ANIMALS AS REFLEXIVE THINKERS: THE APONOIAN PARADIGM

Mark Rowlands
Department of Philosophy
University of Miami
Coral Gables, FL 33124

Susana Monsó
Department of Logic, History and Philosophy of Science
UNED, Spain

1. It's Complicated

Reflexive thought is thought about thought, or thought about other mental states more generally. As such, the ability to engage in reflexive thought is generally regarded as a complex intellectual achievement: one that is beyond the capacities of most animals – indeed, perhaps all animals except humans. Denial of this ability can be made on a variety of grounds. First, many argue, one cannot think a thought about any given mental state without having the concept of that mental state. And so, it is claimed, attributing the capacity for reflexive thought to non-human animals (henceforth "animals") would entail attributing to them an implausibly sophisticated conceptual repertoire. In addition to the issue of burgeoning conceptual repertoire is another – quasi-empirical – objection. If a creature has the ability to engage in reflexive thought, then it must have the ability to attribute mental states both to itself and others. It is argued that there is no empirical evidence for this ability in any non-human animals. Any apparent evidence in favor of

this ability can always, it is argued, be explained in more parsimonious terms: for example in ways that involve an ability for behavioral abstraction (i.e. to form generalizations about behavior and its likely consequences) but not the ability to think thoughts about what is going on in the mind of another.[1] Assessing these arguments against the possibility of reflexive thought in animals is a huge undertaking – both empirically and conceptually – one that has already generated countless books and journal articles, with no sign of resolution on the horizon.

We shall not address this question directly. Instead, our approach will question the importance usually attached to the issue of whether animals are reflexive thinkers. This importance derives from the belief that the capacity for reflexive thought is built into, or required for, many other capacities. If animals lack the capacity for reflexive thought they, therefore, must also lack these other capacities. This is the idea that we shall resist. At the heart of this, we shall argue, is a pervasive tendency, shared by philosopher and scientist alike, toward over-complication.

Suppose there is a property – let us call it P – that, common sense decrees, is widely distributed in nature: being possessed not only by normal adult human beings, but also by children and at least some animals. While property P is, on some level, mundane and familiar – that is the basis of confidence in the widespread distribution of this property – the precise theoretical articulation of P is controversial. There is a range of theoretical options that might be used to capture this property P, running from the simple to the complex. On some of the more complex options, it turns out that the distribution of P will not be as wide as common sense supposes: for example, the possession of P by

[1] Povinelli & Vonk (2003, 2004); Penn & Povinelli (2007, 2013).

children and other animals will be rendered problematic, unlikely or impossible. Thus, there is a clash between common sense ("P is widely distributed through the human and animal world") and theory ("P is probably/definitely restricted to normal adult humans"). We think it is fair to say that, in this sort of case, there has been a persistent historical tendency, in both philosophical and scientific treatments of animals, to favor the restrictive theory over the more liberal common sense. Indeed, in philosophy, a few dissenting voices aside, this tendency is endemic, and almost definitive of the attitude that, historically speaking, this discipline has taken towards animals. It is unclear why this should be. Philosophers are, perhaps, complicated people, and have a natural proclivity to favor the complex over the simple. But this tendency is not restricted to (professional) philosophers. There is, we think, a lot of truth in Wittgenstein's (implicit) claim that we are all philosophers.[2] Scientists also frequently find themselves in the grip of philosophical assumptions and confusions, and the scientists who study animals are no exception.

The approach we are going to defend, we shall refer to as *aponoian*. *Aponoia* comes from the Greek, *apó*, which means 'away from', 'separate', 'without', and *noûs*, which means 'intelligence', 'thought'. An *aponoian* approach to psychological abilities is, accordingly, one that aims to leave intelligence and thought aside.[3] This does not mean – and we can't really emphasize this enough – that animals are lacking in intelligence. *Aponoia* is something that applies to humans and other animals equally. The idea we

[2] Ludwig Wittgenstein, "Philosophy is a tool useful only against philosophers and the *philosopher in us*." MS 219, emphasis is ours.
[3] We would like to thank Luis Gil for this term.

mean to convey is this: seemingly complex psychological abilities are often not as complex as they seem. Intelligence is, in this sense, often not as intelligent as it seems.

2. Two Mistakes

This paper will inveigh against the philosophical art of (over-) complication. The endemic over-complication in question can take a variety of forms but two, in particular, stand out.

Suppose there is a property P, which might, in different contexts, stand for any number of psychological traits and/or abilities, including consciousness, emotions, empathy, beliefs (and other propositional attitudes), action and agency (including moral agency). One type of complication, then, consists in a rather unseemly rush to go *meta*: to assume that articulation of a given psychological phenomenon, P, requires appeal to a meta-level phenomenon of some sort: awareness of P, thoughts about P, the ability to scrutinize P, concepts of P, and so on. We shall refer to this baleful rush to the meta-level – in contexts where it is neither necessary nor fruitful – as a tendency toward *premature meta-articulation*. Obviously, this tendency is directly related to the issue of the importance of reflexive thought in animals. To appeal to the meta-level in explanation of a given mental phenomenon – consciousness, emotions, beliefs, empathy, and so on – is to suppose that it is not possible for a creature to possess or exhibit that phenomenon unless it is capable of engaging in reflexive thought of some sort. Conversely, if

possession of these things does not require a meta-level explanation, then the lack of the ability to engage in reflexive thought does not preclude possession of these phenomena.

Second, there is another tendency that, unfortunately, stubbornly resists our attempts to provide it with a catchy appellation. This tendency is best introduced by example. Suppose we are tempted to say, on the basis of its behavior, that one animal is the subject of a given emotion, say fear. The little dog, let us suppose, fears the big dog. Given an array of behavioral, evolutionary and neurobiological evidence, our attribution of fear to the dog does not seem to be an unreasonable one. However, suppose the 'philosopher in us', as Wittgenstein put it, urges caution: to fear the big dog, the little dog would have to understand that the big dog is worthy of fear – that it is the sort of thing that *should* or *ought* to be feared. But now we seem to be attributing to the little dog the concept of ought or should – the concept of *warrant*, as we might say. And this does seem to be an overly sophisticated concept. Therefore, we might find the philosopher in us urging us to abandon the idea that the dog can possess the emotion of fear.

This, we shall argue, would be an example of confusing the ability to *make* a certain judgment and to *track* a certain judgment. Accordingly, we shall refer to it – and we note, once again, that if there is a catchier appellation for this confusion, it is beyond the abilities of the authors to devise – as the *making/tracking confusion*. The little dog, we shall argue, does not need to make the judgment that the big dog *ought* to be feared, or that it *warrants* fear. Rather, a far weaker condition is all that is required: if the dog's emotion is not, as we shall say, *misguided* (a technical term to be explained later) then the claim that the big dog ought to be feared must be true. In such circumstances, we shall

say, the dog's emotion *tracks* the truth of the claim. The little dog does not need to judge that the big dog ought to be feared, as long as his or her emotion tracks the truth of the claim.[4] The making/tracking confusion is important in its own right. In this paper, however, given that our primary concern is the issue of reflexive thought in animals, we shall tend to focus on the way it has been used to support the error of premature meta-articulation.

