

# THE IMPORTANCE OF BEING ERRONEOUS: PROSPECTS FOR ANIMAL INTENTIONALITY

David Beisecker

## I. A PLEA FOR ECUMENICISM

The question of animal belief (or animal intentionality) often degenerates into a frustrating and unproductive exchange. Foes of animal intentionality point out that non-linguistic animals couldn't possibly possess the kinds of mental states we linguistic beings enjoy. They claim that linguistic ability enables us to become sensitive to intensional contexts or to the states of mind of others in a way that is unavailable to the non-linguistic, and that would be necessary for proper attributions of intentionality. To attribute mental states to non-linguistic brutes, no matter how natural it comes to us, would be grossly anthropomorphic. In the face of these challenges some friends of animal intentionality have attempted to show that at least a few animals (chimpanzees, vervet monkeys, honeybees) are capable of engaging in quasi-linguistic, communicative practices that ought to be accorded at least a minimal degree of intentionality. Others have questioned the foes' necessity claims; linguistic ability, claim these animal friends, isn't required for sensitivity to intensional contexts, surprise, or belief about belief after all, or if it is, then these features aren't really requisite for mental capacity. Indeed, if we focus exclusively upon linguistic ability, then we are apt to miss the primitive kinds of mental capacities from which our own full-blooded intentional capacities likely evolved. Animals certainly seem to interact intelligently with their surroundings, so much so that we ought to follow our natural (brute?) anthropomorphic inclinations to credit them with minds. Failing to recognize their genuine intentional capacities would be "brutishly" anthropocentric.

An ecumenically-minded witness to this exchange would maintain that both sides are partly correct. On the one hand, there is such a thing as "animal" or possibly "pre-linguistic" intentionality. On the other, this type of intentionality isn't compellingly like the sort of intentionality that those blessed with linguistic capacity enjoy. The friend of animal intentionality would be guilty of anthropomorphism only if

we unfairly restrict our attention to intentional states of the second, higher-grade kind. Thus the frustrating exchange over the status of animal belief can be seen to be generated by a tacit assumption that there is just one sort of intentionality, or a single most basic or “original” type from which all other sorts derive, or to which they reduce. So the ecumenical solution would be to acknowledge that there are at least two sorts of intentionality in play, and that each of them deserves to be associated with mental capacity. Moreover, there is no reason to think of either type as more real or genuine than the other.

While such an ecumenical resolution to this debate is certainly tempting, it doesn’t come for free. It is incumbent upon those who would favor the ecumenical route to provide an account which makes sense of talk about varieties of intentionality. Many contemporary philosophers of mind have failed to rise to this challenge. In this paper, I shall do better.

## II. INTENTIONALITY: IT’S ABOUT MAKING MISTAKES

The intentionality thought to be so definitive of mental states is typically glossed in terms of *aboutness* or *directedness* toward objects. The term ‘intentionality’ derives from a Latin word meaning roughly “to aim” - as one might do with a bow. Hence tradition has it that intentional relations are those that hold between items in the mind (ideas) and the possibly non-existent external objects or possibly non-actual states of affairs at which these items are directed. The picture is that intentional states are objects of the mind that are directed at (are about) things and happenings in an external world. Taking their task to be that of describing this “aboutness,” several philosophers of mind have tried to construct intentional relations out of various types of causal or informational connections between signals in the head and objects out in the world.<sup>1</sup> The challenge for these theorists has been to explain the evident possibility that mental states could be directed at or be about non-actual states of affairs or things that don’t actually exist in the subject’s surroundings.

Animals certainly display “directedness” towards objects in their surroundings. So it seems that this approach would be friendly to the idea of animal intentionality. But then again, things we’re not

prepared to credit with thought—for example, heat-seeking missiles and sunflowers—also exhibit directedness towards objects. The challenge then is to find a way to distinguish the special sort of directedness possessed by bonafide thinkers from the more primitive kinds exhibited by these simpler systems. Perhaps we should regard genuine thinkers as directing themselves towards objects *as objects*. But what does that amount to? Rather than shedding light on what intentionality is, this proposal simply relabels the problem. Furthermore, it threatens to require thinkers to have command of sophisticated metaphysical concepts, such as that of an object. So this approach might not be as friendly to animal intentionality as one might have supposed.<sup>2</sup> In general, attempts to unpack intentionality in terms of aboutness seem too obscure to support satisfying taxonomies of mental phenomena.

We can make more progress in distinguishing varieties of intentionality by focussing upon the *normativity* of intentional phenomena. From this perspective, the hallmark of intentional states is their susceptibility to evaluation. For instance, a doxastic (or belief-like) state can be correct or mistaken depending upon whether or not some state of affairs, identified as its *content*, actually obtains. Believers are beholden to the way things are. Similarly, conative states (goals, desires, and other “pro-attitudes”) may be fulfilled or unfulfilled, depending upon the satisfaction of some content condition. So intentional states are those that are associated with conditions of satisfaction or fulfillment. Indeed, the logical features Roderick Chisholm took to be so definitive of intentional locutions would seem to be products of this susceptibility to evaluation.<sup>3</sup> For instance, Chisholm noted that intentional locutions are typically not truth-functional. The content clauses of true belief ascriptions, for example, could be either true or false, which is precisely the distinction between a subject's having a correct or having a mistaken belief. Chisholm also noted that the content clauses of belief ascriptions could refer to non-existent objects. Once again, such circumstances typically arise in the context of a mistaken existential commitment. Curious George might mistakenly believe there to be a leopard in the branches overhead when there is no such thing. He might even fear this non-existent leopard and desire to flee it, or have yet other mental states presupposing this mistaken existential commitment. Finally, the oft-discussed semantic opacity (or intensionality) of propositional-attitude ascriptions arises because it is possible for a subject to become ontologically confused, and thus have incompatible beliefs about the same object. As Frege put it, we want to credit

terms referring to the same object with different senses, in order to accommodate the possibility that a person might mistakenly fail to endorse a true identity.

The point is that the states and the performances (or *actions*) of intentional beings are *accountable to standards*. A system is properly credited with intentional states only if it can be judged as correct or mistaken with respect to some standard or purpose, or according to some intention (hence the term ‘intentionality’<sup>4</sup>). Different species of intentionality thus correspond to different ways in which a subject is susceptible to evaluation.<sup>5</sup> If this is right, then the question of animal belief boils down to one of determining the kinds of mistakes to which animals are susceptible. In what sense could it be proper to understand a non-linguistic creature to be mistaken about the way things are? Now it has seemed perfectly evident to some that animals are capable of getting things wrong. Norman Malcolm, for example, describes an instance in which a dog, which has the typically canine inclination to chase after cats, chases the neighbor’s cat around a hedge, and ends up barking up the old oak tree.<sup>6</sup> Many of us are tempted to explain these reactions in terms of the dog’s having a belief that the cat is up in the tree, which in this case happens to be mistaken. Such a belief serves to show how the dog’s behavior fits an overall rational pattern.

Donald Davidson has famously argued that the ascription of belief in this case isn’t sufficiently sensitive to intensional contexts to warrant an intentional interpretation.<sup>7</sup> Unless the dog has the capacity to speak and to interpret the utterances of others, Davidson argues, we are unable to pick out the precise content of its beliefs, which means that we are unable to determine the conditions in which its belief would be correct or mistaken. The dog’s behavior is simply not complex enough for us to determine how the dog conceives of the situation. For example, would the dog be mistaken if a possum happened to be in the tree instead, or if the oak had been an elm, or if the cat had belonged to someone else? Davidson’s point is that we are unable to justify any particular specifications of the contents of the dog’s beliefs; any description we give of the contents of its belief is bound to inflate its true conceptual capacities. While I’m not wholly taken by Davidson’s argument, it does serve to remind us that if we lose all sense of the dog’s sensitivity to intensional contexts, then we may also lose the sense that it can be determinately correct or mistaken about

the way things are. If we are to justify our practice of attributing beliefs to Malcolm's dog, we ought to be able to pick out conditions for correctness and incorrectness of the dog's belief, even if we allow there to be some imprecision in the specification of their contents.

Nor do I find Malcolm's example to be ultimately compelling. We wouldn't be so tempted to attribute a mistaken belief to the dog, were we to discover that it habitually barked up that tree. It makes a difference that this behavior is somehow odd or unusual. Underdescribed as the example is, Malcolm's conclusion is an expression of our natural inclination to attribute beliefs to creatures like dogs, and our general willingness to fit the behavior of such animals into rational patterns. But Davidson is challenging the appropriateness of this practice. To meet his challenge, we still need an account of how non-linguistic creatures like Malcolm's dog could interact with their environments in ways that would legitimate our propensity to understand them as beholden to the way things are.

