective, when Kant speaks of cognition as acquaintance *with consciousness* (his emphasis), *what* is differentially conscious in cognitive states is not the state itself, but rather the *content* of those states. That is, Kant is suggesting that the differentia of cognitive acts is that the acts of acquaintance in which the sameness and difference of objects is represented are *themselves* consciously represented in cognitive acts. So, to return to my dog, he represents dogs and squirrels differently, and he can even distinguish between them when instances of both are present. He can represent something in comparison with other things, both as to sameness and to difference. But he does not represent that sameness and difference itself *as such*. That is, Sam is incapable of intending *that* he represents dogs and squirrels differently, and that these representations differ from one another in such and such respects. It is for this reason that Sam is incapable of using concepts. To have the concept 'dog' is at least to be potentially conscious of those respects in which representations of all

dogs are similar and the respects in which the representations of all dogs are different from the representations of non-dogs.[8]

The distinguishing feature of human representation is not introduced in the Transcendental Deduction through a contrast with animal representation, as it is in the Lectures on Logic. Nevertheless, the same differentia are suggested there as in the Logic. The Deduction in B begins with the suggestion that the distinguishing 'act of spontaneity' of the faculty of the understanding, an act which has "the general title 'synthesis'", is "the combination of a manifold in general".[9] This way of putting the matter makes it sounds as if what is at issue is the act of putting together representations itself. Fortunately, Kant immediately corrects this misleading impression. For he tells us, first, that it is not mere combination of representations which is the act of understanding, but the *representation* of the combination, and, second, that what is contained in combination is not merely a manifold and its synthesis, but also *the representation* of the *unity* of the combination or synthesis of a manifold: "...of all *representations* combination is the only one which cannot be given through objects." "But the concept of combination includes, besides the concept of the manifold and of its synthesis, also the concept of the unity of the manifold. Combination is the representation of the *synthetic* unity of the manifold. The representation of this unity cannot, therefore, arise out of the combination. On the contrary, it is what, by adding itself to the representation of the manifold, first makes possible the concept of the combination."[10] That is, the understanding combines a manifold in the sense that it represents the manifold as unified in a single representation – it represents the unity of what is manifold. Each of our representations of dogs is itself a synthesis or combination of a manifold of different representations. My dog, insofar as he is acquainted with objects, can have such synthetic representations. Indeed, he can represent two dogs together and note their similarity. But he can not represent that similarity of representation in a single representation by recognizing that both of these synthetic representations have been synthesized in the same way and that they are both instances of the same type of representation, 'dog'. The representation in which we recognize that Sam is similar to Fido and all other dogs in respect of being a dog, is, of course, the judgment that

[8] I discuss these passages in the Jasche *Logic*, and the crucial issues they raise for Kant interpretation, far more fully than I can do here in my paper "Acquaintance and Cognition" in *Aesthetics and Cognition in Kant's Critical Philosophy*, ed, Rebecca Kukla (Cambridge: Cambridge University Press, 2006), pp. 85-108.
[9] Kant, KrV B130.
[10] Ibid., KrV B130-131.

Sammie is a dog. It is for this reason that in the *Logic* Kant explicitly asserts that the distinguishing mark of human cognition is that is it discursive. Only we can form judgments, and, as we will see, according to Kant we can form judgments only if it is possible for the 'I think' to accompany our representations. This is the, relatively modest and circumscribed, Kantian form of the third dogma of rationalism.

In both the A and B Deductions Kant immediately follows his discussions of the consciousness of the unity of synthesis with the first introduction of the necessity of the unity of apperception. This 'I think', which must be capable of accompanying all of my cognitive representations, is itself, for Kant, a representation which embodies a consciousness of the unity of the synthesis of all that is manifold in my experience. "The synthetic proposition, that all the variety of empirical consciousness must be combined in one single self-consciousness, is the absolutely first and synthetic principle of our thought in general. But it must not be forgotten that the bare representation 'I' in relation to all other representations (the collective unity of which it makes possible) is transcendental consciousness."[11] Indeed, this 'I think' is a specific kind of representation, a 'thought'. "On the other hand, in the transcendental synthesis of the manifold of representations in general, and therefore in the synthetic original unity of apperception, I am conscious of myself, not as I appear to myself, nor as I am in myself, but only that I am. This representation is a *thought*, not an *intuition*."[12] *What* I am conscious of in this thought, this "bare representation 'I'", is the 'unity of synthesis', or combination, of my various representations. Putting this all together, the 'I' which must be capable of accompanying all of my representations is the representation of the unitary act of thinking which relates all of my various representations into a single consciousness or experience.

Kant's line of argument here seems to be as follows. What is distinctive about human cognition is the ability to represent or be conscious of the unifying or combining character of our own mental activity in a single unifying representation. Typically, such a representation itself ultimately involves a concept applied in a judgment to a synthesized manifold; e.g., 'That is a dog'. When one represents in this way, what is represented is the type of synthesizing character of one's own activity. As such, every such representing act, no matter what concept is applied, is also an act of self representing, an act in which one conceptually represents one's own combining activity. Since it is the

[11] Ibid., KrV A117n.
[12] Ibid., KrV B157.

synthetic representation of that dog which is conceptually characterized as 'dog', and that representation has that character in part in virtue of the character of the synthesizing activity that constituted that complex representation, it is one's own activity that one types when one types a representation as one of a dog. So to be capable of conceptually cognizing something as a dog, one must be capable of conceptually cognizing one's dog representations as one's own representations, in the sense that they are recognized as the product of a certain sort of combining activity on my part. What I have which my dog Sam lacks is precisely this ability to be acquainted with objects with consciousness, that is, the reflective capacity to cognize and type my own acts. That which *all* such acts of combination share in common is just that they are all my acts. But insofar as I can cognize conceptually I have the reflective capacity to type my own acts, so I have the ability to conceptually represent, to think, my own acts *as* my own acts. That is, I can conceptually cognize, or think, an object only if the thought 'I think' can accompany the act in which I think the object. For Kant, what the 'I think' is necessary for is the capacity to judge and to conceptually represent objects by forming discursive judgments about them.

At the same time the possibility of the 'I think' is also required if any representation or object is to be anything 'to me'. Something is something 'to me' only if it is recognizable by me as something which *I* am cognizing. That is, for a dog to be something to me I must be able to represent *that* the dog is being thought by me as a dog. But this possibility just *is* the possibility of representing the act in which I intend the dog as my act, that is, the possibility of the 'I think' accompanying the cognition of the dog as dog. It is thus analytically true that some thing can be something to me only if I am capable of affixing the 'I think' to its representation.