3. Consciousness

Descartes claimed, notoriously, that animals lack minds, and most have interpreted this claim as a denial that animals are conscious.[5] This encouraged generations of Port-Royal scientists to nail living, conscious animals to boards and dissect them. The screams and screeches, they assured themselves and others, were just the rubbing together of various mechanisms, and should not be taken to indicate anything about the mental life of the animal – for there was no such mental life. Equally notoriously, Descartes had no convincing arguments for this dismissal of the mental lives of animals. When pushed, he resorted to the observation that if animals had minds they would have immortal souls, and they could not be reasonably credited with these.[6]

---

[4] This idea is developed in more detail in Rowlands (2012), chapter 2.

[5] Interpretations of Descartes diverge. Some claim that while he was committed to denying them thought or reason, Descartes did allow that animals were sentient, or could feel. Cottingham (1978). This animal-friendlier interpretation of Descartes has been disputed by Sztybel, and in the opinion of at least one of the authors is dubiously compatible with many of the things Descartes (1927) asserts in 'On the automatism of brutes'. However, Descartes scholarship is not our business here, and so we shall simply note that the denial of any sort of mentality – thought and feeling – to animals is a common interpretation of Descartes.

[6] Descartes, (1927) p. 357.

One might think that, today, we have shaken off this post-medieval nonsense, but the claim that animals lack consciousness has, in fact, been defended in recent philosophical discourse. This defense turns on what is known as the *higher-order thought* (HOT) model of consciousness. We discuss this not because the view is widespread – even amongst philosophers, few today are willing to bite the bullet and deny consciousness to animals – but because it is a glaring, and so for our purposes useful, example of the pitfalls of assuming a phenomenon must be explained by appeal to the meta-level.

The sense of consciousness in question is *phenomenal*: the way it seems or feels when one has or undergoes an experience. The overwhelming preponderance of the scientific evidence suggests that this sort of consciousness is possessed by most, perhaps all, vertebrates, and some invertebrates and so, at a conservative estimate, probably made its first appearance (on this planet at least) 300-500 million years ago. However, some proponents of HOT have contested this claim.[7]

In order to understand the HOT model of consciousness, two distinctions are required. The first is between *creature* and *state* consciousness. Consciousness can be ascribed to both creatures and mental states. A creature can be conscious in the sense that it is awake as opposed to asleep. But a mental state – a desire, for example – can also be conscious or unconscious. The second distinction is between transitive and intransitive consciousness. Transitive consciousness is consciousness *of* something. If I (consciously)

[7] Notably, Peter Carruthers (1989). Not all defenders of HOT, by any means, will endorse this conclusion. Indeed, most regard such an implication as a *reductio* of the HOT account, and so seek to distance themselves from this implausible conclusion by trying to find ways to show why HOT does not entail it. See, for example, David Rosenthal (2004).

think that the cat is on the mat, then I am transitively conscious of this state-of-affairs. Creatures are transitively conscious of things; mental states are not. My thought that the cat is on the mat is not conscious of anything. I am conscious of the cat being on the mat in virtue of having this thought. My thought that the cat is on the mat, on the other hand, is intransitively conscious when I am (consciously) thinking it. Based on these distinctions, we can express the guiding idea behind HOT accounts as follows: Intransitive state consciousness is to be explained in terms of transitive creature consciousness.

According to the HOT model of consciousness, for a given mental state of mine – say pain – to be conscious, it is necessary that I form (or, on some versions, *be able to form*) a thought about this pain – a thought to the effect that I am in pain. This thought confers consciousness on my pain. Until I form the thought – or unless I possess the ability to form the thought – my pain is non-conscious.[8] The thought makes me transitively conscious of my pain, and in doing so makes my pain intransitively conscious.

The HOT model is implausible. In particular, it seems to fall foul of a dilemma. Is the higher-order thought (intransitively) conscious? If it is, then HOT does not explain consciousness but presupposes it. If it is not, then it is utterly mysterious how the thought is supposed to make me transitively conscious of my pain. Intransitively unconscious

[8] HOT accounts come in two forms – actualist and dispositionalist. According to actualist versions, for my pain to be conscious, I must actually think that I am in pain. According to dispositionalist versions, I need only be able to from the thought. The differences between these two versions of the HOT account are not important for our purposes.

thoughts do not make their subjects transitively conscious of their objects – that is precisely what it is for them to be intransitively unconscious.

Suppose, for example, I think unconsciously – perhaps due to various mechanisms of repression – that someone very close to me is seriously ill. What would this mean? We might explain it in terms of various unexplained feelings of melancholy that assail me when I am talking to them, or a vague sense of foreboding that I can't quite pin down. However, what the thought cannot do, in its unconscious form, is make me aware of the fact that the person is seriously ill. Because as soon as it does that it becomes, by definition, a conscious thought. To become aware of the fact that my friend is seriously ill is to consciously think that my friend is seriously ill. That is, the thought has become intransitively conscious. That a thought does not make me aware of what it is about is precisely what it means for the thought to be intransitively unconscious. Conversely, as soon as it does make me aware of what it is about, it becomes a conscious thought – because making me aware of what it is about is precisely what it is for the thought to be conscious.[9]

The appeal to a higher-order thought to explain consciousness does not work. This is an example of premature meta-articulation. In this case, the premature meta-articulation seems to be based on the attribution of seemingly miraculous powers to the meta-level: the idea that the meta-level can somehow magically bestow something on a range of phenomena that is intrinsically lacking in the phenomena themselves. This form

[9] See Rowlands (2001a) and (2001b), chapter 5.

of premature meta-articulation, therefore, is grounded in what we might call the *miracle-of-the-meta*.[10] This is the general form of the miracle:

1. First-order phenomena a, b, c etc. are intrinsically lacking in some property P.

(In this case, P = phenomenal consciousness).

2. Higher-order phenomena x, y, z take a, b, c, as their objects, and in so doing confer P on a, b, c.

The problem – the reason why this schema is miraculous – lies in this dilemma:

3. Do x, y, z possess P?

4. If so, then we have not explained P but simply presupposed it.

5. If not, then it is mysterious how x, y z can confer P on a, b, c.

In other words, the appeal to the meta-level cannot do the work it is supposed to do, and is therefore fruitless.[11] While many prominent forms of premature meta-articulation are grounded in the miracle-of-the meta, as we shall see, this is not true for all cases. Therefore, we shall treat premature meta-articulation and the miracle of the meta as distinct fallacies, with the latter being a category of the former.

The usefulness of the HOT account as an example of premature meta-articulation is limited, because the account has not achieved widespread acceptance. We use it

[10] We understand the relation between premature meta-articulation and the miracle-of-the-meta as one of genus to species. All cases of the miracle-of-the-meta are cases of premature meta-articulation, but not the other way around.
[11] Rowlands (2012), chapter 6.

because it provides an exceptionally clear example of premature meta-articulation. As we shall see, there are other examples of premature meta-articulation that many find more plausible. Plausible or not, we shall argue that they suffer from the same deficits.

4. Belief I: Premature Meta-Articulation

Some have argued that animals cannot have beliefs. In this, and the following section, we shall argue that their case invariably depends either on premature meta-articulation or the making/tracking confusion. This section deals with premature meta-articulation about belief.