### III. BIOLOGICAL APPROACHES: THE MEANING OF LIFE

It is tempting to suppose that a creature ought to be understood as making a mistake when it does something that isn't "good" for it, or when it fails to take advantage of an opportunity to do something that would be "good" for it. On this line, one assumes that there are conditions in which a creature ought to engage in a certain response, and its responding in this manner is an indication that it takes those conditions to obtain. For instance, one might reasonably suspect that frogs ought to snap at flies, so a frog's snapping at something could then be taken as an indication that it takes its target to be a fly. In this fashion, we are in a position to understand frogs to be making some sort of mistake, in view of their legendarily pathetic disposition to snap at lead pellets.<sup>8</sup> To make a story like this work, we need an account of the conditions in which a creature *ought* to engage in the various responses in its behavioral repertoire. Here we might hold that animals ought to respond in ways likely to sustain their survival and enhance their chances at reproductive success. Simply put, animals generally ought to flee predators, chase prey, and attract mates. With a bit of ingenuity, one might even be able to tell such a story about the antics of Malcolm's dog.

From a biological perspective, perhaps the most reasonable thing for a dog to do when it has treed a cat is to stand below and bark.

This is the governing idea behind Ruth Millikan's biological account of intentionality. She begins with an account of how we might understand the organs and internal mechanisms of naturally selected creatures as having purposes-- she calls them "natural purposes" or "proper functions."<sup>9</sup> Roughly, an item's *proper function* is whatever job that that item's *ancestors* have performed that explains the current presence or proliferation of things of that type. For example, the proper function of the human heart is to pump blood. The fact that past hearts have pumped blood explains why blood-pumping hearts have proliferated in the humans that are around nowadays. On the other hand, even though human hearts may serve as crude timekeepers, this feature doesn't explain their proliferation. Rhythmic thumping is thus not one of their proper functions.

According to Millikan, some creatures have been equipped with internal mechanisms that have been selected to detect or to keep track of the way things are in the creature's surroundings. We can thus understand these creatures to be making some sort of mistake about the way things are when one of these internal mechanisms fails to perform its proper function. The detector responsible for triggering a frog's snapping response, for example, is typically sensitive to small, dark objects zipping across the frog's visual field. And there is a plausible explanation of why frogs tend to be so equipped. Frogs that have survived to pass their genes on to future generations have snapped at a lot of flies (and not much else); if they had lacked this trait, they wouldn't have flourished as much as they have. Indeed, frogs that tend not to snap at flies, or that tend to snap at things other than flies, aren't (or haven't been) as likely to pass this trait on to subsequent generations. So the frogs that flourished in the past and so proliferated were the ones equipped with the ability to detect flies. Millikan's account thus holds that the detection of flies is the *proper function* of the mechanism that triggers the frog's snapping response. Frogs are equipped with such detectors, because the activation of these triggers generally has co-varied with the presence of flies.<sup>10</sup> As a result, we can understand a frog to be mistaken about the way things are when this mechanism triggers the snapping response in the presence of something other than a fly. We may thus understand frogs that snap

at lead pellets to be *mistaking* these pellets for flies. In fact, Millikan's account might even have the resources to underwrite errors that turn out to be beneficial for a creature. For example, it might allow us to understand how frogs could mistake edible things (such as the ballistic frog chow fed to laboratory frogs) for flies, as long as these items did not play a significant role in the natural selection of frogs.

Millikan's achievement is remarkable. She has described a sense in which naturally selected creatures (such as frogs) could be beholden to the way things are in their surroundings, and so has shown us how naturally selected creatures exhibit a genuine *biological* sort of intentionality. However, her account has been castigated as analysis of *belief* or *mental* representation. Several opponents of the biological approach to intentionality have argued that determinations of proper function are too *ad hoc* or too indeterminate to underwrite genuine ascriptions of belief and other states with propositional content.<sup>11</sup> Others have objected that this approach generates counterintuitive assignments of intentional content.<sup>12</sup> Still more have claimed that the biological approach renders content too historically contingent to be a plausible account of the intentionality of mental states. After all, wouldn't we still regard ourselves as thinking things were we to discover that we were cosmic coincidences spontaneously generated out of swamp muck a decade ago, and so not products of natural selection after all?<sup>13</sup> Our conception of ourselves as thinking beings doesn't evidently rest upon our being products of natural selection. In other words, the intentionality of mental states doesn't seem to be contingent in the way Millikan's account suggests.

I'm sympathetic to these objections, but I think the intuitions upon which they rest are better explained in *normative* terms. As I see it, the deeper problem with the biological approach to intentionality is that it appeals to the *wrong sort of normativity* to be a compelling account of *mental* representation. Biological accounts acquire their normative oomph by (implicitly) encouraging us to view creatures subject to natural selection as artifacts of Mother Nature or *survival engines*, "designed" as it were for the purpose of surviving to propagate their kind.<sup>14</sup> Now there is a sense in which artifacts are susceptible to evaluation, and thus possess a certain sort of intentionality. They are subject to a kind of failure according to whether and how well they fulfill the purposes for which they have been designed, or to which they are put. For example, chess-playing computers can be evaluated with respect to how well they play chess. Moreover,

some devices have also been equipped with internal mechanisms whose purposes are to detect or to keep track of the way things are in their surroundings, thereby rendering these devices susceptible to a sort of evaluation with respect to the way things are. For instance, thermostats are equipped with bi-metallic strips, whose function is to detect the ambient room temperature; thermostats can thus be “fooled” by drafts and nearby heat sources. Similarly, slot machines can make mistakes when they accept slugs, and also when they fail to accept genuine coins. Finally, heat-seeking missiles can make mistakes when they fail to keep track of the relative positions of their targets. However, the intentionality possessed by these artifacts seems to be of a derivative, second-class sort, for they can be assessed as correct or mistaken only with respect to the standards set for them by their designers. Since we are the ones who ultimately specify how these devices are supposed to perform, we ultimately determine the proper functions of their internal mechanisms. We say that a slot machine’s internal detector has the function of detecting quarters, because that is what would be required for the machine to fulfill the purposes we have assigned it. Since these devices are correct or mistaken only against the background of purposes we have assigned them, their very *intentionality*, now understood as their susceptibility to evaluation, depends upon (or *derives from*) those who assign them these purposes (in other words, us). And so it would seem that these aren’t compelling examples of mindedness. Chess-playing computers, thermostats, and so on can’t be understood as correct or mistaken “by their own lights,” or accountable with respect to standards that *they* set for themselves. As many have pointed out, an item’s artifactual purpose may depend upon more than its mere responsive dispositions.

Seen in this light, the intentionality captured by biological accounts looks very much like the patently derived intentionality exhibited by conventional artifacts (despite the absence of designers). The standards or norms in play here do not appear to be those of the individual creature; instead, they come from the selective wisdom of Mother Nature or from the interests of the entire species. Indeed, the striking thing about the frog of philosophical legend is that it is so unresponsive to its own mistakes. It doesn’t appear to acknowledge them *as mistakes*. As a result, although we can understand it to be mistaking lead pellets for flies, it isn’t clear that it has made mistakes according to standards that it has set on its own; these mistakes depend upon its having an appropriate biological pedigree and thus transcend its own

responsive dispositions. Biological accounts of intentionality get off the ground by assuming that animals have the proper function of surviving or flourishing (perhaps to reproduce their kind) regardless of how they are actually disposed to interact with their surroundings. On these accounts, organisms are assessed as correct or mistaken with respect to norms or standards that they don't plausibly set for themselves. Consequently, biological accounts can be applied to exceedingly simple sorts of organisms. Millikan is even willing to apply her account to certain varieties of bacteria. But this largesse should give us pause, for now the intentionality in the story begins to look too cheap to be a compelling account of mental capacity.

