Forthcoming in *Ergo*. This a pre-copy-edited version.

Erik Nelson

# Kantian Animal Moral Psychology: Empirical Markers for Animal Morality

Are animals[1] capable of acting morally? Answers to this question, as Mark Rowlands (2012: xi) notes, have been almost universally negative. Most philosophers and scientists think that animals lack the reflective capacities that are necessary for moral thought. Despite this blanket denial, sentimentalist philosophers and scientists, such as David Hume (1739), Charles Darwin (1871), and Frans de Waal (2014), have been willing to attribute at least some of the basic building blocks of morality to animals. De Waal (2006: 54), for instance, argues that some animals can be described as participating in proto-moral practices because they have moral emotions, such as empathy. There is also a long lineage of experimental work on animal empathy, altruism, and other morally linked states and behaviours (for an overview, see Monsó and Andrews, 2022). Building off this earlier work, Rowlands (2012: 33–36) has argued that animals *can* be moral subjects if they have moral emotions that can function as reasons for their behaviour.

My aim in this paper is not to critique these sentimentalist approaches. Instead, I will show that sentimentalism is not the only moral tradition that can treat animal morality as an open question with a potentially affirmative answer. I argue that a Kantian inspired investigation into animal morality is both a plausible and coherent research program. This will likely surprise contemporary Kantian philosophers, since many of them have argued that animals are not capable of cognition (Brandom, 1994), inner and outer experience (McDowell, 1996), judgment (Rosenberg, 1997), or conceptual capabilities (Sellars and Chisholm, 1958). Most directly,

[1] For this paper, I will be using the term 'animals' to refer to nonhuman animals.

Christine Korsgaard (2004) has argued that animals are not capable of acting morally because they lack the necessary rational capabilities. My aim in this paper is not to argue that animals *do* act morally, nor to argue that Kant should have given an affirmative answer. Instead, I argue that the philosophical commitments of Kant and his contemporary interpreters should lead them to regard the question as an open one, and that Kant's empirical moral psychology can help shape what an affirmative answer would look like.

To show that a Kantian inspired investigation into animal morality is possible, I argue that philosophers, such as Korsgaard, who use a two-perspectives interpretation to claim that reason demarcates animals from the domain of moral beings are equivocating in their use of the term 'rationality'. Kant certainly regards rationality as necessary for moral responsibility from a practical standpoint, but his distinction between the noumenal and phenomenal means that he can only establish it as a marker for morality from a theoretical standpoint. A theoretical standpoint is a standpoint from which one views the world as made up of objects that causally interact; for example, how one views the world when they are doing scientific work. In addition, I argue that while a two-worlds interpretation of Kant can save the argument from this equivocation, its inability to say anything specific about noumenal objects will still undermine the validity of the argument. This means that rationality from a theoretical perspective can neither be necessary nor sufficient for morality, leaving open the possibility for other empirical markers for moral responsibility. I argue that the higher faculties, character, implicit knowledge of universality, and antecedent practical pleasures (which provide a way to distinguish between morally motivated behaviour and other types of socially motivated behaviour) can all serve as empirical markers for morality. There is empirical evidence that at least some animals have conceptual capabilities and therefore the empirical marker of the higher faculties as well as suggestive evidence that merits

further investigation for the other markers. While this will not provide a definitive answer on whether animals are capable of acting morally, it will provide a Kantian outlook that can be used to evaluate empirical and philosophical work on animal morality.

## 1. Korsgaard's Argument Against Animal Morality

In a series of influential papers (2004, 2006, 2010, 2011, 2018b) and her recent book (2018a), Korsgaard has critiqued recent sentimentalist attempts to attribute moral or proto-moral capabilities to animals. Her argument rests on the claim that the rational capacities necessary for morality are intellectually beyond the capabilities of animals. Inspired by a Kantian conception of both rationality and morality, Korsgaard argues that rationality is a metacognitive process that allows humans to evaluate their reasons for acting (Korsgaard, 2018: 299). In a representative passage, she writes:

> ...rationality, for Kant, is the capacity for normative self-government. Rationality makes us capable of assessing and judging the principles that govern our beliefs and actions, and of regulating our beliefs and actions in accordance with those judgments. Rationality also makes it necessary for us to exercise this capacity, for as long as we are conscious of our principles, to some extent we cannot help but assess them. Once they are before our minds, we must decide whether to endorse or reject them, and act accordingly. According to Kant, the fact that human beings live under this kind of normative self-government is the distinctive difference between human beings and the other animals. And it is clear from this account why Kant thinks that we are the only moral animals, in the sense that we are the only animals whose conduct is subject to moral guidance and moral evaluation (2004: 87).