Donald Davidson argues that "dumb" animals (i.e. animals incapable of engaging in linguistic communication) are incapable of having beliefs: "First, I argue that in order to have a belief, it is necessary to have the concept of belief. Secondly, I argue that in order to have the concept of belief one must have language."[12] The requirement that one possess the concept of belief in order to possess a belief may seem, prima facie, unduly intellectualistic. Davidson thinks otherwise:

> Here I turn for help to the phenomenon of surprise, since I think that surprise requires the concept of a belief. Suppose I believe there is a coin in my pocket. I empty my pocket and find no coin. I am surprised. Clearly enough I could not be surprised (though I could be startled) if I did not have beliefs in the first place. And perhaps it is equally clear that having a belief, at least one of the sort I have

[12] Davidson (1975) and (1985)

> taken for my example, entails the possibility of surprise. If I believe I have a coin in my pocket, something might happen that would change my mind. But surprise involves a further step. It is not enough that I first believe there is a coin in my pocket, and after emptying my pocket I no longer have this belief. *Surprise requires that I be aware of a contrast between what I did believe and what I come to believe*. Such awareness, however, is a belief about a belief: if I am surprised, then among other things I come to believe my original belief was false.[13]

To be surprised, one must be able to have a belief about a belief. However, one cannot have a belief about a belief unless one has the concept of belief. But to have the concept of belief requires the concept of objective truth: "Much of the point of the concept of belief is that it is the concept of a state of an organism which can be true or false, correct or incorrect. To have the concept of belief is therefore to have the concept of objective truth."[14] But the concept of objective truth, Davidson argues, for reasons deriving from his semantic theory, is not possible for creatures lacking in language.

There is much about this argument that can be questioned.[15] However, we shall simply focus on stopping the argument before it starts. Davidson's conception of surprise is a meta-cognitive one. To be surprised requires (i) being aware of two distinct beliefs, and, in virtue of this, (ii) being aware of the contrast between them. The claim that animals cannot be surprised is, of course, itself a rather surprising one in that it contradicts a wealth of evidence, scientific and anecdotal, that suggests surprise is rather

[13] Davidson (1985), p. 478. Emphasis is ours.
[14] Davidson (1985), p. 478.
[15] Indeed, Davidson often puts the term 'argument' here in scare quotes, recognizing that the argument is far from compelling.

widespread in the animal kingdom. Let's us take a case of apparent surprise. Hugo, a dog, requests to be let out of the back door for his nightly constitutional. It has been raining, and a rather large American bullfrog sits outside, a few feet away. Hugo exits in his usual way, but upon noticing the frog, freezes for around thirty seconds, staring intently.

It would be implausible to claim that nowhere in this little tableau is there any element of surprise. It would also be implausible to attribute to Hugo an awareness of the contrast between his initial and subsequent beliefs about the state of the patio vis-à-vis large American bullfrogs. However, to explain Hugo's surprise, it is possible to take an *aponoian* approach in which there is no need to attribute to him any such thing. Within this approach, all that is required to explain surprise is the postulation of a first-order mechanism that records a discrepancy between the content of a belief and the way the world is.[16] Suppose we grant that Hugo has a dispositional belief about the patio of roughly this form: when I go out the door, things will be more or less the same as they usually are. The postulated mechanism works by detecting a discrepancy between this belief and the way the world, in fact, is. Hugo is, as a result of this discrepancy, surprised. Or, better, surprise is the experiential form the detection of this discrepancy takes. Surprise, therefore, does not need to be explained meta-cognitively.

The mistake Davidson has made, in effect, is to confuse awareness of the *contents* of beliefs with awareness of *beliefs*. The content of a belief is what the belief is about. What the belief is about will be, roughly, a *state-of-affairs*: an arrangement of objects and properties in the world. To have a dispositional belief is to be disposed to entertain content – to believe that a certain state-of-affairs is the case – under certain eliciting

[16] This point has been made by Peter Carruthers (2008)

conditions. The content of Hugo's belief, we have supposed, is that *the patio is more or less the way it usually is: no surprises there*. This belief, presumably, exists in dispositional form: Hugo does not need to be consciously thinking this to himself. There is, however, a surprise there, as Hugo quickly perceives. Hugo, thus, becomes aware of the new content: thing, there! And so the surprise-detecting mechanism kicks in to detect the difference between the content of his perception and the content of his dispositional belief. In no part of this account do we need to postulate meta-cognitive abilities or arrangements such as beliefs about beliefs. Davidson's position is, then, an example of unnecessary premature meta-articulation.

5. Belief II: The Making/Tracking Confusion

Davidson, along with Stephen Stich and others, has another argument against the possibility that animals can believe. In this section we shall try to show this argument falls victim to the *making/tracking confusion*.

Consider the following scenario, based on Malcolm (1973). A dog chases a squirrel up a tree. The squirrel jumps from one tree to the next, and eventually disappears. The dog does not see this, and sits at the foot of the tree barking. It is natural to explain the dog's behavior in terms of his or her belief that the squirrel is in the tree (conjoined with, perhaps, its desire to catch the squirrel, or its frustration at not being able to do so, and so on). It cannot, after all, see that the squirrel is in the tree – the squirrel is no longer there.

Davidson (1975, 1985) and Stich (1979) disagree with this interpretation of the situation. Davidson puts his argument as follows:

> Can the dog believe of an object that it is a tree? This would seem impossible unless we suppose that the dog has many general beliefs about trees: that they are growing things, that they need soil and water, that they have leaves or needles, that they burn. There is no fixed list of things someone with the concept of a tree must believe, but without many general beliefs there would be no reason to identify a belief as a belief about a tree, much less an oak tree. Similar considerations apply to the dog's supposed thinking about the cat.[17]

The more general moral of these considerations is:

> We identify thoughts, distinguish between them, describe them for what they are, only as they can be located within a dense network of related beliefs. If we really can intelligibly ascribe single beliefs to a dog, we must be able to imagine how we would decide whether the dog has many other beliefs of the kind necessary for making sense of the first.[18]

Davidson focuses on the tree. We'll focus on the squirrel. Does the dog – let us call him Hugo – believe that the squirrel is a mammal, that it is warm-blooded, that it has a skeleton, and so on? All these beliefs, Davidson claims, are part of our concept of a squirrel, and so without them Hugo cannot share our concept. Therefore, the attribution to Hugo of the belief that there is a squirrel in the tree is problematic. The attribution of a

[17] Davidson (1985), p. 474.
[18] Davidson (1985), p. 475.

belief about the squirrel to Hugo depends on his possession of the concept of squirrel. However, possession of a concept depends on possession of a network of related beliefs. Therefore, attribution of beliefs (and other propositional attitudes) is a holistic enterprise. Roughly:

> Attribution-holism: The attribution of a single belief or other propositional attitude to an individual requires, and only makes sense in terms of, the attribution of a network of related beliefs.

This attribution-holism precludes attribution of beliefs to individuals who do not share our belief-network. Hugo, along with all other animals, is such an individual.

We might call this the 'anchoring' argument. The content of any concept is anchored to a network of related beliefs. The human concept of squirrel is anchored to the network of related beliefs shared by humans. We have to suppose that Hugo does not possess this concept. But, therefore, we cannot attribute beliefs about squirrels to Hugo, for when we do so we employ a concept ('our' concept of squirrel) that he does not possess. More generally, the attribution of individual beliefs to individuals is constrained by the networks of beliefs they hold. If this network is not shared with us, we cannot attribute beliefs to them, for such attribution would be predicated on concepts they do not possess.

The anchoring argument is unconvincing. At the heart of it lies an equivocation between the issue of (i) whether animals have beliefs at all and (ii) which beliefs they have. At most, the argument shows that we may not be able to attribute beliefs to animals

because of a divergence in the content their beliefs would possess and the concepts we would employ in ascribing those beliefs. The argument does not show that animals do not possess beliefs, merely that we cannot specify the content of their beliefs, and so cannot ascribe beliefs to them. But if that is the problem, then there is a well-known apparatus for getting around it.