Moreover, the biological story is psychologically unrealistic, at least as an account of *human* belief or desire. For one thing, the idea that the content of a creature's beliefs should be determined by its actual response to a situation doesn't seem plausible, for the same response can be elicited by several different beliefs. But more importantly, it certainly seems that beings like us can be interpreted as having gotten things right or wrong in ways that don't have much to do with fulfilling the functions for which they've been, as it were, designed or selected. The correctness of many of my beliefs—those about omega chains or the 1960 World Series, for example—seem to have precious little to do with my survival or reproductive interests. We are beings who may be understood as succeeding or failing, or as having gotten things right or wrong, in ways that are not ultimately grounded in our biological purposes. Linguistic beings like us participate in communicative practices that enable us to engage in performances that are meaningful, in the sense that they may be judged as correct or mistaken with respect to standards instituted by the whole community. When we assess the truth or correctness of other speakers' utterances, we don't consider whether or not they've responded in ways likely to enhance their overall biological fitness, nor do we determine whether their utterances accord with behavior that explained the proliferation of their forebears. Indeed, it is in these respects that the interpretation (content-assignment) of human natural languages differs so dramatically from the interpretation of animal communication systems (e.g., those of honeybees or vervet monkeys).<sup>15</sup>

In other words, things like us are equipped to have beliefs and desires of their own, and so capable of exhibiting a goal-directedness that is independent of, and that can even run against, their biological

purposes. We are standard-setting beings, who can pursue counter-(re)productive ends. This insight allows us to cash out the intuition that our intentionality is somehow *original*, or at least not derived in the ways that the intentionality of artifacts and simple organisms is derived.<sup>16</sup> It might even unpack the intuition that our will is in some sense “free.” Calvinists can take heart; we needn’t presume we are mysteriously exempt from the causal or predictive order. In spite of the hegemony of efficient causes, we have a *normative* autonomy. We can be assessed as correct or mistaken in ways that don’t rest upon our biological heritage.

So we ought to look for ways in which animals could interact with their surroundings that could render them susceptible to forms of evaluation that aren’t so closely tied to their naturally selected purposes. How, for instance, could they have beliefs and desires of their own, and so exhibit a goal-directedness that is independent of their natural or proper functions? That is, we should be after an account of how subjects could be beholden to the way things are, where we can discern the standards for evaluation in their own interactions with their environments, and where the normative oomph ultimately derives from some other, non-biological source. The appeal to biology might just give us the meaning of life, but that is all we get. We should ask for more.

#### IV. DENNETT AND THE QUEST FOR REAL MEANING

It is fruitful at this point to compare the approach to intentionality sketched here with that advocated by Daniel Dennett. Dennett is one of the stauncher champions of animal intentionality, a philosopher whose work cognitive ethologists routinely draw upon to justify their attributions of intentional mental states to non-linguistic animals. On Dennett’s view, we are justified attributing intentional states to agents whose behavior falls into a *rational pattern*, as discerned from an “intentional stance.”<sup>17</sup> The attribution of intentionality to a subject (or “system”) is warranted by (and only by!) the predictive and explanatory usefulness of regarding it *as if* it pursued goals based on information available to it. Dennett clearly thinks that animals pass this test. As products of natural selection, they exhibit an overwhelming tendency to behave in ways consonant with the preservation of themselves and their kin. So one can

fruitfully and voluminously predict their behavior simply by assuming that they act *as if* they had their continued survival “in mind.” Similarly, insofar as sophisticated artifacts (such as chess-playing-computers) are the products of intelligent design, we can also profitably predict their behavior by treating them *as if* they had their purposes “in mind.” So Dennett holds that the behavior of both non-linguistic animals and artifacts can fall into a pattern sufficiently *rational* to warrant attributions of intentionality.

Dennett is also willing to taxonomize minds based upon their capacity to construct and evaluate possible responses to situations in their environments. As Dennett sees it, creatures with higher intellectual capacity have the ability to kill off bad plans before those unwise courses of action end up killing them. Abjectly tropistic, or “Darwinian,” creatures fall at the bottom level of his “Tower of Generate and Test.”<sup>18</sup> Lacking the capacity to tailor their responsive dispositions to their environments on their own, their responsive dispositions evolve only through the operations of natural selection. More sophisticated than these are animals equipped with the ability to revise their own responsive dispositions through operant conditioning. Above these “Skinnerian” creatures are “Popperian” creatures that are capable of basing their behavior upon simulated outcomes of the possible responses that they may make. Finally, “Gregorian” creatures occupy the top of Dennett’s hierarchy. These creatures are capable of designing tools of their own, including linguistic tools like words, to discover ever better means of navigating their environments. At first blush, then, Dennett would appear to be quite friendly to the ecumenical solution to the debate over animal intentionality that I’m endorsing here.

However, as is so characteristic of his work, Dennett’s true sentiments are much harder to divine. While he allows that things like us have different kinds of minds from more primitive organisms, he is suspicious of attempts to draw distinctions between the kind of *intentionality* we have and the kind typically attributed to simpler animals and artifacts. Notice that Dennett’s taxonomy is constructed around a creature’s resources to pursue its biological purposes. Some organisms are cognitively superior to others because they are more adaptive or able to cope flexibly with the contingencies of their environments. Dennett does not try to show how creatures occupying the upper storeys of his tower can have goals of their own, or be evaluated as having gotten things right or wrong, in ways that don’t appeal to their biological

purposes. Assuming that intentionality should be botanized along normative lines, we can see that Dennett doesn't show how creatures belonging to different levels in his hierarchy have differing *intentional* capacities.

Indeed, there is reason to suspect Dennett would be hostile to any attempt to draw such distinctions. Nowhere is this unecumenical attitude more evident than in his recent attempt to debunk the notion of "original" intentionality, and his subsequent rejection of "intentional realism." In some of his recent work, Dennett has assembled a line of reasoning purporting to show that we ought to reject the idea that there is any substantive distinction to be drawn between *original* and *merely derived* intentionality.<sup>19</sup> So he seeks to undermine the intuition that we "really" represent the world, or that we represent it *to ourselves*, whereas things like chess-playing computers and frogs do not. As he sees it, the root error of "industrial strength" intentional realism is the assumption that there must be something special about our activity that makes us susceptible to the attribution of a different, non-derived type of intentionality.

Here I am pointing to the moment where I think the wrong turning [on the part of intentional realists] occurs: it occurs when the implied contrast is accepted. My view is that belief and desire are like froggy belief and desire *all the way up*. We human beings are only the most prodigious intentional systems on the planet, and the huge psychological differences between us and the frogs are ill described by the proposed contrast between literal and metaphorical belief ascription.<sup>20</sup>

In other words, subjects cannot be evaluated as correct or mistaken with respect to anything other than biological norms. Attempts to discover something special about persons - some feature that would ground attributions of a different, original, kind of intentionality - are, in his eyes, "a great labor wasted."<sup>21</sup>

As far as I can make out, there are two distinct strands to the case Dennett builds against the intentional realist. First, he takes the ever-popular Twin-Earth thought experiments to show that it can be very difficult to determine precisely what a subject means. By having us consider physically indistinguishable subjects occupying different possible worlds, these thought experiments purport to show

that subjects with the same internal constitutions may nevertheless be interpreted as having thoughts with differing contents, due simply to differences in their respective environments.<sup>22</sup> If so, then belief contents are not completely determined by the activity inside a subject's head. For Dennett, this lack of intrinsic content determination is of a piece with his thesis that interpreters have some latitude in specifying the contents of a subject's intentional states. Just as it makes no substantive difference whether a frog's "fly-concept" includes the ballistic frog chow it encounters in a lab, it can also be unclear just what lies within the extensions of our own concepts. Dennett rhetorically asks: would I be making a mistake, were I ever to apply my concept 'horse' to the "schmorses" inhabiting the pastures around my philosophical doppelganger? He encourages us not to expect determinate answers in such bizarre circumstances.

Now it is one thing to accept Dennett's contention that it can be somewhat unclear what the extensions of our concepts are, but it is quite another to claim that this indeterminacy or interpretive latitude could amount to a compelling case against non-biological, "less derivative" types of intentionality. Here Dennett seems to conflate *original* intentionality with John Searle's conception of *intrinsic* intentionality.<sup>23</sup> While Twin-Earth thought experiments might show that linguistic intentionality isn't wholly internal, and so not "intrinsic" in the sense suggested by Searle's vocabulary, there is no reason to suppose that the contents of non-biological intentional states would have to be completely settled by a subject's internal activity (or invariant across extrinsic changes in a subject's surroundings). We can perfectly well acknowledge indeterminate cases without concluding that all intentionality must be of the biological or artifactual sort.