While it is clear that Korsgaard's Kant views rationality as necessary for morality, this passage also demonstrates that she sees him as claiming that it is also sufficient. Rational beings "cannot help but assess" the principles regulating their beliefs and actions, and therefore morality is unavoidable for them. Animals lack these rational capabilities and therefore cannot be moral. Korsgaard writes that the attitudes of an animal are "invisible to her, because they are a lens *through* which she sees the world, rather than being parts of the world that she sees" (Korsgaard,

2011: 102). Lacking any reflective distance from their maxims, animals cannot evaluate their moral worth and therefore cannot make moral decisions.

Korsgaard's Kantian argument against animal morality is then:

1. Rationality is necessary and sufficient for morality.
2. Animals are not rational.
3. Therefore, animals are not moral.

I will call this argument the Kantian argument against animal morality (KAAAM) because while it is extracted from Korsgaard's interpretation of Kant, I suspect that if asked, many Kantians would agree to some version of this argument. The argument may seem straightforwardly valid; however, if it is meant to rest on Korsgaard's interpretation of Kant, the term 'rationality' cannot have the same meaning in both premises. In addition, I will argue further down that the argument cannot be rescued (at least as a deductively valid argument) by a non-Korsgaardian two-worlds interpretation of Kant. To see why, we need to turn to the question of how to tell if another being is capable of morality.

From Korsgaard's exegesis, it may seem obvious that determining whether another being is capable of morality is a matter of identifying which beings are rational. However, for Kant this is no easy task. Kant presupposes that his readers are rational and tries to convince them that they should regard and treat other rational beings morally, but never explicitly provides a formula for sorting rational beings from nonrational beings. As Korsgaard notes, for Kant, "moral thought is seen as arising from the perspective of the agent who is deciding what to do. Responsibility is in the first instance something taken rather than something assigned" (1996a: 189). This is helpful if one is interested in the transcendental grounds of one's own morality or what those grounds prescribe but is not particularly helpful when it comes to questions about the moral capabilities

of others. However, even from an agential perspective, Kant cannot altogether avoid questions about the minds of others, given some of his most central principles. For example, the second formulation of the categorical imperative tells us to always treat others as ends in themselves (1785: 4:429). For Kant those 'others' are rational beings because only rational beings can set their own ends. So, following the second formulation (at least in a world with both rational and nonrational beings) then seems to require some way to identify which beings are rational.

Korsgaard argues that, for Kant, identifying someone as morally responsible for their actions is a practical, not a theoretical matter (1996a: 197–198). A theoretical attribution of moral responsibility is when "it is a fact about a person that she is responsible for a particular action, or that there is some fact about her condition either at the time of action or during the events which led up to it which fully determines whether it is correct to hold her responsible" (1996a: 197). In contrast, recognizing others as morally responsible from a practical perspective means adopting an attitude or holding a postulate about a morally relevant other. A postulate is a theoretical claim, such as 'X is morally responsible', that is accepted a priori for practical reasons. For example, the very possibility of moral action requires that I recognize that some others are capable of moral responsibility.

Part of the reason for identifying morally responsible others practically, instead of theoretically, is the connection between freedom and moral responsibility. Korsgaard and Kant think that it is only through rationality that one can be moral because it is only through rationality that one can freely adopt a principle of action. In other words, one can only be morally responsible for one's actions if one has freely chosen their maxims. If someone is not acting upon freely chosen maxims and is being influenced by an outside force, then whatever is influencing one from the outside would be responsible for their actions instead of oneself

(Korsgaard, 1996b: 162–163). So, moral responsibility requires freedom, and it is only through rationality that one can be free. This quickly leads to the familiar problem of finding a place for freedom in a deterministic world. To do this, Kant argues that humans are members of two distinct worlds (Korsgaard, 1996a: 201). One of these worlds, which Kant calls the phenomenal world – "the world of things as they appear to us" – is completely causally determined (Korsgaard, 1996a: 201). The other is the noumenal world, which is "the world of things as they are in themselves" (Korsgaard, 1996a: 201). Kant argues that the possibility of freedom in the noumenal world does not conflict with the deterministic nature of the phenomenal one. Since we exist in both worlds, it is possible for one's behaviour to be fully determined but also free.