Here is a famous philosophical thought experiment. It sometimes gives philosophers a bad name, because it seems so far-fetched but, in fact, it is just a way of making graphic a simple point. There is a planet – Twin Earth, which exactly duplicates Earth in every respect bar one: there is no water on Twin Earth. Instead, there is a substance that looks, tastes, and feels exactly like water, and that fills the oceans and rivers, and emerges from the faucets, of Twin Earth. A molecule-for-molecule twin of someone on Earth speaks twin English, and so has beliefs that he or she would express with sentences of the form, 'Water is wet', for example. However, the Twin can have no water-beliefs. There is no water on his or her planet. He or she has only ever been in contact with this other substance. We can call it *retaw*, for ease of identification: The Twin only has retaw-beliefs (though speaking Twin English he or she would express these beliefs using the term 'water'. He/she calls retaw 'water'). The point of the thought experiment is this: a person's beliefs can vary even though everything in his or her head – they are molecule-for-molecule duplicates – remains the same. But that doesn't matter for our purposes.

Suppose we accept the conclusion of this thought experiment. Then, if we explain the Twin's retaw-drinking behavior through postulating a desire to quench his/her thirst

and a belief that water will quench it, our explanation would be false. Nevertheless, there is surely something about it that is right. It is not as if we tried to explain his/her behavior by way of the desire to quench his/her thirst and the belief that water is poisonous, or the belief that colorless green ideas sleep furiously. The explanation may not be strictly correct, but it is *not far off* the truth.

The crux is how to explain the idea of being not far off the truth, and there is a way of doing this. The truth of the claim (or proposition) that water is wet guarantees the truth of the claim that retaw is wet. If the former proposition is true, then the latter must be true also. More than this, the guaranteeing of truth derives from the fact that there is a reliable connection between the properties of water and the properties of retaw: if water is wet, colorless, odorless, transparent, thirst-quenching etc. then retaw must be these things too. This reliable connection between properties is simply a feature of the way the thought experiment is set up.

We can apply this general idea to Hugo the dog. Suppose Hugo represents the squirrel in, for example, affordance-based terms. That is, the dog represents the squirrel as a *chase-able* thing. This is, of course, an empirical matter, but suppose, for the sake of argument, it is correct. Corresponding to the proposition that we entertain, namely that the squirrel is in the tree, Hugo thinks a thought along the following lines: the chase-able thing is up there. Can we still, legitimately use our proposition to explain Hugo's behavior? The answer, we suspect, is that we can, and while we won't be strictly correct, what we say is close enough to the truth to be useful and enlightening. More precisely, what we say will be useful and enlightening if the following two conditions – designed to

parallel the Twin Earth case – hold. First, the truth of the (anchored to us) claim that there is a squirrel in the tree guarantees the truth of the (anchored to Hugo) claim that the chase-able thing is up there. If the former is true, then the latter is true also. Secondly, this guaranteeing of truth holds because there is a reliable connection between the properties of squirrels and chase-able things: squirrels, for Hugo, are reliably chase-able (and things in trees are reliably up and there).

Consider the first condition. If the (anchored to us) claim that the squirrel is in the tree is true then the (anchored to Hugo) claim that the chase-able thing is up there must also be true. The truth of the first anchored claim guarantees the truth of the second one. Moreover, and this is the second condition, the reason the first anchored claim guarantees the truth of the second is because of a reliable connection between the property of being a squirrel and the property of being a chase-able thing. For Hugo, squirrels are reliably chase-able: that is, for any x, if x is a squirrel then x, for Hugo, is chase-able. When these first and second conditions are met we can say that the (anchored to us) claim that the squirrel is in the tree *tracks* the (anchored to Hugo) claim that the chase-able thing is up there. There is a truth-guaranteeing relation between the two claims, where this is grounded in a reliable connection between the properties of the thing (the squirrel) the claims are about.[19]

Note, once this apparatus is accepted, we don't even need to know what beliefs Hugo has vis-à-vis squirrels. He may represent them as chase-able things or by way of some other categories entirely. All that is required for the attribution of the belief that the squirrel is in the tree to be useful and enlightening (if not strictly true) is that the

[19] See Rowlands (2012), chapter 2 for an elaboration of this argument.

(anchored to us) proposition that the squirrel is in the tree track whatever proposition it is that can truly be employed in attributing the belief to Hugo.

Davidson's argument against attributing beliefs to animals, therefore, is a version of the making/tracking confusion. It assumes that attributions of beliefs and other propositional attitudes to an animal can be legitimate only when the animal is capable of entertaining a given claim or proposition – that is, capable of making a given judgment. We have argued that this is too strong. There are various ways in which an attribution of belief to an animal can be legitimate. One of these ways is that the attribution be useful or enlightening – that it allows us to make sense of the animal's behavior. But this condition, we have argued, does not require that the animals be able to make the judgment, or entertain the proposition, implicated in the belief that we attribute to the animals. Rather, all that is required for the belief-attribution to be enlightening – that is, to have explanatory value – is that there is an appropriate relation of tracking – in the sense explained above – between the thought the animal actually thinks (which is an empirical matter) and the thought we attribute to it.

6. Emotion

A Capuchin monkey sees its fellow being rewarded with (highly prized) grapes for completing a given task. Upon completing the same task, this monkey is given a (not at all highly prized) piece of cucumber. After several repetitions, the seemingly enraged

monkey hurls its cucumber out of the cage at the researcher.[20] If the players in this scene were human, it would be natural to describe their behavior by appeal to the emotion of indignation.

Jaak Panksepp has argued, on neurobiological grounds, that basic emotions such as happiness, sadness, fear, anger, surprise and disgust extend beyond the human realm, encompassing all mammals, in all likelihood birds, and possibly reptiles.[21] This is not a minority view in affective neuroscience. Panksepp's view also coincides – with the possible exception of reptiles – quite closely with common sense. Arrayed against common sense and affective neuroscience, however, we find philosophers who regard the attribution of any emotions to animals as deeply problematic – and the attribution of a fairly complex emotion such as indignation especially so.

There is a tendency to think that emotions are somehow more primitive than cognitive states such as belief. It is unclear from where this idea derives, but its legitimacy is very questionable. Emotions are, at least conceptually, more complex than cognitive states. An emotion contains everything a belief contains and more.

Emotions are distinct from moods. Like beliefs and other cognitive states, emotions have intentional content. Fear is fear *of* something, or that something will happen. Anger is directed at someone because they did something. This means that emotions have intentional content. Indeed, they are individuated by this content. If the monkey were indeed indignant, the content of his indignation would be *that* he is being

[20] Brosnan & de Waal (2003)
[21] Panksepp (1998). Panksepp questions whether surprise and disgust should be classified as genuine emotions rather than simpler types of motivational state.

offered a cucumber (when his fellow Capuchin is being offered a grape). This is what individuates the emotion – distinguishes it from other cases of indignation. *That* he is being offered a cucumber is what we might regard as the *factual* content of the Capuchin's emotion. In their possession of factual content, emotions are akin to beliefs.

Emotions are different from beliefs, however, in that there is more to their content than the factual. Implicit in the monkey's indignation would be the evaluative content that his being offered the grape is wrong. The content of the emotion is composed of the factual judgment ('I am being given a cucumber, again'), and the moral judgment ('This is wrong!' or 'I am being wronged!'). This seems to be a moral judgment. Not all emotions involve specifically moral judgments. But all involve evaluative judgments of some sort. If the little dog does, indeed, fear the big dog, then implicit in this, it seems, is the judgment that the big dog is worthy of fear – that it *should* be feared.