In the second part of his case against original intentionality, Dennett seeks to embarrass intentional realists with the observation that we have the biological kind of intentionality by virtue of being products of natural selection. Suppose we were to fangle sophisticated robots that serve our interests by interacting with their environments with all the cleverness of genuine human beings. Even if these robots were capable of talking to one another, Dennett seems to think that an intentional realist must agree with Searle's contention that whatever intentionality they possess could only derive from their designers.<sup>24</sup> By the same token, though, our intentionality could only derive from natural selection, since we too are basically

survival engines, designed to serve the interests of our genes. Insofar as we are “artifacts of Mother Nature,” we clearly have the biological form of intentionality Millikan has described so well. Dennett takes this to entail that *our intentionality* must be every bit as derived as that of the frog of philosophical legend.

But this line isn't compelling either, and our normative account of intentionality allows us to see why. No intentional realist worth his or her salt would let Dennett talk about “our intentionality” as if we could have only one kind. This just goes to show that Dennett is in the grip of the idea that rival accounts of intentionality preclude one another, and this is precisely the bias that our ecumenical philosopher of mind is seeking to overcome.<sup>25</sup> Any credible intentional realist would challenge Dennett's inference that since we are the products of natural selection, the ways in which we may be understood as accountable are intelligible only with respect to Mother Nature's standards or intentions. By no means do I want to deny that we are products of natural selection, and so *can* be understood as subject to the biological accountability that Dennett and Millikan take us to have. Biological accounts of intentionality succeed in showing how we could attribute to humans intentional states like those of the frog, but they don't thereby show that this is the *only* type of intentionality properly attributed to humans. The fact that the products of natural selection can be evaluated according to biological norms doesn't preclude the possibility of there being other ways to evaluate human behavior that turn out to be largely independent (or intelligible apart from) the purposes for which humans have been designed or selected.

Here we ought to keep in mind that not all of an organism's attributes need to be identified with respect to their proper functions. Being green and having a disposition to snap at flies are two examples. One can see that a frog is green or recognize that it has a disposition to snap at flies without having to know how these traits contribute to the frog's biological success. One might want a further explanation of why being green or having a disposition to snap at flies would be evolutionarily advantageous for a frog. But the point here is that an account of *what it is* for an organism to have a certain trait need not include an explanation of *why* that organism has that trait.<sup>26</sup> These can be separate issues. An intentional realist can claim that having thought (or original intentionality) is a behavioral trait (albeit a complex one) that is

similar in the sense that it can also be identified in creatures without having to determine how it contributes to these creatures' biological fitness. In this fashion, an intentional realist can explain how having contentful thought could be independent of a subject's biological purposes. Just as there could be green things that aren't products of natural selection, there could be non-biological things (such as Dennett's robots or Davidson's "swampman") that possess original intentionality.

By no means would I deny that an *organism's* possession of original intentionality calls for biological explanation. One would want to know how being subject to evaluation in the sense requisite for some other kind of intentionality could have enabled (or at least not prevented!) an organism's ancestors to survive to reproduce their kind. Any non-biological intentionality ought to be "biologically plausible" if it is possessed by thoroughly biological beings such as ourselves. But although any capacity we have to harbor intentional states is presumably the product of natural selection, and so in a sense a biological phenomenon, that doesn't mean that *the intentionality* of these states has to be of the biological kind described by Millikan. Even though there is a sense in which our intentionality "derives" from the operation of natural selection, this doesn't mean that the intentionality itself is "derived" in the same sense that the intentionality of frogs is derived-- to wit, that the norms by which they are evaluated must ultimately be biological norms. Dennett's conclusion that our intentionality must be as derived as that of simpler biological organisms runs these senses of "derived" together. In other words, the fact that beings with high-grade, original intentionality have evolved from organisms with low-grade, biological intentionality doesn't make this higher-grade intentionality a form of, or even continuous with, this lower form of intentionality. Things like us may have evolved *additional* intentional capacities (or susceptibilities to evaluation).<sup>27</sup> That we have evolved through natural selection doesn't rule out the possibility of our being susceptible to evaluation according to standards that we set for ourselves, standards that are wholly funded out of our responsive dispositions without appeal to the purposes for which we have, as it were, been designed or selected.

Let's step back for a moment to see where we stand. We're after an understanding of how we can draw distinctions between kinds of intentional, mental capacities so that we can do justice to the idea that

linguistic beings exhibit a special kind of intentionality, while at the same time allowing that non-linguistic creatures can enjoy mental lives of their own. I began by suggesting that we should taxonomize intentionality along its normative dimension. Different kinds of intentional capacities correspond to different ways in which subjects are susceptible to error. Although our mental lives are richer because we are susceptible to different kinds of evaluation, non-linguistic animals are not altogether incapable of getting things right or wrong. After providing some reasons why we might want to look for more than biological intentionality, I showed why we should not accept Dennett's conclusion that the quest for "real meaning" is a quixotic venture best abandoned. Something's being the product of selection or design might explain *why* it has a special sort of accountability, but it doesn't follow that this accountability must be biological or artifactual. It thus remains worthwhile to pursue the question of whether (and if so, how) people and animals could exhibit non-biological intentionality.

## V. NON-BIOLOGICAL INTENTIONALITY: THE RATIONALITY OF EDUCABILITY

Let's begin with Dennett's thesis that the attribution of intentional states is appropriate for those beings whose behavior falls into rational patterns discernible from an intentional stance. Dennett himself isn't entirely clear about what exactly constitutes a rational pattern, and he fails to entertain the possibility that there might be several such patterns. In fact, this oversight is of a piece with his insistence that there is only one kind of intentionality and only one kind of intentional stance. Once we have rejected Dennett's reasons for abandoning the quest for real meaning, however, we are in a position to look for other kinds of possible rational patterns corresponding to different ways in which one may adopt an intentional stance. People evidently engage in a sophisticated, rational pattern of behavior when they go about justifying (or "rationalizing") their claims and actions to one another. However, we are looking for less-sophisticated, rational patterns of behavior that could plausibly be exhibited by non-linguistic animals. What might they be?

It is commonly held that creatures with educable capacity enjoy richer mental lives than tropistic beings. I have already mentioned that Dennett places organisms capable of operant conditioning and latent learning at higher levels of his “Tower of Generate and Test” than creatures that are unable to modify their responsive dispositions through experience. Fred Dretske is another philosopher who thinks that we should restrict the attribution of “genuine belief” to creatures capable of simple reinforcement learning. On Dretske’s account, learning involves the selection of internal detectors to control a creature’s behavior based upon the states of affairs that these detectors indicate. Creatures come to possess intentional states that mean what they do *to those creatures* only after their behavior has been conditioned by this selection of internal indicators.<sup>28</sup> Unfortunately, both Dennett and Dretske concentrate on how educable capacity renders an organism better able to fulfill its natural purposes in the face of environmental contingency. Dretske, for instance, focuses on how providing creatures with the ability to conduct their own selection of appropriate internal indicators may be the best way for a designer (including Mother Nature) to construct creatures that are likely to fulfill their intended purposes.<sup>29</sup> Unable to anticipate the relevant regularities of a creature’s environment, a designer might find it useful to equip creatures with some ability to tailor their own responsive dispositions to their particular surroundings. While this thought is probably correct, it doesn’t show how creatures with such educable capacity exhibit any *intentionality* of their own, apart from that attributed to simple biological organisms or artifacts. At best, it explains why behavioral plasticity might have arisen through natural selection. Although there is a sense in which educable creatures are responsive to error, and so “learn from their mistakes,” it remains to be seen how these mistakes could be intelligible as such, apart from appeal to the creatures’ biological purposes.

Now if we had a story about how the flexibility of educable creatures gives rise to a special sort of accountability or susceptibility to evaluation, then we might begin to see how non-biological intentionality could be a product of natural selection. I’d like to illustrate how a very simple story like this might go. Several learning theorists have argued that the apparent educability of some creatures is best explained in terms of the adjustment of certain cognitive structures called *expectations*.<sup>30</sup> For example, expectations are a reasonable explanation for the *blocking* phenomena often observed in actual creatures. Animals that have been trained to associate a conditioned stimulus with an unconditioned stimulus will subsequently fail to

associate other stimuli with the unconditioned stimulus, when the latter are presented along with the original conditioned stimulus. For example, rats that have been trained to associate a bell tone with an electric shock will not come to associate a red light with a shock, as long as the red light is consistently paired with the bell tone. The prior conditioning prevents (or “blocks”) subsequent conditioning to other, co-varying stimuli. If their learning were merely a matter of the frequency of stimulus-pairing, then one would expect the animal to become conditioned to the new stimulus as well. One would expect the rats eventually to associate the red light with a shock, as indeed they do when they aren't subjected to the earlier training. Many learning theorists have argued that the failure of previously conditioned animals to become conditioned to the new stimulus arises because the animal already uses the original conditioned stimulus to *predict* the occurrence of the unconditioned stimulus, and with a reasonable degree of success. When a previously conditioned rat encounters the compound tone and light stimulus, it *expects* that the shock will occur (because it heard the bell tone), and so the subsequent shock isn't a surprise. Since events are as they were *expected* to be (they were not novel), there is no pressure to develop new associations, and there is no subsequent conditioning to the light. Thus these theorists conclude that the rats are responding to *surprise*, to things not being as they *expected* them to be.