We can thus see that for Kant the ability to self-consciously represent oneself as thinking, or judging, is a necessary condition on the possibility of thinking, or judging, at all. What needs to be added to the Kantian position to generate a full-blown statement of the third dogma of rationalism is an additional commitment, the claim that no creature who is incapable of forming judgments is capable of intending objects as objects. I have argued elsewhere the Kant himself waffles on just this point. At times he speaks as if he is committed to the view that only agents who are capable of asserting judgments are capable of intending objects as objects. In other moods Kant seems to admit the possibility that non-judging

creatures are indeed capable of intending objects.[13] Regardless of whether or not the historical Kant believed that there is no intentionality without the possibility of judgment, however, there is no question that many of his twentieth century successors committed themselves to just this view. And with this commitment they arrived at the complete third dogma position that self-conscious second order intentionality is a necessary condition on *all* intentionality. In the remainder of this paper I consider the structure of the arguments that twentieth century Kantians have used to support this conclusion.[14]

Necessary Conditions on Intentionality

In the next section of this paper I will take Davidson's discussion of the necessary conditions on an agent having intentions as representative of the twentieth century version of the 'third dogma' tradition. There are three reasons for doing so. First, Davidson's discussions of this issue are clear. Second, as opposed to similar discussions in Sellars and Brandom, his discussions are brief. And third, the most distinguished contemporary Sellarsian, Robert Brandom, has explicitly endorsed Davidson's views, (though, as we will see, he adds to them), so we might with confidence abstract from whatever differences remain between Davidson's fundamentally Quinean approach and the Sellarsian version of the story. As Brandom has developed the tradition beyond the point at which Davidson stops, the following part of my discussion will focus on his views.

Davidson's argument proceeds in two steps, and each step is clearly transcendental in form. First he argues that the intentional attitude of believing plays a central role in intentional life. For Davidson, an agent has any intentions at all only if among those intentions some are beliefs. Second, he argues that having some other kind of intentional state is a necessary condition on an agent having any beliefs. As will become clear in the following, Davidson takes the class of agents that have beliefs to be coextensive with the class of agents that can form judgments, and all of his arguments for the second stage in his argument, and thus for the third dogma of rationalism, turn on that identification.

Davidson offers a variety of intentional attitudes as necessary for

[13] Cf. my "Acquaintance and Cognition" cited above.

[14] An earlier version of much of the material in this section has appeared in my "The 'I Think' and the For-the-Sake-of-Which" in "*Transcendental Heidegger*, S. Crowell and J. Malpas, eds. (Stanford: Stanford University Press, 2007) pp. 151-168.

belief. The canonical candidate for this role is the concept of belief: "... I argue that in order to have a belief, it is necessary to have the concept of belief."[15] This condition provides the focus for all of Davidson's other formulations of the conditions on belief. Because one can't have a concept of belief unless at the minimum one can group beliefs together as distinct from non-beliefs in virtue of being beliefs, in order to have a concept of belief one must be able at a minimum to believe of some beliefs that they are beliefs. That is, one must be capable of having some beliefs about beliefs if one is to have a concept of belief. "...in order to have any propositional attitude at all, it is necessary to have the concept of a belief, to have a belief about some belief." But, for Davidson, that which distinguishes beliefs as beliefs is that they are intentional states of an agent that, in virtue of their content, could be true or false. Since this is what a belief *is* no agent could have a concept of belief or recognize a belief as a belief unless she also could intend the possibility of error and understand being true and being false, and their contrast. "Someone cannot have a belief unless he understands the possibility of being mistaken, and this requires grasping the contrast between truth and error – true belief and false belief."[16] But the contrast between truth and error is grounded in the distinction between the way the world is and the way an agent takes the world to be. So no agent who lacks a concept of an objective world or of objective truth can have a concept of belief. And, finally, Davidson holds that the capacity to be surprised trades on an agent's ability to recognize that her previous beliefs were false, and so he concludes that this capacity for surprise is an infallible marker of a concept of falsity and thus of the presence of the concept of belief. So, if an agent has intentional states at all only if she has a concept of belief, then an agent has intentional states only if she is capable of surprise.

But why does Davidson hold that only agents who possess the concept of belief can have beliefs? Ordinarily one needn't have the concept of X in order to have or be X, even for intentional states. The clue to Davidson's adherence to this version of the third dogma is contained in the quote from "Thought and Talk" that I just cited. The mediating term relating having a belief and having the concept of belief is the ability to understand the possibility of being mistaken. For Davidson, it is not only the case that if an agent has the concept of belief, she also has the capacity to understand that she is mistaken. It is also the case, for Davidson, that if an agent has the ability to understand that she is mistaken, then she has the concept of

[15] D. Davidson, "Rational Animals", p. 102.
[16] Donald Davidson, "Thought and Talk" in *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1984), p. 170.

belief. And, Davidson argues, since only an agent who 'understands the possibility of being mistaken' counts as having beliefs, and only agents who have the concept of belief can understand the possibility of being mistaken, having the concept of belief is a necessary condition on having beliefs.

So Davidson's argument turns on two distinct claims. First, he holds that the ability of an agent to understand being mistaken is necessary for that agent to have beliefs. And second, he holds that having a concept of belief, and thus second order beliefs, is necessary for grasping the possibility of being mistaken. From these premises Davidson infers his version of the Third Dogma: Having beliefs about beliefs, beliefs of the second-order, is necessary for having any beliefs at all. But what supports these premises? In one sense of 'understanding being mistaken', it is easy enough to see why someone might think that having a concept of belief is necessary for understanding being mistaken. If 'understanding being mistaken' consists in forming the *judgment* that one has committed an error or has had a false belief, then an agent can understand that they are mistaken only if that agent has a concept of belief. To judge of some state that it is false I must at least take that item to be a candidate for truth and falsity, that is, I must judge that the item satisfies the concept of belief. But, then, having a concept of belief is necessary for one to judge of a belief that it is a mistake. But why do Davidson and other adherents of the third dogma think that the ability to understand the possibility of making a mistake, in this *judgmental* sense, is necessary for having beliefs? On its face, the concept of belief demands that to be a belief a state must be a candidate for truth or falsity, and this requires that the state has some 'content', specify some way the world might be, and that the state is true if the world is that way and false if it is not. The belief B that some belief A is mistaken, then, is a state that has the 'content', that specifies the possibility, that the content of A not line up with the way the world actually is. So to claim, as Davidson does and the third dogma demands, that an agent can't have beliefs without being capable of judging that some belief is mistaken, is to claim that an agent can't have beliefs about the world unless that agent has some beliefs about beliefs that specify that her beliefs about the world might be mistaken. But what is it about belief that, for the adherents of the third dogma, motivates *this* claim?