Because emotions have both factual and evaluative content, philosophers skeptical of the idea that animals can have emotions have two different options for developing their case. They might contest the claim that animals can entertain factual content. We have already discussed this idea in the previous two sections. The other version of the case turns on hostility to the idea that animals can make moral or other evaluative judgments required for possession of emotions. That is the avenue of hostility we shall examine in this section. We shall argue that this idea is an example of the *making/tracking confusion*.

To make a moral judgment seems to require the possession of the moral concepts of right and wrong. And it is not unreasonable to suppose that Capuchin monkeys do not

possess these concepts. Underlying this thought is the distinction between concept possession, on the one hand, and the ability to discriminate on the other. If an ant is sprayed with oleic acid, its fellow ants will remove it from the colony – oleic acid is given off when an ant dies. Ants can discriminate, with a reasonable level of precision, which of their fellows are dead from which of them are not. But it would be implausible to suppose that they possess the concept of death. Similarly, it might be argued, animals might be trained to discriminate things that are good from things that are bad without possessing the concept of good and bad. To possess that concept, one would need to know not merely which things are good and bad but in what their goodness or badness consists.

If we accept this, then it seems that (1) if emotions such as indignation involve moral judgments, and (2) moral judgments require moral concepts, then (3) Capuchin monkeys, it would seem, cannot possess emotions of this sort. The account of emotions assumed here is a *cognitivist* one: emotions are seen as requiring (on some implausibly strong versions, reducing to) judgments. One option for the defender of emotions in animals, therefore, is to attack the cognitivist account of emotions. We shall not pursue this strategy, largely because we think cognitivism about emotions is correct. Instead, we shall argue that even if one assumes cognitivism about emotions, and so sees emotions as bound up with evaluative, and sometimes moral, judgments, this is compatible with animals possessing emotions. The key to the argument we shall develop is the difference between *making* moral judgments and *tracking* moral judgments. Making moral judgments is not required for possession of emotions such as indignation. All that is

required is that the emotion, in a sense to be made clear, track moral judgments – judgments that the animal need not be able to make.

Smith is indignant that Jones snubbed him. There are two ways in which this emotion might, let us say, *misfire* – roughly the analogue of what it is for a belief to be false. The category of a misfire is a conjunctive one. An emotion can misfire either because it is *misplaced* or *misguided*. Smith is indignant because he believes Jones snubbed him, but he is, in fact, mistaken. Jones didn't snub him at all. Smith was being his usual hypersensitive self, imagining slights where there are none. Let us say that, in this case, Smith's indignation is *misplaced*. An emotion is misplaced when it depends on a factual assertion's being true when that assertion is, in fact, false. The other source of failure would occur if Jones has every right to snub Smith – say because of Smith's boorish behavior on their most recent encounter. Smith, as we might say, deserved no better from Jones in this case. Let us say that Smith's indignation is, in this case, *misguided*. An emotion such as this is misguided when it depends on a claim of entitlement where there is, in fact, no such entitlement. More generally, an emotion is misguided when (i) it requires the truth of a given evaluative claim, and (ii) this evaluative claim is not, in fact, true.

The Capuchin's indignation – which we have supposed, for the sake of argument, it possesses – can misfire in the same ways. It might be *misplaced*: for example, his fellow Capuchin has not been offered a grape at all, merely a cucumber. In such a case, the Capuchin might be angry that the researcher has nothing better to offer. But he cannot be indignant at the unfair way it is being treated in comparison with his fellow Capuchin.

Or it might be *misguided*: for example, the monkey has not performed the task for which the grape is the reward, and therefore the implicit evaluative judgment that it deserves better treatment is not, in fact, true.

The idea of an emotion being misguided allows us to understand the location in logical space of the evaluative component of the emotion. *If an emotion, E, is not to be misguided, then a certain evaluative proposition, p, must be true*. The truth of this proposition, as we might say, makes sense of the emotion. We need not think of emotions as reducible to evaluations. Rather, for any emotion, there is a certain evaluative proposition that must be true in order for the emotion to not be misguided.

In this sense, possession of an emotion *tracks* a true evaluative proposition. If an emotion is not misguided, then there exists a certain evaluative claim, *p*, and *p* must be true. More precisely, there exists one and only one evaluative clam whose truth is *guaranteed* by the non-misguided status of the emotion: the indexical proposition that 'In being offered a cucumber rather than a grape, I have been wronged'. The non-misguided status of an emotion, therefore, guarantees the truth of a given evaluative proposition. When we have emotions that are not misguided, these emotions track true evaluative propositions. To track a proposition does not require that one entertain, or even be capable of entertaining, it – that one be able to *make* the moral judgment that is tracked. Thus, this *aponoian* account of emotions avoids the charge of over-intellectualization, and explains how they can be spread as widely through the animal domain as both science and common sense take them to be.

The fallacy embodied in the argument that animals cannot have emotions is, therefore, the *making/tracking confusion*. An emotional state that is thought to require that animals make, or be capable of making, a given judgment in fact requires no such thing. It is sufficient for possession of the state that a given judgment or proposition be tracked in the sense that, if the emotion does not misfire (i.e. is not misguided or misplaced or both), there is a certain proposition that must be true (or two propositions – one factual one evaluative).

7. Empathy and Moral Motivation

The combined effects of both premature meta-articulation and the making/tracking confusion are nowhere more evident than in many treatments of empathy and moral motivation.

The Capuchin monkey's indignation is a self-directed emotion in that it concerns his or her own well-being. An important category of other-directed emotion – one that concerns the welfare of another – is *empathy.* Other-directed moral emotions are based on a concern – which can take either positive or negative form – about the well-being of another individual.

The term 'empathy' is notoriously ambiguous. In particular, it is used, as a proposed *explanans*, in two quite different theoretical contexts. Sometimes empathy is understood as a mechanism involved in social cognition. For instance, in some versions of the simulation theory of mind, empathy is understood as an ability that enables the

attribution of mental states to others. By empathizing with someone, we place ourselves in her mental shoes and come to understand what she thinks and feels. In this theoretical context, the motivational or emotive connotations of the concept of empathy are usually bracketed.[22] We shall not discuss this sense of empathy.

Our concern, rather, will be with the concept of empathy as it is employed in moral contexts, specifically in contexts of moral motivation. Here, the concept is strikingly variegated: at one end of the spectrum, empathy requires breathtakingly complicated cognitive and conceptual abilities; at the other it is little more than a brute physiological reaction. Leslie Jamison, in her wonderful book, *The Empathy Exams*, veers about as far as one can in the direction of complexity:

> Empathy isn't just listening, it's asking the questions whose answers need to be listened to. Empathy requires inquiry as much as imagination. Empathy requires knowing that you know nothing. Empathy means acknowledging a horizon of context that extends perpetually beyond what you can see … Empathy means realizing no trauma has discrete edges. Trauma bleeds. Out of wounds and across boundaries ... you enter another person's pain as you'd enter another country, through immigration and customs, border crossing by way of query: *What grows where you are? What are the laws? What animals graze there?*[23]

Empathy, in this sense, requires mind-reading – the ability to attribute mental states both to others and to oneself –, which is a form of reflexive thought, and much more besides

[22] See, for example, Goldman (2006), p. 17: "mindreading is an extended form of empathy (where this term's emotive and caring connotation is bracketed)".
[23] Jamison (2014), p. 5.