On accounts of expectation-based educability, animals revise their responsive dispositions over time by being sensitive to the consequences of their responses. For instance, rats that have been trained to press a lever while hungry to obtain a fructose solution are more likely to press this lever when they are thirsty than similarly trained rats that have been trained to press the lever to obtain food pellets. Evidently, the first group of rats has learned that lever pressing can be more than just an appropriate response to hunger. The intuitive interpretation is that these animals have learned that pressing the lever brings about an outcome that satisfies thirst as well as hunger. Similarly, expectations are also thought to be required to account for cases of “latent learning” that resist explanation in terms of simple operant conditioning. For example, rats that have been given time to explore a maze are (unsurprisingly) better able to navigate their way through it to new food sources than rats unfamiliar with the maze. Not only does exploration seem to be its own “reward”, but rats also seem capable of exploiting this knowledge of their surroundings in their pursuit of new ends.

Accounts of expectation-based educability differ in detail, but share a common structure. Viewed abstractly, expectations have three basic components. First, individual expectations possess *conditions of activation and deactivation* that specify, as it were, when these expectations are turned on or off. Next, expectations require some kind of *consequence condition* that picks out the expected states of affairs associated with the activation of an expectation. Finally, expectations require some kind of *response component* that specifies the responses expected to bring about the satisfaction of the consequence condition. *Expectation-mongering creatures* can then be defined as those whose responses are governed in part by the consequence conditions of their currently activated expectations.<sup>31</sup> By adjusting these basic components when specific conditions have been met (for example, whenever engaging in the response component of an activated expectation fails to bring about the satisfaction of its consequence condition), expectation-mongering creatures can exhibit the sensitivity to the consequences of their own responses that learning theorists have sought to describe.<sup>32</sup>

Given the intuition that the activation of expectations can be evaluated as correct or mistaken, this would seem to be a promising beginning of a story about intentionality that isn't biologically grounded. However, to sustain the claim that expectation-mongering creatures exhibit a non-biological sort of intentional capacity, we need to show how their behavior fits an overall rational pattern. By that, one might mean that their activity is directed toward the attainment of non-biological goals or ends. And indeed, by adopting a certain stance, we can see the behavior of expectation-mongering creatures as purposeful activity, and even explain their behavior in terms of goals different from the biological purposes animating Millikan's account. To the extent that a creature is disposed to engage in responses *expected* to bring about a certain outcome, we can regard that outcome as one of its current goals.<sup>33</sup> So construed, such goals work in conjunction with a creature's expectations to explain its particular responses to situations. By characterizing responses in terms of the outcomes they are expected (by the subject) to bring about, these explanations show how a particular response fits a creature's overall pattern of responsive dispositions.<sup>34</sup> We need not think of such explanations as empty, because they point out that a subject might have done

otherwise, had that response not been expected to bring about a certain outcome, or had some other response been expected to bring about that outcome.<sup>35</sup>

In light of our discussion of biological intentionality, it is worth noting that the attribution of these goals to a creature does not obviously rest upon a determination of the purposes for which a creature has been designed or selected. In fact, not only is this account of goals intelligible apart from considerations of a creature's biological purposes, the goals on this account may even collide with those purposes. For instance, there is no reason why a creature couldn't be disposed to respond in ways expected to bring about some self-destructive outcome. Such a creature would have a goal that is, *biologically speaking*, remarkably maladaptive. The important thing is that these goals are completely determined by how a creature is disposed to respond to the activation of its expectations.<sup>36</sup> Nevertheless, this account does not simply identify these goals with the outcomes the creature actually brings about. In other words, creatures do not have to be disposed to bring about the eventual attainment of their goals. For one thing, a creature might not ever find itself in circumstances where the attainment of its goals is possible. But more significantly, even when their attainment is possible, the fulfillment of a creature's goals may be hindered by the activation of the wrong expectations. An expectation-mongering creature will be disposed to attain its goals (whenever such attainment is possible) to the extent that its expectations are configured *correctly*.<sup>37</sup> The nice thing is that we can pick out unfavorable expectation configurations likely to hinder a creature's attainment of its goals, and so have reason to regard these configurations as expectation *errors*. For instance, a creature is liable not to fulfill a goal if one of its expectations is activated in a situation in which the expectation's response would fail to bring about the satisfaction of its consequence condition. We can thus think of such an occurrence as an *error of commission*. Similarly, an *error of omission* arises whenever the response of an expectation that is not activated would bring about the satisfaction of its consequence condition (that is, were its activation not to be an error of commission). Since these two expectation configurations are liable to prevent a creature from attaining its goals, as they've been construed above, expectation-mongering creatures are susceptible of two distinct sorts of mistakes about the way things are in their environment.<sup>38</sup> They can be evaluated as having gotten things right or wrong,

and so may be understood to exhibit an intentionality above and beyond that typically attributed to artifacts and simple organisms.

According to this theory, expectations ought to be activated just in case their consequence conditions would be satisfied, were the creature to engage in the response picked out by that expectation's response component. As an account of error, this account has several appealing features. First, unlike what happens in biological accounts of intentionality, the errors in this account do not depend upon a creature actually making appropriate responses to the situations it encounters. Moreover, the conditions for the activation of one expectation to be appropriate can be quite different from the conditions of appropriate activation for another. That is, the activation of separate expectations can be beholden to distinct features of a creature's environment. As a result of this *feature selectivity*, expectation-mongering creatures can thus be correct with respect to some features of their environment, yet mistaken with respect to other features. They may be right or wrong *in a variety of respects* consequent upon the simultaneous activation of several expectations. In fact, an expectation-mongering creature may even be massively mistaken about the way things are.<sup>39</sup> Indeed, since the situations in which one expectation would be satisfied may happen to line-up or co-vary with those in which another would be satisfied, distinct expectations may, "extensionally speaking", share the same circumstances of appropriate application. However, the *particular means* by which these circumstances are picked out would differ for each such expectation, simply because they would be comprised of different expectation components. So even though their circumstances of appropriate activation may be the same, their content ("intensionally speaking") may remain quite distinct. Had the subject's environment been otherwise, these expectations may not have shared circumstances of appropriate activation. It would thus appear that attributions of expectation states could exhibit something like the ballyhooed semantic opacity or sensitivity to intensional contexts so often associated with the attribution of genuine intentional states. To attribute an expectation to a creature is not tantamount to attributing to it other expectations sharing the same circumstances of appropriate activation. It would seem then that we've found grounds to challenge the popular contention that a fine-grained sensitivity to intensional contexts would require linguistic capacity.<sup>40</sup> This account thus meets what might be called "Davidson's challenge." At the very least, advocates of Davidson's position would need to clarify

just what they mean by the sensitivity to intensional contexts alleged to be required for legitimate ascription of intentional states.

There is also a respect in which the behavior of expectation-mongering creatures may be seen to exhibit a certain measure of *critical* rationality on top of the practical (goal-driven) rationality I've been describing. Insofar as they are disposed to revise their expectations in the wake of the errors described above, educable creatures would be disposed to take steps to avoid similar mistakes in the future. There is of course no guarantee that these revisions will yield future success.<sup>41</sup> The point is just that creatures displaying this sort of educable capacity would take expectation correctness or aptness to be a regulative ideal, at least in the sense that they are disposed to revise error-prone expectations while leaving correct expectations as they are. And so it seems that they exhibit something akin to the capacity for rational responsiveness to error that Davidson argues must be possessed by any rational animal.<sup>42</sup> By responding in a more or less reasonable fashion when the outcomes of their responses aren't as they were expected to be, such creatures manifest an apparent capacity to be "surprised."