In order to answer this question, and thus to see why the adherents of the Third Dogma tradition think that second order belief is necessary for belief *tout court*, it is necessary to understand what adherents to this tradition think that beliefs *are*. I think that Haugeland points the direction to the answer to this question when he points out that for Davidson (and

Dennett, Sellars, Brandom, etc.) "...cognitive states and processes are determinate only relative to the interpretability of manifest behavior". This is the case for these thinkers (and, I might as well admit, for me as well) because for them, and for me, what a state with intentional content *is* is a state that potentially plays a certain definite role in a certain kind of interpretation and understanding of the behavior of an agent. The kind of interpretation in which such states figure is the kind of interpretation that sees what the agent does as having a *point*, as occurring *in order to* reach a goal. On this view, intentional states in general, that is, states with content, and beliefs in particular, that is, states that are to be evaluated in terms of truth and falsity, have the intentional content that they have in virtue of their roles in 'fleshing out' this kind of interpretation. So to understand why the adherents of the third dogma think that understanding the possibility of error, in the sense of being capable of judging that some belief is false, is necessary for having intentionality, we must understand how they interpret the interpretation of the purposeful activity of agents.

Action and Reasons

According to the interpretationalist view of intentional states that stands behind the twentieth century version of the Third Dogma position, only agents whose actions have a point, who attempt to achieve goals, have intentionality. This is a traditional Aristotelian view. But not all things that react differentially to their environment do so in order to achieve some end. An iron bar reacts differentially to the presence or absence of water vapor in its neighborhood (by rusting or not), but its behavior has no point. The behavior of my 13 year old, on the other hand, clearly marks her as acting so as to achieve ends of her own. There must, then, be features of the behavior of an entity that mark that behavior as having a goal and it is the presence of such features that alert us to the possibility that the entity is an agent who has intentionality and beliefs.

The features in the behavior of an agent that marks that behavior as action that has a point are holistic in the sense that they have to do with the overall structure of the behavior of the agent. We are only justified in suggesting that an agent is attempting to achieve goals by acting for reasons related to those goals when we can discern a certain pattern in the agent's behavior. At a minimum, the behavior must be generally 'successful' at resulting in certain definite states of affairs in varying situations, and what the agent does must itself vary flexibly in accordance with variations in the environment so as to result in those preferred outcomes. Take my dog Sammie, for example, which periodically picks up

a chewy toy and drops it at my feet. *Partially* on the basis of this act I am inclined to attribute to him the desire to play fetch, together with the belief that cuing me in this way will get me to play the game with him. After all, this behavior on Sammie's part frequently results in my picking up the toy and throwing it, and if this were the point of Sammie's behavior, this would be an *appropriate* thing for him to do. Now, of course, I don't always play fetch with Sam when he cues me in this fashion. Sometimes, as we say, Sam has the false belief that I will play with him if he cues me in this way. But unless what the agent, Sammie, does is *in general* flexibly suitable for bringing about certain preferred 'ends' or goals in changing circumstances we would not even consider the possibility that what that agent does has the goal of bringing about those ends.

But, as the example of the iron bar always producing iron oxide in the presence of water reminds us, not every response to a given environment that generally results in a certain outcome counts as an act that has that outcome as a goal. Acting for a goal is a phenomenon that essentially occurs over time. To count as having a point most of what the agent does must keep changing in response to variations in the environment so as to keep honing in on the goal. At the most fundamental level it is this consistent appropriateness of behavior for bringing about preferred results in shifting circumstances, even though the physically described behavior keeps changing, that marks the difference between the iron bar and my dog.

The fact that activity with a goal involves a holistically suitable response to a changing environment has enormous implications for what it is for an agent to act in order to achieve an end. For every time the agent does something what it does alters its environment. So every time the agent acts it changes how it should act in the future so as to achieve its goals. And, since an agent acts for a goal only if in general it acts suitably for bringing about that goal in changing circumstances, an agent acts for a goal only if it can in general string together various acts in such a way that the entire string of acts fits together so as to achieve that goal.

It is this fact, that to interpret an agent as acting in order to achieve an end one must interpret the agent as doing a variety of things that together serve to progressively realize that end, that provides the anchor for the interpretationist view of intentionality. As opposed to iron bars, agents don't do the same things in the same circumstances at different times. Rather, what they do at a time in a given environment varies as a function of the overall point of the agent's activity at that time. A given leopard in a given physical environment might, or might not, head for a tree full of monkeys depending on whether its goal is to eat or not. But if its goal is to

eat, what it does at each successive stage of its quest must vary in response to the way in which its own activity changes the situation, or might change the situation. Going towards the tree, for example, might alert the monkeys to its presence, and the monkeys might then take appropriate countermeasures that would make it difficult for the leopard to achieve its goal, to eat. So the leopard attempts to hide. That is, achieving the ultimate goal of eating involves the leopard in acting with a more proximate goal, to hide. But, the interpretationist argues, when the agent, the leopard in this case, acts in order to hide, there must be some fact about the agent that disposes it to act in order to hide, that is, act in ways that have the goal of hiding from the monkeys. Any agent that has *that* kind of state, that is, a state that disposes it to act in order to hide from the monkeys, can be said to have the *desire* to hide from the monkeys. The 'content' of the desire is just the 'in order to' of the acts that the desire would motivate if it motivated behavior. And, since the attempt to hide counts as having that goal only as part of a larger pattern of purposeful action in which the ends of each bit of behavior fit together into a generally effective instrumental chain, no agent can count as having a single 'desire' unless most of her desires fit together in a way that 'makes sense', that is, fit together in a generally consistent instrumental pattern.