Korsgaard argues for what is known as the two-perspectives interpretation of the noumenal/phenomenal distinction. This approach takes the distinction between the noumenal and the phenomenal to be a perspectival distinction instead of a metaphysical one.[2] Korsgaard writes that:

> As thinkers and choosers we must regard ourselves as active beings, and so we place ourselves among the noumena, necessarily, whenever we think and act. According to this interpretation, the laws of the phenomenal world are laws that describe and explain our behaviour. But the laws of the noumenal world are laws which are *addressed to us* as active beings; their business is not to describe and explain at all, but to govern what we do. Reason has two employments, theoretical and practical. We view ourselves as phenomena when we take on the theoretical task of describing and explaining our behaviour; we view ourselves as noumena when our practical task is one of deciding what to do. The two standpoints cannot be mixed because these two enterprises – explanation and decision – are mutually exclusive (1996a: 204).

Moral responsibility is something that beings recognize themselves as having when they view themselves from a practical perspective. They view themselves as having choices which is incomprehensible unless they also regard themselves as being free to make choices. However,

[2] I will consider the metaphysical version, known as the two-worlds interpretation, and its potential ramifications for KAAAM further down.

this creates a problem for Korsgaard's argument against animal morality, since it is only when regarding oneself from this first-person perspective that one can recognize moral responsibility.

A theoretical perspective cannot establish moral responsibility because a theoretical perspective on oneself takes oneself as entirely determined. From this perspective, there are no choices to make because every decision can be fully explained through causally determined events. So, no theoretical fact can fully establish that another being is morally responsible because moral responsibility requires that the being is freely making choices. Therefore, moral responsibility is something that can only be recognized from a practical viewpoint. If establishing which beings are morally responsible relied on identifying some phenomenal trait, capability, or psychological process as necessary and sufficient, the distance between the theoretical and practical standpoints would doom any ascription of moral responsibility to failure. While taking up an attitude or postulate might seem to mitigate this concern, we are still lacking guidance for whom those attitudes and postulates should be directed towards.

So given these difficulties in identifying morality, rationality, and freedom in others for Korsgaard and Kant, it is far from clear how Korsgaard can argue that animals are incapable of morality. Again, Korsgaard's argument is:

1. Rationality is necessary and sufficient for morality.
2. Animals are not rational.
3. Therefore, animals are not moral.

There are now two ways to interpret the term 'rationality' in her argument: rationality is either phenomenal or noumenal.

If she is using 'rationality' to refer to a psychological process that can be identified from a theoretical standpoint, then Kant would regard Premise 1 as false because the phenomenal

process of rationality is not sufficient for morality. Rationality, as a psychological process viewed from a theoretical perspective, is causally determined. Therefore, it can be used to provide explanations for the behaviour of oneself and others, but it cannot provide the freedom necessary for moral decision making. If Korsgaard is using the term 'rationality' to mean noumenal reason, then Kant would certainly regard the first premise as true. However, viewed from a practical perspective, it is not clear how one can know that the second premise is true. For other humans, Korsgaard (1996a) claimed that there is no theoretical fact of the matter that can be used to identify the rationality of others, and she suggested that ascribing rationality was a matter of taking up a certain attitude towards someone. The first-person practical perspective through which we identify our own rationality is not available when it comes to others, so we must take it on as a postulate because there are no phenomenal facts about them that could fully justify one's belief. However, if this is the case, it gives one no basis for arguing that certain beings are not rational. So, it is not clear what grounds Korsgaard has for asserting Premise 2, if 'rationality' is interpreted noumenally.