In sum, we can discern a sort of rational structure in the behavior of expectation-mongering creatures. As I've shown, they can be seen to pursue goals based upon possibly mistaken ideas about how to attain them, and their ability to attain these goals can develop in a more or less reasonable fashion. Since they can be evaluated as having gotten things right or wrong, we are justified in crediting these creatures with some sort of intentional capacity. Moreover, this intentionality is clearly distinct from the biological and artifactual kinds described by Millikan and Dennett. Expectation-mongering creatures do not have to be products of any sort of selection, natural or otherwise, and their expectation errors are intelligible without our having to consider the purposes for which they have, as it were, been designed. So this account shows how non-biological "creatures" such as one spontaneously generated out of swamp muck could nevertheless possess a certain kind of intentional capacity. Thus the account avoids the somewhat disconcerting conclusion that a being physically indistinguishable from something capable of bonafide mental representation could nevertheless be wholly incapable of getting things wrong, on account of its lacking a suitable pedigree.<sup>43</sup>

## VI. HIGHER-ORDER INTENTIONALITY

So far, I've described two routes (or stances) by which non-linguistic animals could appropriately be understood to be intentional, or capable of being right or wrong about the way things are. One might be inclined to stop at this point and try to show, for instance, how higher-order intentional capacity is just a sophisticated form of expectation mongering. But a truly ecumenically-minded philosopher would resist such reductionist temptations. We have no reason to suppose that there couldn't be yet other rational patterns of behavior, such as those underlying our capacity to justify claims and actions to one another. I'd like to wrap this discussion up, therefore, with a few remarks on more advanced intentional capacities.

It has been argued that persons are distinct by virtue of having *second-order intentionality*, the capacity to have intentional states about intentional states. Some have thought that this capacity is a critical part of language use. For example, Griceans maintain that genuinely meaningful utterances must be made with an intention to induce (possibly mistaken) beliefs in others.<sup>44</sup> Dennett attributes the sense that persons are distinct from other intentional systems to the fact that it is fruitful to regard them as themselves adopting "the" intentional stance to predict and explain the behavior of others. Similarly, Davidson argues that "genuine believers" must themselves have the concept of belief, and so be capable of having beliefs about beliefs. Notoriously, he further argues that such capacity could make itself manifest only in the context of a subject's interpreting the utterances of others; that is, thought requires talk.<sup>45</sup>

Non-linguistic animals would not be able to attribute *linguistic* intentionality to others. But might they be capable of regarding others as having simpler states of mind? Several ethologists have resisted Davidson's conclusion that second-order intentionality could only arise in a linguistic context, and so have sought to show that non-linguistic animals could have beliefs about beliefs or at least a rudimentary "theory of mind."<sup>46</sup> Accordingly, there has been interest in discovering how well different creatures can exploit or be trained to discriminate the evident attentional or perceptual knowledge of others. There has also been a

great deal of excitement over purported cases of deception amongst animals.<sup>47</sup> For example, ground-nesting birds typically feign injuries in order to draw predators away from their nests. But more intriguing are instances of apparent intra-specific deception amongst primates. Vervet monkeys, for example, have occasionally been observed to emit “false” alarm-calls, seemingly to avoid potentially unfavorable confrontations with rivals. Lower-ranking chimps and bonobos have been observed not to reveal (or even to conceal) food caches, apparently in order not to have to share them with higher-ranking conspecifics. Subordinates have also been observed to suppress evidence of sexual activity in order to avoid discovery by dominant individuals. Finally, there are reports of younger individuals feigning threat or injury from a rival, so that family members will chase the other away from some desirable what-not. It has become increasingly clear that some primates have to navigate a turbulent sea of shifting relationships and coalitions. According to the so-called *Machiavellian Intelligence Hypothesis*, greater cognitive capacity evolved in the context of an intellectual arms race, where creatures under selective pressure to manipulate alliances to their advantage acquired ever-greater powers to deceive one another, while avoiding getting duped themselves.<sup>48</sup>

If, as has been argued here, normativity is the hallmark of intentionality, this attention to animal deception is entirely appropriate. Cases of deception require a creature to comport itself so as to induce another into making a *mistake* about how things are.<sup>49</sup> Given the debate over animal intentionality, however, it should come as no surprise that there is little consensus on which cases of purportedly deceptive activity are good evidence for second-order intentionality. Happily, we are in a position to explain this disagreement, for just as there may be several different kinds of intentionality, there may be several corresponding varieties of deception. Just as there may be different kinds of goals and goal-directedness, there may be different ways in which deception can be the goal of an animal’s behavior. For instance, a plover’s injury-feigning behavior hasn’t struck many as an instance of “genuine” deception. Even though its (biological) purpose is evidently to deceive the predator (and so qualifies as a kind of deceptive behavior), it isn’t clear that the bird has this goal appropriately in mind, or even whether such animals are capable of engaging in behavior that is goal-directed in more than just the biological sense. It seems too automatic or too hard-wired to count as an instance of behavior exhibiting the *kind* of second-

order intentionality that these ethologists are after. “Genuine” cases of deception require a creature to grasp the possibility that another is capable of getting things wrong, and so to possess some sort of intentional capacity. The animal must engage in such behavior with the other’s error as its aim or goal, which in turn may redound to the perpetrator’s advantage.

What about the much less instinctive, probably learned behavior of the primates? Here the problem has been to rule out the possibility that subjects have simply cottoned onto a relatively effective trick for manipulating others, conceived simply as mobile obstacles in their environment. For example, an enterprising individual might discover that simply emitting a certain call gets rivals to head for the trees or gets momma to come to its aid. If so, then even though the behavior is in effect deceptive, it might not really be intended to induce a mistaken belief. Instead, it would merely be a response meant to get the other to do something conducive to one’s aims, and so wouldn’t be a case of genuine deception after all.<sup>50</sup> Emphasizing the difficulty of ruling out such “killjoy” interpretations of a creature’s deceptive behavior, especially when treated on a case-by-case or anecdotal basis, Dennett has argued that we should look for more systematic evidence of deception in order to warrant an interpretation that the perpetrator has *deception* relevantly in mind.<sup>51</sup> For instance, one might like to see evidence of a creature’s engaging in a certain bit of deception in the pursuit of *different* ends and, in particular, for a creature to engage in a certain bit of deceptive behavior in different kinds of situations with the evident purpose of getting the deceived creature to make *different* kinds responses. This way, a creature could register an appreciation that its target may make different responses to the possession of the same belief, and so begin to register a behaviorally robust theory of mind.<sup>52</sup> However, the very systematicity Dennett urges us to seek poses challenges of its own. By its very nature, we would expect deceptive behavior to be relatively infrequent or unusual, lest deceived individuals come to realize that they are being played for dupes. Indeed, to the extent that a subject repeatedly engages in deceptive behavior, one begins to suspect that it has simply latched onto a neat trick, and so isn’t genuinely treating others as having minds of their own.

In section V above, I explained how certain educable creatures can be understood to exhibit a special, non-biological sort of intentionality. So insofar as a creature treats another *as educable*, it could

exhibit some sort of second-order intentionality. Indeed, part of the difficulty in finding systematic evidence for intentional deception among animals arises because chronic deceivers wouldn't seem to be treating their marks as particularly educable. Are there patterns of behavior in which we could discern a creature's treating another as educable, and so as adopting an intentional stance of sorts to others? Are there actual cases where a creature's (non-biological) goal or purpose is to get others to learn from their mistakes? When the issue is thus cast, the obvious place to look would be at cases where animals appear to be trying to *instruct* one another, cases in which it seems to be a subject's purpose for another to get better over time.<sup>53</sup> Such social practices are inherently more systematic than one-shot cases of self-deception, and they are happily more cooperative. In fact, an instruction routine might even include attempts to *fool* or to *test* a subject. But unlike the cases of *tactical* deception described earlier, such efforts to induce a trainee into making a mistake wouldn't necessarily be for the trainer's immediate benefit, but rather to get the trainee to *learn* from its errors. So this would also seem to be a good place to look for systematic occurrences of *deception*. Of course, most cases of animal instruction are probably too instinctive for this instruction to be counted among an animal's non-biological goals. Mother cats, for instance, might train their offspring to hunt simply out of instinct. This instruction might not fulfill any purposes beyond the biological. But it would be interesting to find cases in which an animal apparently instructs another for non-biological ends as well. I wouldn't be surprised if some non-linguistic animals are capable of acting with another's eventual improvement appropriately "in mind," but until ethologists have explored the matter in greater detail, I'm not in any position to make pronouncements one way or the other.