(perspective taking, imaginative reconstruction, and so on). We shall assume, with more than a little confidence, that animals are incapable of empathy in Jamison's sense. Not all cases of empathy need be this complex, of course. Nevertheless, one can detect a pronounced tendency to suppose that empathy involves mind-reading abilities: the ability to understand the minds of another by way of the attribution of mental states to them. Thus, de Vignemont and Jacob (2012) write:

> The motivational role of empathetic pain for moral and prosocial behavior ... has often been stressed. ... In order to react appropriately to another's pain, one needs to understand the fact (or to believe) that she is in pain. Hence, prosocial behavior requires third-person mind reading.[24]

A similar idea can be found in Batson (2009):

> Feeling for another person who is suffering ... is the form of empathy most often invoked to explain what leads one person to respond with sensitive care to the suffering of another. ... To feel for another, one must think one knows the other's internal state ... because feeling *for* is based on a perception of the other's welfare[25]

The idea here is that, without the ability to attribute mental states to others, an empathically motivated helping reaction simply *cannot occur*. If a creature cannot understand the pain and suffering of the individual whose misfortune she is witnessing,

[24] de Vignemont & Jacob (2012), p. 310.
[25] Batson (2009), pp. 9-10.

then she has no *reason* to help her. These are all examples of fairly complex models of empathy.[26]

Consider, now, the other end of the spectrum. The most basic form of empathy, so basic that one might legitimately question whether it is indeed empathy, is emotional contagion: an involuntary affective resonance that occurs in presence of another individual who is undergoing a certain emotion. It requires no understanding of what initiated the reaction and yields a form of personal distress that is either non-intentional or directed at one's own well-being. There is no reason to suppose that this sort of reaction has a moral character, for neither the emotional reaction nor the behavior triggered by it are directed towards the other individual, so it is not an expression of other-directed concern.

If these two sorts of cases were exhaustive, the prospects for empathy occurring in animals, as a specifically moral motivational state, would be bleak. The more complex forms of empathy might be moral, but animals cannot possess them. The simpler forms of empathy animals can possess but have no moral import. In accordance with our *aponoian* paradigm, we shall argue for the existence of intermediate forms of empathy that are both moral and can plausibly be thought to be possessed by at least some animals. We shall designate the least cognitively demanding of these forms as *minimal moral empathy*.

What makes a version of empathy complex – non-minimal in our sense – is that it essentially involves judgment. In order to be empathically motivated, in this complex

[26] In our view, the assumption that empathy always *requires* mind-reading abilities is an example of premature meta-articulation. We do not deny, of course, that mind-reading abilities are implicated in some cases of empathy.

sense, a creature must have the ability to make certain pertinent moral judgments. This is a common assumption. For example, as Dixon (2008) puts it:

> [W]hat matters to empathy understood as a moral concept is that the subject perceives what is morally salient about another's situation. Even this additional requirement doesn't quite capture what needs to be added to cognitive empathy to make it a morally significant concept, and that is its relation to sympathy or compassion. These states are genuine moral emotions in the case where they motivate a subject to help or to alleviate need, distress, or suffering when this is *judged* to be "serious" and undeserved. The *moral significance of the emotional states of sympathy and compassion is explained by the presence of evaluative judgments* as well as the motivations to act on these evaluations or appraisals.[27]

Empathy, therefore, cannot be a moral emotion unless it is accompanied by explicit evaluative judgments by means of which one reflects upon the morally salient features of a situation (judgments of seriousness, deservedness, and so on). Hauser (2001) seems to have something similar in mind:

> At present, we have no convincing evidence that animals attribute beliefs and desires to others. ... Similarly, we also lack evidence that animals have access to their own beliefs, reflect on them, and contemplate how particular events in the future might change what they believe. If this lack of evidence correctly reveals a lack of capacity, then animals can certainly cooperate, beat each

[27] Dixon (2008). p. 140. Emphasis is ours.

> other to a pulp, and make up after a war. But they can't evaluate whether an act of reciprocation is *fair*, whether killing someone is *wrong*, and whether an act of kindness should be rewarded because it was the *right* thing to do.[28]

A creature's empathic reaction to the plight of another would not be a moral emotion unless the creature is able to view this reaction in, as Hauser puts it, "a context of right and wrong". But to view it in such a context seems to require that the animal be able to locate its reaction in a network of judgments about right and wrong.[29]

If his is the case against animals possessing empathy as a moral motivation, then we have already outlined an apparatus that can be used to undermine this case: the distinction between *making* a judgment and *tracking* a judgment. To see how this works in the case of empathy, consider the (notoriously immoral) experiment performed by Masserman *et al*. (1964). Monkeys would refuse to pull a chain that delivered food to them when they found out that, by pulling said chain, a conspecific situated within their sight received an electric shock. One of these monkeys – let us call him M – famously refrained from eating for twelve days in a row.

Supposed M's refusal to pull the chain was the result of a feeling of distress. This feeling is not simply *caused* by the suffering of his fellow; it is also *intentionally directed* towards that suffering. That is, M is distressed *that* the other is suffering. Even though this distress motivates M to take steps to mitigate his fellow's suffering, this would not, according to Hauser and Dixon, qualify as a moral motivation. To do so, M would need

[28] Hauser (2001), p. 312. Emphasis is his.
[29] In this passage, we can also scent, in the requirement of being able to attribute beliefs to oneself and others, the pungent aroma of premature meta-articulation.

to be able to make judgments concerning the moral status of his motivations and/or resulting behavior. This idea, however, is grounded in the making/tracking confusion.

M's empathic motivation, like emotions in general, can be misplaced or misguided. It is *misplaced* if it is based on a factual assertion that is not true: let us suppose its fellow monkey is not suffering. M is mistaken. It is *misguided* if it is based on a morally evaluative judgment that is false. This could arguably be the case if his fellow monkey, due to some extraordinarily improbable circumstances, somehow deserved to suffer, or if the electric shock was being delivered to him for his own good. Thus, if we assume that M's emotional response is *not* misguided, then there exists a moral proposition that *must* be true: namely, the proposition that the monkey's suffering is wrong or bad. This does not require that M be able to *make* this judgment. All that is required is that it possesses an empathic motivational state that *tracks* this judgment, in the sense that the non-misguided status of M's motivation guarantees the truth of the relevant moral proposition. Even if M's emotional reaction is a form of contagion triggered by watching his conspecific suffer, this does not preclude its status as a moral motivation. This idea of a truth-preserving or truth-guaranteeing relation between the emotion and a given moral judgment lies at the heart of the concept of minimal moral empathy.

The key to understanding minimal moral empathy lies in the idea of a *reliable emotional response* to *morally relevant features* of a situation. Let us suppose, a supposition that seems entirely reasonable, that the suffering of M's fellow monkey is a bad thing. This suffering is, therefore, a morally relevant feature of the situation: it is

what we might call a *bad-making* feature of this local situation. M's response to this morally relevant, bad-making, feature is an emotional one: it takes the form of distress, built into which is an urge to mitigate the situation. Thus, M refuses to pull the chain. Let us make one further assumption: M's response is not a random or arbitrary one. Rather, he responds to situations such as this in a reasonably reliable way. When monkeys are tortured with electric shocks, M reliably feels distress and an urge to mitigate the suffering. This, we might suppose, is the result of some mechanism that reliably produces emotional responses to at least some morally salient features of situations. In these circumstances, all that is required for M's emotional response to be a moral one is this: if M's emotional response is not misguided, then the moral claim, "The suffering of this monkey is bad" must be true. M does not need to be able to make this moral judgment, or entertain this moral proposition. To suppose that he does is to fall victim to the making/tracking confusion. It is enough that M's emotional response tracks the moral proposition in the sense explained above. If it does, M's response is an example of minimal moral empathy.