However, one could take the first premise to be established practically and the second premise to be established theoretically. While Korsgaard sometimes makes it sound like theoretical considerations cannot factor into a practical viewpoint (e.g. 1996a: 204), there is nothing in the two-perspectives interpretation that inherently rules out this possibility. Theoretical considerations can and must play a role in at least some practical deliberations (such as when considering the consequences of a particular action). As Onora O'Neill, a proponent of the two-perspectives interpretation, writes: "the actions that agents perform assume a causally ordered and knowable world that provides an arena for action" (1989: 68). Interpreting passages from Korsgaard and O'Neill, Patrick R. Frierson argues that the practical standpoint does not

mean that "one sees the world as free of causal influence" but that "one sees its causal relations as tracing back to one's own, undetermined choices" (2010: 86). So, Korsgaard's two-perspectives interpretation of Kant does not rule out taking the second premise as theoretical if the first premise is established from a practical perspective. The problem for taking the second premise to be a theoretical one is that the argument is now equivocating in its use of the term 'rationality.' In the first premise, 'rationality' is used to refer to the transcendental grounds necessary for morality and in the second 'rationality' refers to a psychological capability. Phenomenal facts, such as the existence or lack thereof of a causally determined psychological capability, cannot rule out noumenal possibilities, such as the freedom necessary for morality. Otherwise, morality would be impossible. Therefore, KAAAM is invalid.

Turning from Korsgaard's two-perspective interpretation to a two-worlds interpretation may seem to provide a possible avenue for a valid interpretation of KAAAM.[3] A two-worlds interpretation of the noumenal/phenomenal distinction takes the distinction to be a metaphysical one. Tobias Rosefeldt (2022: 23) has argued that noumenal objects play a role in explaining even our perceptual experiences because the idea of an appearance only makes sense if there is "something beyond our representations of it." If this claim is right, then it could provide a way to refer to noumenal objects in theoretical statements. When it comes to questions about rationality and minds, Rosefeldt (2006) argues that the "I" in Kant's interpretation of the "I think" picks out a kind of 'nonreal' object. Expounding on this approach, James Hutton (2020: 995) argues that "the thinking being is picked out procedurally, in a topic-neutral fashion, through the limits of possible self-ascription in first-person thought."

[3] I would like to thank an anonymous reviewer for stressing the importance of and suggesting ways to consider a two-worlds interpretation of KAAAM.

Kant holds that ascribing minds to other humans always involves judging "that it is possible for there to be a first-person thought in which that mental representation is self-ascribed" (Hutton, 2020: 995). As Kant puts it:

> Now I cannot have the least representation of a thinking being through an external experience, but only through self-consciousness. Thus such objects are nothing further than the transference of this consciousness of mine to others, which can be represented as thinking beings only in this way (1781: B405/A347).

However, when it comes to ascribing minds to other animals, Kant argues that ascriptions of mental content come from analogical reasoning about the states that lie behind observed behaviours. For example, the similarities between the dam-building behaviours of beavers and the building activities of craftsmen allow us to attribute causally analogous mental states (Kant, 2000: 5:464n). Hutton writes that "the criterion for a mental representation's belonging to an animal mind is simply that it plays a certain causal role within the animal's life" (2020: 995).

It seems like this approach could potentially provide a positive valid argument for the moral capabilities of other humans where Korsgaard's two-perspective approach could only provide a postulate. The argument could go something like:

A. Rationality is necessary and sufficient for morality.

B. Other humans are rational.

C. Therefore, other humans are moral.

One could potentially regard this argument as valid even if the second premise is theoretical because rationality picks out a kind of noumenal object in the same way that, according to Rosefeldt, our perceptual experiences of phenomenal objects are undergirded by noumenal ones. The problem comes when we ask: how do we know the truth of premise B? As others have pointed out, Kant wrote surprisingly little on this subject (Walker, 2017: 216). Presumably, there will need to be a sub-premise that links certain types of behaviours to certain types of noumenal

objects. The problem is that, as Rosefeldt has pointed out, we will never be able to say anything specific or determinate about the properties of noumenal objects:

> Not only do we not know anything about the general nature of such properties, because we have no idea what properties that cannot be constructed in space and time are like; we cannot even know anything about their identity conditions under some very general description because *for all we know the same re-identifiable response-independent appearance property could be grounded in completely different response-independent properties* (2022: 40, emphasis mine).

So, while one can know that their experiences are determined by noumenal objects, it is not possible for one to say what that object is like. In other words, if Rosefeldt is right, then it is possible to say that phenomenally rational behaviours have noumenal object(s) behind them, but one cannot say with any certainty that the object picked out in premise A is the same as the object picked out in premise B. As Kant himself argues, any ascription of the 'I' to other humans should be taken "problematically" and not "apodictically" (1781: A347/B406). The possibility that they could be picking out different objects means that the positive argument for the morality of other humans is not deductively valid, even if noumenal objects can play an explanatory role in theoretical statements.