There is a more compelling reason why we should be interested in cases of apparent animal instruction. Beings like us participate in complex social practices that enable us to engage in performances that are correct or mistaken with respect to standards that are instituted and maintained across entire communities. Insofar as these performances are susceptible to such evaluation, these practices possess a *social* type of intentionality. Language-use is perhaps the clearest example of such a social practice. The truth-conditions (or meanings) of utterances in a natural language are in part determined by the speech dispositions of the language's speakers; linguistic meaning is somewhat conventional. Now some animals have been trained to participate in linguistic practices in at least a limited way. Primates can be taught to

associate different signs with different types of objects in their surroundings, and also to use such signs to make “requests” of sorts. More impressively, they can be trained to use such signs to coordinate one another’s behavior.<sup>54</sup> Nevertheless, it isn’t clear that they can ever be full-fledged participants in these social practices. The most striking thing about these attempts to discern the linguistic ability of apes is that *we* set the meanings or standards for the correct use of the signs they employ. That is, their performances can be evaluated with respect to standards that we have set *for them*; these standards come, as it were, from “on high.” Of course, much the same can be said of young children; they are ushered into linguistic practices where the correctness of their utterances are determined by the other members of their communities. But while they might not begin their linguistic careers as full-fledged members of their linguistic communities, eventually they mature into beings that *enforce* these standards as well. They too become standard-setting beings, capable of maintaining and enforcing the linguistic norms of their culture, by instructing and correcting the performances of others, as well as by *debating* with one another the appropriateness of such performances. It is much less evident that the apes who have been trained to manipulate different sorts of signs could ever be similarly capable of enforcing these standards of correctness on their own, say, by *acculturating* others into their practices. Left to their own devices, it is most improbable that they would maintain these practices and pass them along to others.<sup>55</sup> In this sense, they don’t evidently get the *point* of these cultural practices, and so it isn’t clear that they are capable of becoming fully-fledged linguistic beings.<sup>56</sup> But while apes might not be capable of the discursive intentionality we linguistic beings enjoy, the ecumenical philosopher of mind would remind the friend of animal intentionality that there is no reason to despair. For as I’ve shown, there are other stances from which we can discern intentionality in the behavior of non-linguistic animals.<sup>57</sup>

---

<sup>1</sup>Notice the slide from mental representations being objects of the mind to items in the head. Examples of this strategy include Dretske, *Knowledge and the Flow of Information* (Cambridge, MA: MIT Bradford, 1981), Fodor, *Psychosemantics* (Cambridge, MA: MIT Bradford, 1987), Lloyd, *Simple Minds* (Cambridge, MA: MIT Bradford, 1989), and Cantwell-Smith, *On the Origins of Objects* (Cambridge, MA: MIT Bradford, 1996). It bears mentioning that these folk generally don't consider that there might be many different kinds of aboutness, corresponding to different kinds of intentionality.

<sup>2</sup> For example, in *Making It Explicit* (Cambridge, MA: Harvard, 1994), Brandom argues that we comport ourselves toward objects *as objects* by (and only by!) mastering the use of singular terms in a public language, something beyond the apparent capacities of (non-linguistic) animals.

<sup>3</sup> See Chisholm, *Perceiving: A Philosophical Study* (Ithaca: Cornell, 1957), p. 168ff.

<sup>4</sup> Philosophers like to warn others not to confuse the philosophical notion of intentionality with the more common notion of intentional action. I'm suggesting here that drawing a sharp distinction between these notions may not be such a good idea, for it makes one apt to miss the essential normativity of intentional states. After all, the ultimate purpose of attributing intentional states to subjects is to make sense of their behavior, as intentional activity.

<sup>5</sup> At this level of abstraction, there is no reason to presume that intentional beings must become beholden to the way things are by virtue of having discrete internal states in their heads that purport to represent or stand for certain states of affairs. Also note that this is more than just the claim that certain belief contents require linguistic ability.

<sup>6</sup> See Malcolm, "Thoughtless Brutes," *Proceedings and Addresses of the American Philosophical Society* 46 (1972).

<sup>7</sup> Davidson discusses this case in "Rational Animals," in LePore and McLaughlin (eds.), *Actions and Events* (Oxford: Basil Blackwell, 1985).

<sup>8</sup> This legend traces back to Lettvin, et al., "What the Frog's Eye Tells the Frog's Brain," *Proceedings of the Institute of Radio Engineers* (1959): pp. 1940-51.

<sup>9</sup> Millikan spells out her account in *Language, Thought, and Other Biological Categories* (Cambridge, MA: MIT Bradford, 1984) and *White Queen psychology and Other Essays For Alice* (Cambridge, MA: MIT Bradford, 1993).

<sup>10</sup> As several philosophers of biology point out, an internal detector can be selected to indicate a certain state of affairs, despite its having not been perfectly reliable. See, for example, Godfrey-Smith, "Indication and Adaptation," *Synthese* 92 (1992): 283-312.

<sup>11</sup> Fodor is one of the most trenchant critics of biological accounts of intentionality. See in particular *A Theory of Content and Other Essays* (Cambridge, MA: MIT Bradford, 1990) *Mind and Language*, 11 (1996): 246-62. For a useful response to some of these worries, see Rowlands, "Teleological Semantics," *Mind*, 106 (1997): 279-303.

<sup>12</sup> For example, see Pietroski, "Intentionality and Teleological Error," *Pacific Philosophical Quarterly*, 73 (1992): 262-82.

<sup>13</sup> The "swampman" fantasy originated with Davidson's "Knowing One's Own Mind," *Proceedings and Addresses of the American Philosophical Association*, 60 (1987): 441-58.

<sup>14</sup> See *Language, Thought, and Other Biological Categories*, p. 17. Millikan often seems to shy away from addressing the question of why misrepresentations on her view ought to be understood as mistakes. My guess is that this is because she wants to avoid a charge of excessive adaptationism. Dennett, by contrast, is perfectly happy to claim that we are justified to regard biological organisms as if they were artifacts of Mother Nature. It's this kind of talk that raises Fodor's hackles.

<sup>15</sup> For an extensive discussion of the vervet communication system, see Cheney and Seyfarth, *How Monkeys See the World* (Chicago: Chicago, 1990).

<sup>16</sup> Notice that this sense of "original" intentionality must be distinguished from Searle's notion of "intrinsic intentionality" (This will be important in what follows.). There is no reason why machines couldn't exhibit "original intentionality", so construed.

<sup>17</sup> See *The Intentional Stance* (Cambridge, MA: MIT Bradford, 1987), especially chapter 2.

<sup>18</sup> See chapter 13 of *Darwin's Dangerous Idea* (New York: Simon & Schuster, 1995), and chapter 4 of *Kinds of Minds* (New York: Basic Books, 1996).

---

<sup>19</sup> Versions of his argument can be found in chapter 8 of *The Intentional Stance*, “The Myth of Original Intentionality,” in Said et. al. (eds.) *Modelling the Mind*. (Oxford: Clarendon, 1990), chapter 14 of *Darwin’s Dangerous Idea*, and chapter 2 of *Kinds of Minds*.

<sup>20</sup> *The Intentional Stance*, p. 112.

<sup>21</sup> *The Intentional Stance*, p. 294. It is little wonder why Dennett would find it important to debunk any notion of a special “original”, or non-derivative, type of intentionality above and beyond that typically attributed to organisms and artifacts. Throughout his work, Dennett’s favorite tactic has been to consider the attribution of intentionality to “blessedly simple” things like frogs and chess-playing computers, and then to apply the conclusions he reaches to the attribution of intentionality to much more complicated kinds of things like people. See, for instance, *The Intentional Stance*, pp. 32-3.

<sup>22</sup> For the classic presentations of these thought experiments, see Putnam, “The Meaning of ‘Meaning’”, in *Mind Language, and Reality: Philosophical Papers, volume 2* (Cambridge: Cambridge: 215-271, 1975) and Burge, “Individualism and Psychology,” *Philosophical Review* 95 (1987).

<sup>23</sup> See Searle’s *Intentionality* (Cambridge: Cambridge, 1983) and “Minds, Brains, and Programs,” reprinted in Haugeland (ed.), *Mind Design II* (Cambridge, MA: MIT Bradford, 1997, pp. 183-204). Searle explicitly addresses (unecumenically!) the topic of animal intentionality in “Animal Minds,” *Midwest Studies in Philosophy* XIX (1994): 206-219.

<sup>24</sup> This is supposed to be the moral Searle draws from his infamous “Chinese Room” thought experiment. There Searle argues that no machine following a program could possibly exhibit the “intrinsic” intentionality humans possess. See in particular “Minds, Brains, and Programs.”

<sup>25</sup> This idea also lies behind the claim that robots could only have derived intentionality.

<sup>26</sup> Similarly, an account of what language is *for* need not be an account of what language *is*.