We chose the example of Masserman's monkeys for a reason. It is commonly thought that their behavior is open to another interpretation. As Hauser (2001) puts it:

> What is most remarkable about these experiments is the observation that some rhesus monkeys refrained from eating in order to avoid injuring another individual. Perhaps the actors empathized, feeling what it would be like to be shocked, what it would be like to be the other monkey in pain. Alternatively, perhaps seeing someone shocked is unpleasant, and rhesus will do whatever they

> can to avoid unpleasant conditions. Although this has the superficial appearance of an empathic or sympathetic response, it may actually be selfish.[30]

This claim is, if our arguments are correct, based on a false opposition. First, minimal moral empathy does not require the ability to imaginatively feel "what it would be like to be shocked." Some cases of empathy are undoubtedly like this, but minimal moral empathy requires no such ability. Second, the supposition that, if M is motivated by the unpleasant nature of his or her emotion, this automatically disqualifies it from being moral is a supposition that is also unwarranted. M no doubt would find the shrieks of its fellow distressing. But this experiential unpleasantness may be precisely the form M's concern for the other monkey takes. Compare: one would find the shrieks of distress of one's children distressing, and would take immediate steps to try and stop them. Does this mean one is merely engaging in a selfish attempt to stop this unpleasant noise? This claim would be ridiculous. Of course, one finds the shrieks of distress of one's children unpleasant. This is precisely the experiential form one's concern for them takes.

The debate over whether the motivation of M is moral has, thus, been based on a false dichotomy: between a moral motivation and an aversive stimulus. The assumption has, typically, been that if an emotion is the result of vicarious aversive arousal, this precludes its qualifying as a moral emotion.[31] If the idea of minimal moral empathy is correct, this assumption is unwarranted. The actual motivation of the monkey is an

[30] Hauser (2001), p. 276.

[31] See also, for example, de Waal (2008), p. 283: "Perhaps the most compelling evidence for emotional contagion came from Wechkin et al. (1964) and Masserman et al. (1964), who found that monkeys refuse to pull a chain that delivers food to them if doing so delivers an electric shock to and triggers pain reactions in a companion. Whether their sacrifice reflects concern for the other ... remains unclear, however, as it might also be explained as avoidance of aversive vicarious arousal."

empirical matter, on which we do not take a stand here. Our point is this: even if its emotional response is a case of vicarious aversive arousal, this is perfectly compatible with its being a moral emotion – a case of what we call minimal moral empathy.

In addition to the making/tracking confusion, the issue of moral motivation in animals is also clouded by an unfortunate tendency toward premature meta-articulation. This emerges, in particular, in connection with the common idea that animals cannot act morally because they lack *control* over their motivations and actions. Here is a way of thinking about motivation in general, and moral motivation in particular, a way made admirably clear by Christine Korsgaard. In her response to de Waal's Tanner Lectures that were published as *Primates and Philosophers*, Korsgaard comments on the status of the lower animals – a spider crawling towards a moth that is caught in the middle of her web:

> Here we begin to be tempted to use the to use the language of action, and it is clear enough why: when an animal's movements are guided by her perceptions, they are *under the control* of her mind, and when they are *under the control* of her mind, we are tempted to say they are under the animal's own control. And this, after all, is what makes the difference between an action and a mere movement – that an action can be attributed to the agent, that it is done under the agent's own control.[32]

As the animal in question becomes more complex, the degree of control it is capable of exerting over its movements becomes correspondingly greater:

[32] Korsgaard (2006), p. 108. Emphasis is ours.

> Even if there is a gradual continuum, it seems right to say that an animal that can entertain his purposes before his mind, and perhaps even entertain thoughts about how to achieve those purposes, *is exerting a greater degree of conscious control* over his movements than, say, the spider, and is therefore in a deeper sense an agent.[33]

This is the first appeal to the meta-level: an animal that can think about its purposes, perhaps even its thoughts, is more in control of its movements than an animal that cannot do this. With humans, however, Korsgaard believes there is a qualitative leap. The reason is that we can choose our ends, and not merely choose how to achieve ends antecedently given to us by our nature and the demands of our environment: "For we exert a deeper level of control over [our] own movements when we choose our ends as well as the means to them than that exhibited by an animal that pursues ends that are given to her by her affective states."[34] As Korsgaard notes, this ability to choose ends – to assess and adopt them rather than merely have them – is what Kant called "autonomy". And it is only when we have autonomy, Korsgaard claims, that specifically moral agency emerges.[35] The reason this is a qualitative leap, Korsgaard claims, is because it requires a specific form of self-consciousness that only humans, in fact, have.

> What I mean is this: a nonhuman agent may be conscious of the object of his fear or desire, and conscious of it as *fearful* or *desirable*, and so as something to be avoided or sought. That is the ground of his action. But a rational animal is, in addition, conscious *that* she fears or desires the object, and *that* she is inclined to

[33] Korsgaard (2006), p. 109. Emphasis is ours.
[34] Korsgaard (2006), p. 112.
[35] Korsgaard (2006), p. 112.

> act in a certain way as a result … Once you are aware that you are being moved in a certain way, you have a certain reflective distance from the motive, and you are in a position to ask yourself, ”but should I be moved in that way?” Wanting that end inclines me to do that act, but it does it really give me a reason to do that act? You are now in a position to raise a normative question about what you *ought* to do.[36]

The centrality of the concept of control is very evident in Korsgaard's argument. The greater the degree of control an individual has over his actions, the greater the warrant there is for regarding that individual as an agent. In the case of humans, we have a type of control over our motivations ("ends") that no other creature has: we can choose our ends. This is grounded in a uniquely human form of self-consciousness that provides us with "reflective distance" between us and our motives, thereby allowing us to scrutinize those motives and ask ourselves whether they are ones we should endorse or reject. This is what makes moral action possible. In short, we have a form of control over our motives that no other creature has; and it is this control that allows us to act morally.

It is easy to feel the intuitive pull of this idea. It is tempting to suppose that in the absence of the relevant meta-cognitive abilities – the ability to form higher-order thoughts about our motivations and purposes – we are at their "mercy". They push us this way and that. Unable to critically scrutinize these motivations, we have no control over what they cause us to do. Meta-cognitive abilities, however, would transform us. Armed with these abilities, we can sit above the motivational fray: observing, judging and evaluating our motivations, coolly deciding the extent to which we will allow them to

[36] Korsgaard (2006), p. 113. Emphasis is hers.

determine our decisions and actions. This gives us a control over our motivations that we would otherwise lack.

Nevertheless, despite its intuitive appeal, it is doubtful that this picture of control can work. Indeed, the problems here precisely parallel those faced by the HOT model of consciousness. Specifically, there is a recalcitrant property – the property of being under the control of the agent – that first-order states (motivations, purposes, etc.) lack. We introduce higher order states – thoughts about those motivations and purposes – to supply this control. But then the same issue of control will, logically, arise at this higher-order too. Meta-cognition was supposed to allow us to sit above the motivational fray, and calmly pass judgment on our motivations, thus providing us with control over them. However, there is no reason to suppose that meta-cognition is above this motivational fray. If first-order motivations can pull us this us this way and that, then second-order evaluations of those motivations can do exactly the same.