While I have focused on a positive argument for the morality of other humans in this section, it should be clear that KAAAM will face the same problem. From a two-worlds interpretation, the mental states of animals will also be grounded in some kind of noumenal object, but our lack of epistemic access to the actual properties of that noumenal object will limit our ability to say anything apodictic about it not being the same sort of object that our use of the 'I' picks out. This means that, at the very least, it is possible for the use of 'rationality' in the premises to be picking out different noumenal objects; once again, undermining the validity of KAAAM.

## 2. Empirical Markers for Morality

Where does that leave Kant on the question of animal morality? One could argue that Kant does not have anything productive to say on the matter. From a two-perspective view, his argument that freedom is necessary for moral responsibility means that morality can only be identified from a practical perspective.[4] So, either one could say that Kant's practical philosophy is irrelevant because science investigates the world from a theoretical perspective that has no room for freedom, or one could say that Kant's distinction between the theoretical and practical viewpoints shows why any scientific investigation of animal morality is doomed to failure, even if some animals can view the world from a practical vantage point. Science's theoretical perspective means that it will never be able to identify the necessary practical ingredients for moral responsibility.[5]

However, the positive argument in the section above and Kant's discussion of arguments from analogy provide a useful clue to how Kant can still have something to say on the matter. Given that this paper is interested in contributing to theoretical investigations on animal morality, the search for a deductively valid argument for either side is wrongheaded. Instead, we should be looking for the sorts of reasons that could support an inductive argument.

In *Kant's Empirical Psychology*, Frierson (2014: 171) argues that evidence for certain psychological capabilities can function as empirical markers for moral responsibility. Markers, for Kant (1803: 9:58), are "*that in a thing which constitutes a part of the cognition of it*, or - what

[4] Much of the rest of the paper will be framed in terms of the two-perspective interpretation of the noumenal/phenomenal distinction. I do this because I suspect that writing from a perspective that is entirely agnostic about the frameworks will lead to less instead of greater clarity, my own sympathies lie with the two-perspectives interpretation, and as Frierson (2010: 83) points out, it has become the dominant interpretation in discussions of Kantian morality and freedom. That said, I also suspect that most of the claims below can be translated into a two-worlds interpretation. For example, instead of taking empirical markers to represent a possible perspective, one could take them as marking a possible noumenal object.

[5] Ralph C.S. Walker (2017: 40) has argued that the inductive risk for not ascribing self-consciousness/rationality to animals is so great that from a Kantian perspective there are moral reasons to just assume that animals are self-conscious/rational. This may be so, but this still leaves the theoretical or scientific questions about animal morality open.

is the same - *a partial representation, insofar as it is considered as ground of cognition of the whole representation.*" Frierson claims that empirical markers can identify a "set of empirical elements" that fall "under a more complete concept" which serve as a marker "for the whole" (2014: 171). In this case, a marker (such as a mental capacity) can provide evidence for the whole (moral responsibility).

However, an empirical marker cannot be necessary or sufficient for ascribing moral responsibility because empirical markers are phenomenal and moral responsibility is noumenal. This means that empirical markers for transcendental freedom can give us reasons for thinking a being is morally responsible, but they cannot provide certainty. Frierson demonstrates the potentially radical consequences of this by pointing out that it remains entirely possible that a sapling that destroys its parent tree (as in Hume's famous example) "is transcendentally free and thus potentially both morally responsible and guilty of patricide" (2014: 183). While Kant's distinction between the noumenal and phenomenal does not allow us to rule out such a possibility, it does show the usefulness of empirical markers. If the sapling does not provide any markers that fit a psychological account of moral behaviour, then we have no reason to regard it as a moral agent. Those beings that have empirical markers for morality, provide reasons for potentially taking up a moral attitude towards them or ascribing a type of noumenal object to them. As Frierson points out, markers for moral responsibility "perform epistemic rather than metaphysical functions, giving viable methods for determining moral responsibility rather than transcendental conditions of its possibility" (Frierson 2014: 175). This might seem like a point of weakness for a Kantian investigation, but the underdetermination of markers for practical viewpoints or noumenal objects is not all that different from the underdetermination inherent to any inductive argument.