The role of beliefs on the interpretationist model is just to pick up the slack between the 'in order to' of acts and what the acts actually achieve. There is variability in response by agents in identical situations that is not captured by differences in ends. Two leopards might be identically situated relative to a tree full of monkeys, and both might want to eat them, but they still might act differently. One might go to the left of the tree, the other might go right, and the first might be successful in catching and eating a monkey and the other fail in its attempt, perhaps because the left side of the tree is downwind. For the interpretationist there must be something about the two leopards that explains this difference in action, but by hypothesis it can't be that they have different desires or goals. For the interpretationist, what is different is the 'beliefs' of the two agents. It is possible for agents with the same goals to act differently in identical situations because not everything an agent does to achieve an end need be successful for that agent to be acting so as to achieve that goal. That is, an agent can be acting for some end without reaching it. All that is necessary for an agent's behavior to have a point is that most of what it does is successful at achieving its goals. But then much of what it does can fail, that is, result in states that it does not desire. Now, when an agent is successful its acts mesh with the actual environment so as to bring about the agent's preferred end. But this is not the case with failed acts. On the other hand, had the initial state of the world at the time that the agent acted been different in some definite way, what the agent did *would* have resulted in success. So, in our example, what is common to the actions of the successful and unsuccessful leopard is that both of those behaviors would result in the leopard eating a monkey under some definite wind conditions. That is, the leopards' actions share a goal and the leopards share the desire to eat monkeys. What is different is that those conditions under which the behavior would be successful are different and only one of those set of conditions is actual. That is, the first leopard has true beliefs regarding its environment and the second has false beliefs, but both act in ways that would be successful were their respective beliefs true. So beliefs are states of agents that interact with desires to help to interpret the variations in the behavior of an agent in a given environment. The 'content' of a belief is the possible state of the world in which, if actual, the agent's act would be successful at achieving its goal.

The fact that agents who act in order to achieve ends generally act in interconnected ways that over time are flexibly suitable to achieve those ends in changing environments has a further, crucial consequence. Such agents must be good at 'figuring out' the potential consequences of their acts. Otherwise they could not perform in a manner appropriate for achieving their ends as conditions vary in response to their own activity. But to do this amounts to coming to have new beliefs about the world that are themselves *appropriate*, or *justified* by the agent's prior beliefs and the sensory information concerning its environment that is available to it. For example, over time I find that Sammie drops toys at my feet far more frequently than he drops them at the feet of my son, and this makes sense if Sammie believes that I am more likely to play with him than my son is. But that he believes this in turn makes sense in light of the fact that in Sam's past experience I have been far more likely to play with him than my son has. That is, if Sammie had been rational he would have inferred, from his prior beliefs and sensory experience, just this belief, that I, but not my son, would respond to this cue on this occasion. Generalizing, we can see that to interpret an agent as acting for goals we must be able to assign beliefs and desires to that agent that fit together into a pattern that reflects proper inferential connections, If what an agent does has some point, then that agent's intentional states must be related to each other in inferentially appropriate ways.

We can thus see that according to the interpretationist model for understanding intentionality an agent acts in order to accomplish goals only if it is possible to interpret that agent as having beliefs and desires the contents of which are both capable of explaining the overall activity of the

agent and are interrelated to each other in inferentially appropriate ways. As Haugeland says, for the interpretationist, "cognitive states and processes are determinate only relative to the interpretability of manifest behavior'. This is because on this model the content of intentional states just is the peculiar role of those states in the interpretation of the agent's behavior as goal directed. But why, then, is the paradigmatic interpretationist, Donald Davidson, committed to the view that an agent cannot have intentional states without understanding the possibility of being mistaken? Nothing in the story we have been telling seems to require that some of the intentional states that are assigned to an agent in the course of teleologically interpreting her behavior must intend other states as intentional, and thus intend them as potentially mistaken.

But the story we have been telling is incomplete. A list of necessary conditions is not a sufficient condition. If an agent acts for a goal then it must be possible to interpret that agent as acting as if it had intentional states that are linked in inferentially appropriate patterns, but it doesn't follow from it being possible to so interpret an agent that that agent is acting in order to accomplish goals. In addition, the agent must be acting as she does *because* she has those intentional states. In fact, Davidson, Sellars, and Brandom implicitly appeal to an additional requirement on intentionality that they assert that non-self-conscious agents cannot satisfy: Any agent that acts for reasons, and thus possesses beliefs and other intentions, must act *because* of her reasons.

Consider a case that is strictly analogous with Sammie's fetch inducing behavior. I often ask my son to play catch with me, but I less frequently ask my daughter to do so. I *think* that I do this because I have evidence that my son is a better bet for this cuing behavior then my daughter is. After all, when asked, he has played with me in the past more frequently than my daughter has, or so I believe. And that justifies my current belief that my son is a better bet. But even if it is true that in the past Nick has played catch with me more frequently than Valerie it doesn't follow that this is the correct explanation for my behavior on this occasion. In fact, the correct explanation might be simply that I am sexist, and have a stereotypical and false view of women, and this view *causes* me to act as I do. That is, it doesn't follow from the fact that my beliefs line up in inferentially appropriate patterns, that they do so *because* those beliefs are appropriately inferentially connected. And it doesn't follow from the fact that what I do is what I would do if I were acting rationally, that what I do I do because I am acting rationally. And, finally, it doesn't appear to follow from the fact that I can assign beliefs and desires to Sam in an appropriate inferential pattern so as to explain his actions, as if he has

reasons for what he does, that he in fact does what he does because of his reasons.

The moral of this little story is that an agent acts for reasons, and thus is a candidate for having beliefs, only if she does what she does because of her reasons. And it does not seem to follow simply from the fact that an agent acts *as if* it acted for reasons that it acts *because* of those reasons. It is this condition, that rational agents must act because of their reasons, Davidson (and Sellars, and Brandom) think non-self-conscious agents can't possibly meet. To see why they think this we must look briefly at Kant's analysis of what it is to act because of a reason.

Self-Correction And The Conception Of Law

Because agents whose behavior has a point must generally act appropriately to achieve their ends in changing circumstances, and circumstances change in response to their own ongoing behavior, such agents must act differently when they succeed from how they act when they fail. The leopard that walks west in order to reach a hole that contained water last week for the sake of drinking acts differently if the hole still contains water from how she acts if it is empty: in the first case she laps with her tongue, in the second she moves with her legs. For any successful teleological agent, such differential behavior must correlate pretty well with the correctness or the mistakenness of its prior action, where success is measured by the agent's goals, and mistakes are responded to appropriately so as to achieve those goals. But an agent can be self-correcting in this way to a fairly great extent and still not be acting *because* what it does is a response to its own mistakes as mistakes. And it is only agents who vary their behavior because, or *for the reason that*, they have made a mistake (or not), that are candidates for intentional ascription. Consider the Sphex wasp that was made philosophically famous by Dennett:

> When the time comes for egg laying, the wasp Sphex builds a burrow for the purpose and seeks out a cricket which she stings in such a way as to paralyze but not kill it. She drags the cricket into the burrow, lays her eggs alongside, closes the burrow, then flies away, never to return. In due course, the eggs hatch and the wasp grubs feed off the paralyzed cricket, which has not decayed, having been kept in the wasp equivalent of deep freeze. To the human mind, such an elaborately organized and seemingly purposeful routine conveys a convincing flavor of logic and thoughtfulness - until more details are examined. For example, the Wasp's routine is to bring the paralyzed cricket to the burrow, leave it on the threshold, go

> inside to see that all is well, emerge, and then drag the cricket in. If the cricket is moved a few inches away while the wasp is inside making her preliminary inspection, the wasp, on emerging from the burrow, will bring the cricket back to the threshold, but not inside, and then will repeat the preparatory procedure of entering the burrow to see that everything is all right.[17]

The punch line of the story, of course, is that if the biologist is sufficiently persistent she can induce the wasp to repeat the pattern without end so that she never lays her eggs. Now, although I have used this story for other purposes at other times,[18] in this context the moral of the story is just the moral that Dennett wishes to draw. An agent can act as if she is acting for reasons without in fact acting because of her reasons. We know that the wasp is not acting because of reasons, even though it seems that she might be, because she is entirely unresponsive to the singular fact that what she does when she moves the cricket back to the doorstep is completely unsuccessful in furthering her ends. That is, she does not treat the *failure* of her acts to achieve her proximate ends as a *reason* to alter her behavior. If this failure gives us reason to believe that the wasp is not acting because of reasons, as it does, then we have reason to think that only an agent who corrects her mistakes because they are mistakes can act because of her reasons. But only an agent that acts because of reasons can have beliefs. So it is a necessary condition on an agent having beliefs that she act to correct her mistakes because they are mistakes. That is, the ability to respond to the mistakenness of actions as reasons for self-correction must be implicit in the practical behavior of an agent for that agent to be a candidate for rationality or intentionality. But the wasp's repeated routine in the face of recurrent failure indicates that she is not responsive to the *mistakenness* of her own behavior. So it follows that she has no beliefs.

We have now reached the bedrock intuition that stands behind the third dogma of rationalism. And the intuition itself is not an error. The intentionality of an agent demands a differential responsiveness to error as error on the part of the agent. But the implications of this intuition are not quite as clear as they might seem. For this assertion is not quite equivalent to the premise for Davidson's version of the third dogma that we have found in "Thought and Talk".

[17] Dean E. Wooldridge, *The Machinery of the Brain* (New York: McGraw-Hill, 1963), 82.

[18] Mark Okrent, *Rational Animals: The Teleological Roots of Intentionality* (Athens: Ohio University Press, 2007).

That premise was specified as follows: "Someone cannot have a belief unless he understands the possibility of being mistaken, and this requires grasping the contrast between truth and error..." The requirement on belief that we have uncovered, on the other hand, is that someone cannot have a belief unless the ability to respond to mistakes as mistakes, that is, as reasons for self-correction, is implicit in her activity. These are equivalent conditions only if an agent can implicitly respond in her practical activity to her mistakes as mistakes just in case she *understands the possibility of being mistaken*, and that this is impossible unless the agent *grasps the contrast between truth and error*. Now, understanding the possibility of being mistaken is a reflexive, second order intention that does indeed require grasping the contrast of truth and error, that is, intending an intention specifically as correct or incorrect. So if the capacity to respond to mistakes as mistakes can be implicit in behavior only if the agent understands the possibility of being mistaken, then the capacity for such second order intentions is indeed a necessary condition for intentionality. But why would one think this?

For an agent to respond to a mistake as a mistake is for the agent to take the fact that the mistake is a mistake as a *reason* to alter its behavior accordingly. We know that the wasp does not take the fact that the replacement of the cricket in its original position does not permanently alter the situation to be grounds for her to change her pattern because she does not change her pattern in the face of her failures. She does not treat her failure as a reason to change. So if we could discover what it is for an agent to treat a situation as a reason to act in a certain way we could also discover what it is for an agent to respond to a mistake as a mistake, and thereby discover what it is for an agent to satisfy the behavioral conditions on intentionality. Perhaps the clearest and best worked out suggestion regarding what it is to treat a situation as a reason is found in Kant's analysis of practical reason, and it is this analysis that leads to the third dogma of rationalism, by way of the implications of this analysis for the issue of what it is for an agent to treat a mistake as a mistake.

In summary, here is Kant's suggestion. "Everything in nature works according to law. Only a rational being has the capacity of acting according to the conception of laws, i.e., according to principles. This capacity is will. Since reason is required for the derivation of actions from laws, will is nothing else than practical reason."[19]

"Everything in nature works according to law." But according to Kant,

[19] I. Kant, *Foundations of the Metaphysics of Morals* (G), trans. L.W. Beck (Indianapolis: Bobbs-Merrill, 1959), p. 29.

the rational actions of rational beings, as rational acts, satisfy a different condition, although the condition also has to do with a relation to a rule or law. When rational agents act rationally what they do is according to a *conception* of law. That is, the rational acts of agents are mediated by the agent's conception of or grasp on a law. For example, I find myself to be angry with my daughter, but I recognize the principle that parents love, support, and do not hurt their children. I understand myself to be a parent. So I support my daughter, rather than hurting her, as I would if I were caused to act by my anger. I act as I do *because* I accept that I am a parent and I acknowledge the law that parents support their children. As Robert Brandom puts Kant's point, "What makes us act as we do is not the rule itself but our *acknowledgement* of it."[20]

It is crucial to note that on this analysis the normative force of the principle, what makes it a reason for the agent to act, is wrapped up with the agent's acknowledgement of the law as a law for him. Since it is the acknowledgement of the law that accounts for the action, if the agent is acting because of his reasons, it is the fact that the agent accepts the principle as applying to him that is the rational motive for the act. This fact is displayed in the distinctive status of the law that I acknowledge, and the equally distinctive status of my acknowledgement of that law. The law I accept is not descriptively true, and, what is more, I know it to be descriptively false. Nevertheless, what it is to be a parent entails that anyone who is a parent ought to act according to this principle, and my self-understanding as a parent, together with my acknowledgment of the law, *is* an acceptance of the obligation, as a parent, to act according to the law. The concept of a parent, which conditions must be satisfied for someone to count as a parent of someone else, combines in exceedingly complex ways certain factual biological relations, which are neither necessary nor sufficient for someone to be a parent, with certain social roles and functions that parents ought to fulfill, (a fact that is especially obvious to me, as one of my children is adopted). To even understand oneself to be a parent one must understand that one stands under these defining obligations, even if one need not fulfill them, and even if one can be a parent without understanding oneself as a parent. The fact that I am a parent has normative force, provides me with a reason to act, only insofar as I acknowledge myself to be a parent and understand what it is to be a parent in terms of a set of principles of parental action.