<sup>27</sup> So there is still room for an intentional realist to claim that cosmic coincidences like Davidson’s “swampman” could have original intentionality.

<sup>28</sup> See in particular Dretske, *Explaining Behavior* (Cambridge, MA: MIT Bradford, 1988). In “Discrimination Without Indication: Why Dretske Can’t Lean On Learning,” *Mind and Language*, 9 (1992): 163-180, Slater argues that Dretske has misdescribed what goes on in reinforcement learning.

<sup>29</sup> See Dretske’s discussion of the so-called “design problem” in *Explaining Behavior*, pp. 96ff.

<sup>30</sup> See, for instance, Staddon, *Adaptive Behavior and Learning* (Cambridge: Cambridge, 1983), pp. 414ff. and Dickinson, “Expectancy Theory in Animal Conditioning,” in Klein and Mowrer (eds.), *Contemporary Learning Theories* (Hillsdale, NJ: Lawrence Erlbaum, 1989).

<sup>31</sup> It bears mentioning that I’m not trying to show that any particular creatures are expectation-mongers. That is the work of ethologists, not philosophers. Notice that I’ve defined expectation-mongering in terms of how a creature would behave in various possible situations. Since any pattern of actually observed behavior could be the product of tropisms, showing that a creature is an expectation-monger would have to involve establishing that certain counterfactuals hold. It turns out, then, that those who *design* devices would have a much easier time justifying the attribution of expectations to their subjects, simply because they have a much better sense of what goes on inside the “black boxes” they are studying, and so would have a better grasp of the relevant counterfactuals. For a useful discussion of the difficulties attributing to wild subjects states similar to the expectations described here, see Heyes and Dickinson, “The Intentionality of Animal Action,” *Mind & Language*, 5 (1990): 87-104.

<sup>32</sup> Different accounts of expectation-based educability differ with respect to which expectation components may vary from expectation to expectation, which components are capable of being altered, and also the conditions in which they stand to be adjusted. A fully general account of expectation would leave as much of this up for grabs as possible.

<sup>33</sup> In a similar fashion, we can determine a (possibly non-transitive) preference ordering among outcomes.

<sup>34</sup> I am aware that certain hard-nosed physicalists (e.g., Kim) might look askance at such “dispositional” explanations of behavior as, at best, incomplete. I have not set out to assuage the concerns of physicalists, who are generally suspicious of intentional explanations anyway. There are plenty of others (e.g., Dretske) engaged in that project! In particular, see Bennett “Folk-psychological explanations,” in Greenwood (ed.), *The Future of Folk Psychology: Intentionality and Cognitive Science* (Cambridge: Cambridge, 1991): 176-195 and Dretske, *Explaining Behavior*.

<sup>35</sup> Thus there is some basis for thinking that expectations (and even goals) can be *causes* of a creature’s behavior.

---

<sup>36</sup> Notice that this story begins with an account of expectation and then works out an accompanying account of goals; it does *not* go the other way around. That is, this account reverses the strategy advocated by Ramsey and actually pursued by most learning theorists.

<sup>37</sup> As Bennett might claim, a critter will be disposed to attain its goals “all things being equal,” and having correctly configured expectations is part of things being equal. See *Linguistic Behavior* (Indianapolis, IN: Hackett, 1990), pp. 42ff.

<sup>38</sup> Notice that this account doesn’t rule out “accidental” (or unexpected) success at attaining goals.

<sup>39</sup> That is, this story does not appeal to so-called “normative constraints” on the attribution of intentionality, such as Dennett’s “assumption of rationality” or Davidson’s “principle of charity”; interpreters are not constrained to attribute expectations that are for the most part correct. See Dennett, *The Intentional Stance*, pp. 17ff. and Davidson, “On the Very Idea of a Conceptual Scheme,” in *Inquiries into Truth & Interpretation* (Oxford: Oxford, 1984), p. 196.

<sup>40</sup> See, for instance, “Rational Animals” pp. 474-476.

<sup>41</sup> Against the background of this account of expectation error, we can understand educable creatures to be making revision errors whenever they adjust an expectation in ways that would render it more susceptible to either errors of commission or errors of omission.

<sup>42</sup> See Davidson’s “Thought and Talk,” reprinted in his *Inquiries into Truth and Interpretation* (Oxford: Clarendon, 1984) and “Rational Animals.” To be sure, Davidson tries there to argue that the conceptual resources required to be surprised in turn require an animal to be capable of interpreting the utterances of others; thought requires talk. However, we don’t have to accept this argument to take Davidson’s point that the capacity to be surprised, or to recognize when the way things are aren’t as one took them to be, is an important part of being a rational animal.

<sup>43</sup> Of course, such a being would lack an appropriate history to possess the relatively derived, biological sorts of intentionality described by philosophers like Millikan and Dretske.

<sup>44</sup> Cheney and Seyfarth thus devote a significant portion of *How Monkeys See the World* to the question of whether vervets make their alarm-calls with the intention of communicating to other vervets.

<sup>45</sup> Once again, see “Rational Animals” and “Thought and Talk.” It is worth mentioning in Davidson’s defense that attempts to determine at what point children acquire a “theory of mind” plausibly amount to determinations of what age they typically master the concept of belief in natural language.

<sup>46</sup> The classic source is Premack and Woodruff, “Does the Chimpanzee Have a Theory of Mind?” *Behavioral and Brain Sciences*, 1 (1978): 515-26.

<sup>47</sup> See in particular, Whiten and Byrne, “Tactical Deception in Primates,” *Behavioral and Brain Sciences* (1988): 233-44 and Byrne, “What’s the Use of Anecdotes? Distinguishing Psychological Mechanisms in Primate Tactical Deception,” in Mitchell, Thompson, and Miles (eds.), *Anthropomorphism, Anecdotes, and Animals* (Albany, NY: SUNY, 1997).

<sup>48</sup> See Byrne and Whiten (eds.), *Machiavellian Intelligence* (Oxford: Oxford, 1988) and de Waal, *Chimpanzee Politics* (Baltimore, MD: Johns Hopkins, 1989).

<sup>49</sup> Compare this to Byrne’s definition of “tactical deception” as “goal-directed acts, whose success depends upon another animal being deceived.” (“What’s the Use of Anecdotes? . . .,” p. 135)

<sup>50</sup> Byrne’s definition of “tactical deception” does not adequately distinguish these two possibilities. Indeed, Byrne seems to despair of ever getting evidence that primates truly deceive their conspecifics.

<sup>51</sup> See in particular Dennett, *The Intentional Stance*, chapter 7. Note that one doesn’t really need to *rule out* these lower-level interpretations so much as to show that they are less plausible or more *ad hoc* than their higher-level rivals. Furthermore, there is no need to work under the *presumption* (often attributed to Lloyd Morgan) that animals do *not* have higher-order intentional capacities. We don’t approach other people that way; why should we approach animals any differently? Instead, we simply ought to attribute to animals the intentionality that they have!

<sup>52</sup> That is, this would be evidence of a creature’s adopting more than just a “biological stance” toward another. Note also that Byrne overlooks this type of systematicity when he discusses what kind of behavior would signify an animal’s *intentionally* trying to deceive another. See Byrne, *op. cit.*, pp. 146ff.

<sup>53</sup> Note that this implies more activity on the part of the instructor than what goes on in mere imitative learning.

<sup>54</sup> See Savage-Rumbaugh, Rumbaugh, and Boysen, “Linguistically Mediated Tool Use and Exchange by Chimpanzees (Pan troglodytes),” *Behavioral and Brain Sciences*, 1 (1978): 539-54.

---

<sup>55</sup> Of course, this is ultimately an empirical issue. Evidence here is anecdotal at best, but now we have an account of why we *should* be intrigued by Kanzi's apparent attempts to get his sibling to sign along with him. See Savage-Rumbaugh and Lewin, *Kanzi: The Ape at the Brink of the Human Mind* (John Wiley, 1994).

<sup>56</sup> I think it worth pointing out that I've arrived at this conclusion without having to talk about anything as esoteric as grammar. Of course, apes don't seem particularly capable of following complex grammatical rules, let alone enforcing them. However, I'm perfectly willing to countenance the possibility of languages that are largely bereft of grammatical structure.

<sup>57</sup> I'd like to thank Bob Brandom, Laura Ruetsche, and especially Jerry Massey for kindly taking the time to comment upon various stages of this work. I've also benefited from presenting earlier drafts of the first part of this paper to Jerry's seminar in zoological philosophy at the University of Pittsburgh and also to an audience at the University of Nevada, Las Vegas.