One way of understanding Korsgaard's emphasis on the connection between metacognition and morality is then to take KAAAM as an inductive argument where the metacognitive capabilities she identifies as rationality are an empirical marker for morality. However one could argue that this version of the argument will not hold up either because despite Korsgaard's contention that animals lack those capacities, there is a wealth of empirical research that has led to a "general acceptance" among animal behavioural scientists that nonhuman primates and other animals are capable of metacognition (Beran, 2019: 224–225). The problem for this type of critique of KAAAM is that while Korsgaard's account of rationality is metacognitive, mere metacognition will not be sufficient for Korsgaardian rationality.

For Korsgaard, rationality is a form of self-consciousness that allows a being to recognize their reasons for acting and evaluate whether those reasons are good ones on the basis of principles (2018: 299). While the ability to have metacognitive representations does show cognitive sophistication, it does not show the ability to normatively evaluate one's motivations for action. I am not arguing that animals necessarily lack this ability, some of the empirical markers I suggest below will either imply or presuppose these sorts of deliberative capacities. Instead, my point is that the metacognitive capacities studied by scientists are not the same thing as Korsgaardian rationality. The real problem for this interpretation of KAAAM is that identifying Korsgaardian rationality, even if it is an entirely theoretical phenomenon, would seem to require the ability to get fairly specific about the contents of others' maxims and the relations between them. Kant was skeptical about our ability to do this for even our own maxims, let alone the maxims of others (1785: 407), and contemporary empirical work seems to side with him on this (e.g. Nisbett and Wilson, 1977). So, while the deliberative capacities Korsgaard identifies as rationality *could* be an empirical marker for morality, I think that there are better

ones that can be drawn from Kant's empirical psychology. Contra Korsgaard, one of these markers can be established for at least some animals, in addition to suggestive evidence for three others that will require further investigation before ruling one way or the other.

## 3. The Higher Faculties: Concepts & Character

Kant understands the mind as being divided into three faculties: cognition, feeling, and desire (Frierson, 2014: 53–54). He uses these faculties to give an empirical account of both human and animal action. Both human and animal actions are caused by desires and the causes of those desires can be traced back through the faculties of feeling and cognition (Frierson 2014: 56). An object or state of affairs is cognized, the mental representation that results from the cognitive act leads to a feeling of pleasure or displeasure, and then that feeling leads to a desire. The difference between explanations for human and animal actions is that humans have the higher versions of these faculties, whereas animals only have the lower. The higher faculties of cognition and desire provide humans with "a kind of empirical freedom that animals lack" (Frierson, 2014: 56). This "psychological freedom" gives humans a level of control, 'self-activity', or 'spontaneity' but is nevertheless entirely causally determined (Kant, 1788: 5:96). While spontaneity is a part of theoretical explanations and should not be equated with transcendental freedom, Kant regards it as an empirical marker for moral responsibility (Frierson, 2014: 170). In other words, psychological spontaneity and the higher faculties that make spontaneity possible provide us with reasons for thinking that one has moral capabilities.

Despite Kant's contention that animals lack the higher faculties, contemporary empirical work provides evidence that some animals have capabilities that Kant took to be their defining features. For Kant, the lower faculties are receptive and reactive, lacking the spontaneous character of the higher faculties. The lower faculty of cognition involves the senses and the

imagination, and the actions caused by the lower faculty of desire are those that lack any conscious deliberation, such as reflexive, instinctual, or merely habitual actions (Frierson, 2014: 66). What elevates the higher faculties from their lower counterparts is conceptual cognition. Once a being is capable of forming and using concepts then it becomes possible for them to be motivated by "principles or concepts", instead of just "immediate sensation" (Frierson, 2014: 62). Since the lower faculty of cognition is limited to representations caused by the sensations or imagination, it is linked to the lower faculties of feeling and desire through the natural predispositions of instinct and inclination (Frierson, 2014: 70). In contrast, consciousness of principles and concepts through the higher faculty of cognition allows beings with the higher faculties the deliberative space to connect the higher faculty of cognition to the higher faculties of feeling and desire through character which is the ability to bind oneself to practical principles. In this section, I argue that there is empirical evidence that at least some animals are capable of conceptual cognition. In addition, I will argue that there is at least suggestive evidence some animals have the beginnings of the power of character or the types of psychological capacities that we would expect the power of character to develop out of. While I think the evidence for character is a little shakier than the evidence for conceptual cognition, examining it does provide ways for thinking through what an investigation of animal character would look like.