But as Kant makes clear in the continuation of the famous passage

[20] R. Brandom, *Making It Explicit* (Cambridge: Harvard University Press, 1994), p. 31.

from the *Groundwork* that I have been discussing, on his analysis, while such a conceptual understanding is necessary for the possibility of rational action, it is not sufficient. Rational action occurs *because* of the agent's reasons. And in Kant, this 'because' is understood in a distinctive fashion. For an action to be rationally motivated by the agent's reasons, for the action to occur because of the reasons, it is not enough that those reasons *cause* the action, or *explain* the action. Rather, the action must be *derived* from the reasons. Kant's name for this ability of rational agents to derive actions from laws and reasons is 'will', and will is essentially the rational capacity to infer a mode of action from a law. "Since reason is required for the derivation of actions from laws, will is nothing else than practical reason." That is, on Kant's analysis, no agent is capable of acting for reasons unless she is also capable of inferring that she ought to act in a certain way from a set of propositions that include a law and an acknowledgment that she stands under that law.

Now, if what it is for an agent to act for a reason is for that agent to rationally infer from her beliefs regarding a situation and some law that one ought to act in a certain definite fashion, as Kant holds, and one acts for a reason just in case one treats the mistakenness of mistakes as a reason to act, then Kant also offers us a theory regarding what it is to treat the mistakenness of mistakes as a reason to act. On this Kantian view to treat a mistake as a mistake and to alter one's behavior accordingly is just to explicitly *infer* some mode of action from one's acknowledgment of the fact that one has made a mistake together with one's acceptance of some law. That is, the paradigm case of correcting a mistake is the inference from the judgment that some belief is false to the assertion of that belief's negation, via the law of the excluded middle. On this rationalist view to act because of one's reasons, and thus to have intentionality, an agent must be capable of recognizing a mistake as a mistake so that the mistakenness of the mistake can serve as a *premise* in an argument. But an agent can be capable of such recognition only if she has a concept of belief and beliefs about beliefs. So Davidson's form of the third dogma of rationalism follows directly from the interpretationist understanding of intentionality combined with the Kantian account of what it is to act for reasons. Roughly, the argument goes as follows. From the first Critique one gets the premise that no agent is capable of forming judgments without being capable of second order intentions, of affixing the 'I think' to its own representations. From the *Groundwork* one gets the premise that no agent can act for reasons unless it is capable of forming judgments. And from a certain development of the interpretationist analysis of belief, one gets the premises that no agent can have beliefs unless it is capable of responding

to its mistakes as mistakes, and no agent can respond to its mistakes as mistakes without being capable of acting for reasons. From all of this the third dogma of rationalism validly follows: No agent can have intentional states if it is incapable of having second order intentional states that intend its own intentional states as its own intentional states.

Evaluating The Third Dogma

My immediate target here is the premise of this argument that has the Kantian analysis of acting because of a reason as its point of origin. On that account, to act because of a reason is essentially to derive a consequence from a law together with an appreciation of a situation. On the Kantian view, to act for a reason an agent must be capable of correcting her mistakes and to correct one's mistakes is to recognize that some situation, belief or act fails to satisfy some law or principle, and to infer some new belief or act from this failure. But to recognize that some belief fails to satisfy some principle is to form a judgment, and judgments essentially involve second order intentional attitudes, so only agents who can have such second order acts can correct mistakes, act because of reasons, or have any intentions at all.

Unfortunately for the proponents of the third dogma, as an analysis of what it is to act because of a reason this Kantian account is wrong. The late Wittgenstein has given us a strong reason to believe that deriving action from law can't be the only way in which an agent can be responsive to, or act because of, a reason. The argument is given in the context of Wittgenstein's discussion of language use, is familiar, and I won't belabor it here. Roughly, the problem has to do with a regress regarding the application of laws to situations, or what Kant calls our faculty of judgment. On Kant's view, when we act for reasons, that action is mediated by our understanding of a law or rule, that is, by our interpretation of the law. Given Kant's own understanding of concepts as rules for sorting, understanding a law amounts to knowing how to apply it to given situations. Such understanding necessarily involves knowing what one ought to do in different situations. One can rationally act in accordance with a rule only if one understands the rule, but there are correct and incorrect ways to understand the rule. Now, if all rational action, or action because of reasons or norms, demands that one act because one acknowledges a rule, then one can correctly adopt an interpretation or understanding of our first rule, from among alternative deviant ways of understanding the import of the rule, only if one has a reason to adopt that understanding, which, on the analysis, amounts to

choosing this interpretation because one acknowledges some other rule that prescribes this choice. But, by symmetry of reasoning, our application of this further law demands a prior understanding of some third law, and off we go.

The moral of this little Wittgensteinian story is that acting because of reasons cannot be explicit inference all the way down. If anything we, or anyone else, does is done because of our reasons, then some of what we do must be done because of reasons that are not fully articulate, and we must be able to act because of reasons that we do not use as premises in explicit arguments. There must be a way to act because of reasons that does not demand explicit inference.

It would seem to follow from this conclusion, that since there must be a way to act because of reasons that does not demand explicit inference, that there must be a way to respond to mistakes as mistakes that also does not require explicit inference, as there is no acting for reasons without responsiveness to mistakes as mistakes. But from this it would seem to follow that it is also possible that there might be agents with beliefs who lack any beliefs about beliefs, as the only justification we have found for this dogma is the view, now seen to be false, that there is no possibility of responding to mistakes as mistakes without the kind of explicit inference that requires second order beliefs. Nevertheless, some of those who accept the Wittgensteinian argument we have been considering still deny that non-reflective agents are capable of having beliefs, or acting because of reasons. In particular, Robert Brandom emphasizes just this point derived from Wittgenstein. Nevertheless, he still uses a Kantian style analysis of acting because of reasons to deny beliefs and other intentional states to non-reflexive agents. The remainder of this paper is devoted to a consideration of the structure of his argument.