The appeal to meta-cognition to imbue us with control over our motivations, thus, faces a dilemma: essentially the same dilemma we discovered with the HOT model of consciousness. Do we have control over our meta-cognitive assessments of our purposes and motivations? If so, then we have not explained the notion of control, but simply assumed it. But if not, then it is difficult to see how these meta-cognitive assessments could supply us with control over our motivations and purposes. The appeal to the meta-level as a way of explaining control over motivations, therefore, gets us nowhere. It is another example of premature meta-articulation (specifically, in its *miracle-of-the-meta* form).

This is a well-known problem with the idea that we can explain autonomy by the appeal to meta-level phenomena. It is common to respond to this problem through the addition of further factors concerning the conditions under which this meta-cognition takes place. For example, a common response is to insist that the reflection must take place under conditions free of distorting factors, or must reflect an adequate causal history, and so on. These are perfectly reasonable ways of trying to safeguard the idea of autonomy. But, if they work, it is only by divorcing the concept of autonomy from that of control. Whether or not one's meta-cognizing takes place under conditions free of distorting factors, or reflects an adequate causal history, is not something that is under the agent's control – indeed these are things that may remain unknown, perhaps even unknowable, to the agent.

The appeal to the meta-level in order to explain control is not only fruitless, it is also, in the eyes of many, unnecessary. We can explain autonomy without venturing outside the first-order. On the contrary, all that is required is, postulation of a choice mechanism that translates beliefs about our alternatives, coupled with our desires, into plans of action that are designed to realize those desires. This alternative strategy is a, broadly, compatibilist one. For the compatibilist, being produced in this way by the appropriate choice mechanism – a mechanism that is 'responsive to reasons' – is precisely what it is to be an autonomous subject. There is, as yet, no reason to suppose that animals cannot possess such mechanisms.

Discussions of empathy and moral motivation, thus, habitually fall victim to both premature meta-articulation and the making/tracking confusion. To understand the extent

to which animals are capable of empathic and even moral behavior, these confusions must be expunged from the debate.

8. Conclusion

There is a pronounced tendency in empirical studies and theoretical treatments of animals to underestimate their abilities not because some distinctive lack or *aporia* has been discovered in them, but because of implausibly intellectualist accounts of the abilities they are supposed to lack. If animals are thought to lack consciousness, for example, this will stem not from anything we have discovered about animals themselves, but from an implausibly (over) intellectualist account of what consciousness is. The same is true of their purported lack of beliefs, emotions, empathy and the ability to act morally. In all these cases, the alleged deficiency of animals derives not from our discovery of some deficiency in them but, rather, from an unreasonably over-intellectualized conception of these phenomena. At the heart of this intellectualization, we have identified two errors: the tendency toward premature meta-articulation and the making/tracking confusion. As an antidote towards this sort over-intellectualization, and resulting underestimation of animals, we recommend adoption of the *aponoian* paradigm. Intellect is rarely as intellectual as we think it is.

Acknowledgments

This research was partially funded by an FPI scholarship awarded to Susana Monsó by the Spanish Ministry of Economy and Competitiveness (research project FFI2011-23267).

REFERENCES

Batson, C. D. (2009). 'These things called empathy: Eight related but distinct phenomena'. In J. Decety & W. Ickes (Eds.), *The Social Neuroscience of Empathy* (pp. 3–15). (Cambridge, MA, US: MIT Press.)

Carruthers, P. (1989) 'Brute experience', *The Journal of Philosophy* 86, 5, 258-69.

Carruthers, P. (2008) 'Meta-cognition in animals: a sceptical look', *Mind and Language* 23, 1, 58-89.

Cottingham, J. (1978) 'A brute to the brutes? Descartes treatment of animals', *Philosophy* 53, 206, 551-59.

Davidson, D. (1975) 'Thought and talk' in *Mind and Language,* S. Guttenplan ed., (Oxford: Oxford University Press)

Davidson, D. (1985) ‘Rational animals’ in *Actions and Events: perspectoves on the Philosophy of Donald Davidson*, E. LePore and B. McLaughlin eds., (Oxford: Blackwell).

Descartes, R. (1927) *Descartes Selections* ed., R. Eaton (New York: Scribner & Sons)

De Vignemont, F., & Jacob, P. (2012). ‘What is it like to feel another’s pain?’ *Philosophy of Science*, 79(2), 295–316. doi:10.1086/664742

De Waal, F. B. M. (2008). ‘Putting the altruism back into altruism: The evolution of empathy.’ *Annual Review of Psychology*, 59, 279–300. doi:10.1146/annurev.psych.59.103006.093625

Dixon, B. A. (2008) *Animals, Emotion & Morality: Marking the Boundary* (Prometheus Books)

Goldman, A. I. (2006) *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading* (1st ed.) (Oxford University Press, USA)

Hauser, M. (2001) *Wild Minds: What Animals Really Think* (Holt Paperbacks)

Jamison, L. (2014) *The Empathy Exams: Essays* (Minneapolis, MN: Graywolf Press)

Korsgaard, C. M. (2006) ‘Morality and the distinctiveness of human action’. In S. Macedo & J. Ober (Eds.), *Primates and Philosophers: How Morality Evolved* (Princeton University Press)

Masserman, J., Wechkin, S., & Terris, W. (1964) ‘“Altruistic” behaviour in rhesus monkeys’, *American Journal of Psychiatry*, 121(6), 584–585.

Malcolm, N. (1973) ‘Thoughtless Brutes’, *Proceedings and Addresses of the American Philosophical Association*, 46(September), 5–20.

Panksepp, J. (1998) *Affective Neuroscience: The Foundations of Human and Animal Emotions* (Oxford: Oxford University Press)

Penn, D. C., & Povinelli, D. J. (2007) 'On the lack of evidence that non-human animals possess anything remotely resembling a "theory of mind"', *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 362(1480), 731–744. doi:10.1098/rstb.2006.2023

Penn, D. C., & Povinelli, D. J. (2013) 'The comparative delusion: The "behavioristic"/ "mentalistic" dichotomy in comparative theory of mind research'. In H. A. Terrace & J. Metcalfe (Eds.), *Agency and Joint Attention* (Oxford University Press, USA)

Povinelli, D. J., & Vonk, J. (2003) 'Chimpanzee minds: Suspiciously human?', *Trends in Cognitive Sciences*, 7(4), 157–160.

Povinelli, D. J., & Vonk, J. (2004) 'We don't need a microscope to explore the chimpanzee's mind', *Mind and Language*, 19(1), 1–28.

Rosenthal, D. (2004) 'Varieties of higher-order theory' in R. Gennaro ed., *Higher-Order Theories of Consciousness* (Amsterdam: John Benjamins)

Rowlands, M. (2001a) 'Consciousness and higher-order thoughts', *Mind and Language* 16, 3, 290-310.

Rowlands, M. (2001b) *The Nature of Consciousness* (Cambridge: Cambridge University Press)

Rowlands, M. (2012) *Can Animals Be Moral?* (New York: Oxford University Press)

Stich, S. (1979) 'Do animals have beliefs', *Australasian Journal of Philosophy* 57, (1978), 15-28.

Sztybel, S. ‘Did Descartes believe that non-human animals cannot feel pain?’ http://sztybel.tripod.com/animal_feelings.html (accessed August 26th 2014).

Wechkin, S., Masserman, J., & Terris, W. (1964) ‘Shock to a conspecific as an aversive stimulus’, *Psychonomic Science*, 1, 17–18.