The ability to grasp and use concepts is one that Kant exclusively attributes to sufficiently developed human beings. Considering the role that concepts play in many other mental states and processes, such as experience, this leaves Kant in a seemingly difficult position when it comes to explaining animal behaviour. However, Naomi Fisher (2017) has argued that the lower faculties allow Kant to explain seemingly sophisticated animal behaviour through an account of

obscure representations. Kant defines obscure representations as "those of which one is not conscious" (2012: 25:479). If one is conscious of a representation, then that representation is conceptual. Obscure representations then provide Kant with a type of mental content for animals that is not conceptual.[6] An example from Kant is the sorts of features that one uses to tell that an object in the distance is a person (Fisher, 2017: 444). If one is not conscious of the markers that have convinced them that the object is a person, then the markers are obscure.

Obscure representations can explain animal action, when combined with a type of reflection that Kant was willing to attribute to animals. In the third *Critique*, Kant writes that "reflecting . . . goes on even in animals, although only instinctively, namely not in relation to a concept…but rather in relation to some inclination…" (2000: 20:211). Where human reflection allows one to bring singular representations under a general concept, animal reflection is analogous in the sense that it brings singular representations under a general inclination. Fisher argues that for Kant:

> …an animal need not bring an object under the concept of food in order to develop the inclination to eat it. Instead, the animal has an obscure representation of the food. Reflection, then operating "instinctually" and according to rules without the awareness of the animal, picks out that representation as relevant and appropriate for an inclination to eat, and that inclination to eat that object is determined (or produced), which gives rise to the action in the animal (2017: 449).

[6] This story is potentially complicated by Stefanie Grüne's (2009) claim that, for Kant, concepts can be obscure. While this is an interesting interpretive wrinkle, it does not affect Fisher's use of Kant to explain animal cognition because Kant denies that animals have conceptual capabilities (obscure or not). If animals have obscure content, it cannot involve obscure concepts because animals lack concepts. In a response to Colin McLear's argument that her account does not fit with Kant's views on the mental lives of animals, Grüne (2014) argues for many of the same sorts of cognitive limitations that Fisher does. For example, Grüne reads Kant as saying that animals are conscious of individual sensations whereas it is not possible for them to have consciousness of complex representations, such as the relation between "individual sensations to the unity of the representation of their object." Whereas (as discussed below) Fisher claims that the lack of consciousness involved in the obscure representations of animals means that they cannot attend to the relations between representations. Either way, the behavioural consequences of these cognitive limitations for animals should be the same and my argument for attributing conceptual capabilities to some animals should still apply, even if one prefers Grüne's interpretation.

Furthermore, the non-spontaneous part of the imagination, that Kant is willing to attribute to animals, can construct associations between representations through behavioural processes, like conditioning (Fisher, 2017: 450).

While the explanation for animal action that Fisher extracts from Kant provides an account of animal behaviour, the lack of conceptual capabilities means that there will be limitations on their behavioural abilities. The lack of conceptual capabilities means that animals lack a unified consciousness, so any awareness that animals have will be "disconnected and episodic" (Fisher, 2017: 452). A result of this lack of unity is that animals lack an inner sense, which is the ability to be conscious of one's own representations, since Kant defines inner sense as the ability to distinguish a given representation from one's other representations (Fisher, 2017: 454). Fisher argues that Kant can still explain an animal's ability to distinguish between different objects, but this ability will be through physical differentiation, not through any sort of logical comparison. She writes that "animals may be acquainted with objects according to the similarity and difference without conscious comparison of representations: through inclination determining reflection, which will produce different (or similar) responses to different (or similar) obscure representations" (Fisher, 2017: 455). Animals are able to react differently to different objects because they cause different obscure representations. They lack the ability to compare those representations, so they will lack any consciousness of what that difference is.