Brandom draws two conclusions from the above line of Wittgensteinian argument. First, in the terms of our own discussion here, there must be a way of acting because of a reason that is implicit in the practice of an agent, rather than demanding an explicit inference from premises by that agent. (Brandom's own way of putting the point has to do with the necessity that there be some way in which an agent is responsive to norms in their practice that does not require explicitly understanding propositional claims.) Second, he insists that such a practical grasp of reasons is best understood in terms of *assessments* of propriety. "...there is another move available for understanding what it is for norms to be implicit in practices. This is to look not just at what is done – the performances that might or might not accord with a norm (appropriate or inappropriate) – but also at *assessments* of propriety. These are attitudes of

taking or treating performances *as* correct or incorrect."[21] That is, Brandom's suggestion is that the primary, non-explicitly inferential way in which an agent is acting because of a reason, or (in Brandom's terms) is acting because she is responsive to a norm, is by *assessing* an act as correct or incorrect. This amounts to the view that responsiveness to reasons can implicitly manifest itself in the practical behavior of an agent by the agent somehow directly assessing an action as mistaken, or, in Brandom's terms, incorrect, without that assessment being mediated by an explicit cognitive inference.

This is the linchpin of Brandom's discussion of these issues. From here he argues two successive points. First, he argues that the norms or reasons in light of which we assess the actions of others are only instituted as reasons or norms by the very attitudes of assessment we apply. For Brandom, what we ought to do is what we ought to do *only* in virtue of the fact that it is *acknowledged* as what we ought to do in the practices of assessment. This is Brandom's residual adherence to the Kantian analysis of acting because of a reason. Second, he infers from this that such assessment is primarily social in character: Assessment of correctness and incorrectness of performance is something we do to and for each other.

Here is how the argument goes. Since Brandom agrees with Kant that reasons only have normative force insofar as they are acknowledged, it follows that *any* behavior by a single agent that merely accords with a norm but does not include explicit acknowledgement of the reason fails to be an action performed because of the norm, precisely because the agent lacks such an acknowledgement of the reason. Nevertheless, according to the Wittgensteinian argument that we have been discussing, there *must* be some way that acknowledgement of norms can be merely implicit in the practice of an agent. So Brandom thinks that we are faced with an aporia.

Brandom thinks that the *only* way in which this aporia can be overcome is by appealing to a social context. Each of the agents in a community responds to the acts of others as if those acts were correct or incorrect according to some rule by approving and sanctioning those acts. So, for example, I might intervene and punish any member of my group that I observe crossing a street when a light is red, and other members of my community might intervene in a similar way. That we do so is a function of the kind of training that we have undergone in the past combined with the fact that we are the type of social creatures that we are. But, given that for Brandom reasons are reasons only by being

[21] R. Brandom, *Making It Explicit* (Cambridge, Harvard University Press, 1994), p. 63.

acknowledged, and none of us, my self or my peers, are capable of acknowledging the principle 'don't cross street when light is red', because none of us is capable of explicitly formulating the principle, none of this really counts as acting because of reasons. Nevertheless, each new member of the community is trained by the same sanctioning practices that trained us to act *as if* they were responsive to the reason for not crossing at a time that is partially embodied in the rule concerning red lights. Similarly, the same kind of sanctioning and rewarding practices could also be efficacious in training the members of a community of the right type of animal to engage in *vocal* activities, activities that were structured *as if* the members of the community were responding to one another's utterances for appropriate inferential reasons. So, to continue our example, at this stage the members of the community not only sanction individuals who cross when the light is red, and reward those who don't, (thereby training the individuals both to act as if they are following a rule and to get those individuals to train yet other individuals in the same way), but the members of the community now also train each other so as to *say* 'light is red' when it is red, and 'no cross the street', both when someone refrains from crossing and when the light is red. For Brandom, this still does not count as acting for a reason, for as yet there can be no explicit acknowledgement of any rule or reason. But once the members of the community have reached the stage at which their verbal practices amount to acting as if they were following the logical rules for the conditional, everything changes. At this point, those practices themselves allow the members of the community to make the practices they are following explicit. For at this stage, it becomes possible to *say* that 'If the light is red, don't cross the street', and, given all of the background training involved in reaching this point, this allows for the possibility of explicitly *acknowledging* this as a principle to be followed, by treating this statement as a premise in an argument that results in an action as a conclusion. At that point, the back of the aporia is broken, and it becomes possible for the members of the linguistic community to become truly rational, because they are now capable of really acting because of reasons by inferring correctness and incorrectness of performances from rules in the approved Kantian manner. What had been mere animal conditioning becomes acting for reasons by the very act of making explicit the rules that the agents have been conditioned to follow so that those rules can be acknowledged as norms.

The aporia that Brandom constructs turns on the incompatibility of two claims, the Kantian assertion that if an agent is acting for reasons then that agent must infer its action from the acknowledgement of a principle, and

the Wittgensteinian observation that if *all* actions for reasons obeyed the Kantian dictum, we would be faced with an impossible regress of reasons. Brandom's solution is to posit a set of community practices that train members of the society to act as if they are acting for reasons, even though they do not act *because* of those reasons. That is, Brandom dissolves the aporia by keeping the Kantian notion of acting because of a reason in full force, but recognizing that all acting because of a reason has as its necessary condition that the agents who act for reasons have been trained, by other members of their animal communities, to act *as if* they had reasons for what they are doing, *even though, strictly speaking, they do not.*

Now, whatever one thinks about the prospects for this way of dissolving Brandom's dilemma, there is a prior issue about whether the aporia that this line of argument is meant to resolve is a real dilemma at all. There is, after all, a problem to be solved here only if Brandom is right in agreeing with Kant that a reason has normative force only if it is acknowledged as a reason. But there is good reason to think that this claim is implausible. First, consider the fact that in many of the paradigm cases in which we take ourselves to be acting for reasons we simply cannot even *formulate* the rule or norm the acknowledgement of which on this Kantian view constitutes an essential part of our reason for acting. If we could formulate these rules it would be much easier to pass the Turing test than it has turned out to be. But if in many paradigm cases, such as language use, we can't explicitly formulate the principles we are following in acting because of our reasons, how can it be the case that the explicit acknowledgement of those principles is necessary for an agent to act because of a reason?

Beyond this fact, there are familiar cases in which we respond to mistakes as mistakes, and *self-correct* our behavior, solely on the basis of our perceptual interaction with the world, short-circuiting the supposed necessity for acknowledgement of the reason as a reason, without interfering with the fact that we are acting because of reasons. Consider the following example taken from Michael Tye's discussion of Davidson's version of the third dogma of rationalism. "Suppose, for example, I believe that my car had been stolen upon finding it missing in the carpark. I start to walk to the security office on campus. As I do so, I see my car parked on the other side of the street, and I suddenly remember that I left it there on this occasion. No longer believing that my car has been stolen, I change direction and head directly to it. Here surely I revise my beliefs in light of my perceptual evidence and thereby my behavior. But I need not have any (explicit) belief about a belief. In this sense, I need not (explicitly) recognize my mistake: I need not consider my belief that my car has been stolen as such at all. I am certainly acting for reasons, however; and as my reasons change, by behavior changes too."[22]

Now what is it that makes it so obvious to Tye, and to us, that when he changes his behavior here in response to a perceptual encounter that he is 'certainly acting for reasons'? The answer, I think, is clear. The perception provides perceptual *information* about the world, and this information leads to a *self-assessment* of Tye's *own* behavior, and a correction of that behavior as wrong. Assessment begins at home. The primary form of assessment is the self assessment of an agent's own behavior as appropriate or inappropriate in light of the ends of the agent and the changing state of the world. The very pattern of action that provides the sine-qua-non for attribution of purposeful action at all, flexibly suitable behavior by the agent for achieving her goals, in changing circumstances, guarantees that the agent is self correcting, and in that sense, self assessing. It is of course true that any agent must be capable of responding appropriately to the mistakenness of mistakes. But any agent that acts as if it is responsive to reasons must also satisfy this condition as well, at least to the extent necessary in order to display suitably flexible behavior for achieving its minimal goals. The normativity of reasons is dependent upon the goals that direct purposeful action. It is a mere dogma of rationalism that reasons only become reasons if they are reflexively acknowledged as such. And failing this dogma, there is no reason to accept the further rationalist dogma that there is no intentionality without second order intentions.

As we have seen, the bedrock intuition behind the third dogma of rationalism is that only agents who can respond appropriately to mistakes as mistakes can have intentions. Traditional rationalism interprets this requirement as demanding that only agents that reflexively understand the concept of being mistaken can be responsive to mistakes as mistakes. But this is an overly intellectualized way of understanding the self-assessment and self-correction that is inherent in all purposeful action. One of the great lessons to be learned from Heidegger is that purposeful action, understanding, and self-correction are primarily practical and circumspectively perceptual, and that for that reason specific inference and reflexive self-awareness, self-assessment, and self-correction are late secondary growths.

---

[22] M. Tye, *Consciousness, Color, and Content*, (Cambridge: MIT Press, 2000), pp. 177-178.

Self-assessment and self-correction comes in many different degrees, of course. At one extreme lie adult, linguistically competent human beings who can explicitly formulate laws that specify correct behavior and can correct the actions (as well as assessing those laws themselves), of themselves and others in light of those principles. At the other extreme lie plants and simple, genetically programmed animals such as Dennett's sphex wasp who, while capable of achieving apparent ends in standard conditions, have little flexibility for change of behavior in light of differing circumstances. As such organisms fail to display much in the way of self-assessment and self-correction in light of sensory evidence, they certainly do not merit the honorific of 'rational', and there is little or no reason to attribute intentional states to them. But there is a wide range of intermediate cases. And any agent who is capable of self-assessment and correction in the light of changing evidence, that is, any agent capable of learning from its mistakes, is also capable of acting in response to reasons, and thus capable of intentionality.

Epilogue

Doesn't the suggestion that I have advanced, that any agent who is capable of flexibly altering her behavior in the light of changing circumstances so as to achieve her goals counts as assessing that behavior and thus counts as acting because of her reasons, fail to account for the difference between acting because of a reason and acting as if one had reasons for what one does? I don't think so. I think that here Kant and his rationalist successors have fallen into the same kind of scope fallacy that afflicted Descartes, a fallacy that Kant and these same successors have done a great deal to reveal. As I am sure that you recall, Descartes argued that since any given belief might be false, it follows that all beliefs might be false simultaneously. But, as Davidson in particular often pointed out, this inference doesn't go through. It is only against a background of massive truth that the possibility of error can emerge. Without that background there is neither truth nor error, because the agent neither has beliefs nor acts because of her reasons. It is rather ironic that Davidson himself seems to have fallen into the same fallacy. He implicitly argues that because any given act of a self correcting non-reflective agent might not be performed because of the agent's reasons, that all of those acts simultaneously might not be performed because of the agent's reasons. But the same charity considerations which preclude the possibility of massive error of belief also preclude the possibility that all of a successful agent's acts might be only 'as if' performed because of her reasons. The

successful agent who learns from her mistakes, that continues to update her behavior so as to increase her success and correct her failures, just *is* acting in response to reasons, acting because of her reasons. For this is what it is to act because of reasons.

Works Cited

Brandom, Robert. *Making It Explicit.* Cambridge: Harvard University Press, 1994.

Davidson, Donald. "Rational Animals." In *Subjective, Intersubjective, Objective*, by Donald Davidson, pp. 95-105. Oxford: Oxford University Press, 2001.

—. "Thought and Talk." In *Inquiries into Truth and Interpretation*, by Donald Davidson, pp. 155-170. Oxford: Clarendon Press, 1984.

Haugeland, John. "Two Dogmas of Rationalism." Unpublished paper, delivered at The Sixth Annual Meeting of The International Society for Phenomenological Studies, Pacific Grove, CA, July 16, 2004.

Kant, Immanuel. *Critique of Pure Reason.* Trans. Norman Kemp Smith. London: Macmillan, 1968.

—. *Foundations of the Metaphysics of Morals.* Trans. L.W. Beck. Indianapolis: Bobbs-Merrill.

—. *Lectures on Logic.* Ed. J.M. Young. Cambridge: Cambridge University Press, 1992.

Okrent, Mark. *Rational Animals: The Teleological Roots of Intentionality.* Athens: Ohio University Press, 2007.

—. "Acquaintance and Cognition." In *Aesthetics and Cognition in Kant's Critical Philosophy*, ed., Rebecca Kukla, pp. 85-108. Cambridge: Cambridge University Press, 2006.

—. "The 'I Think' and the For-the-Sake-of-Which." In *"Transcendental Heidegger*, ed. S. Crowell and J. Malpas, pp. 151-168. Stanford: Stanford University Press, 2007.)

Tye, Michael. *Consciousness, Color, and Content.* Cambridge: MIT Press.

Wooldridge, Dean. *The Machinery of the Brain.* New York: McGraw-Hill, 1963.