The inability to compare representations means that beings that lack conceptual capabilities should not be able to succeed at behavioural tasks that require them to use the representation of a similarity or difference between two or more objects to make a choice. Now, this does not mean that Kant's account cannot explain some types of behavioural experiments that might initially seem to require such a representation. For example, Kant's account of

nonconceptual cognition is sophisticated enough to provide a plausible explanation for identity matching-to-sample (IMTS) tasks. IMTS tasks require the subject to match a sample card to another card that it shares some feature with. For example, in Smirnova et al. (2021), crows and amazons were shown a sample card, and then shown two possible match cards. If the subject is matching on the basis of colour, a sample card might be white, while one of the matching cards is white and the other is black. If testing for the ability to make similarity judgements, the correct choice is the white card, while if testing for the ability to make judgements about difference, the correct choice is the black card. The birds were trained and then tested using novel, never seen before, cards on tasks for matching colour, shape, and number, before being tested on size with no additional training. In most of these tasks, the feature to be matched was one of several features on the card. For example, numbers were represented by the number of shapes on a card.[7] A sample card representing two could have a red square and a green triangle. The only feature that the correct matching card would share with the sample card is the number of shapes, and not the type of shapes or colours.

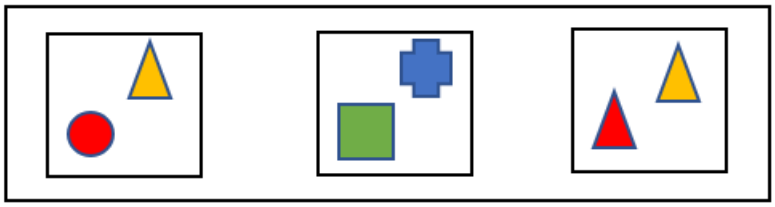
Kant's account of particulars falling under inclinations can offer a possible explanation for success at these behavioural tasks. In IMTS tasks, an obscure representation of the sample could lead to an inclination to match with whatever feature the subjects have just been shown. This inclination could be created and reinforced by conditioning from the training procedures. This sort of explanation works best when there is only one feature on each card (e.g. colour), but it can also potentially work for more complicated tasks, where the cards have more than one feature on them (e.g. coloured shapes that represent numbers by amount). The birds could have

[7] If Frege (1950) is right that numbers are second-order concepts, then the ability of crows and amazons to match cards based on the number of objects on them already shows some ability with abstract relational concepts. For explorations of what Frege's understanding of number concepts means for animal behavioural tasks, see Nelson (2020) and Clarke and Beck (2021).

an obscure representation of the sample card that represents multiple features. The obscure representation could lead to an inclination to pick a card that matches at least one of those cards, or perhaps each feature leads to its own inclination, and the one that gets selected is based on the possible matching cards that are presented to the birds after the sample. If one of these stories works, then Kant can explain the success of the birds at IMTS tasks, while holding that the birds lack representations of similarity and difference, as well as any consciousness of the particular features guiding their inclinations. In other words, these explanations do not require conceptual mental states or processes.

However, the limits of Kant's story are shown by the next set of tasks that Smirnova et al. (2021) tested the crows and amazons on. Without any additional training, the birds were tested on relational matching-to-sample (RMTS) tasks. RMTS tasks require the subject to match cards that share relations instead of physical features. For example, the sample card could have a blue cross and a green square on it, whereas the potential matches could then have a red circle and a yellow triangle on one, and a red triangle and a yellow triangle on the other (see Figure 1). The correct match is the first card because, like the sample card, it has two different types of shapes on it. In other words, the similarity they share is difference. The crows and amazons did almost as well on the RMTS tasks as they did on the IMTS tasks.[8]

**Figure 1.**



[8] Symbol trained chimpanzees have also succeeded at RMTS tasks (Thompson, Oden and Boysen, 1997).

*Figure 1. The author's representation of a possible set of cards for an RMTS task from Smirnova et al. (2021).*

It is not clear how Kant's account of obscure representations can provide an explanation for RMTS tasks. A story about how obscure representations fall under inclinations that then explain the birds' choices needs to identify what parts of the representations are doing the work (even if the subject is not conscious of them). The cards share none of the same physical features; they only share the relations of similarity or difference. If one tries to claim that a sample card provides a singular representation that then falls under an inclination to pick the matching card, then it seems like it would need to be a representation of similarity or difference, something Kant claims is not possible without conceptual capabilities. If one tries to deny this possibility, then we are pushed back into an explanation that refers to a comparison of representations, something that Kant also claims is impossible without conceptual capabilities. The usual way to explain these types of abilities is to treat the subject as bringing something to the task themselves. In other words, the relation would be identified through the application of a concept. There is no need to select which of these explanations is the right one because all three, according to Kant, require conceptual capabilities. Considering that Kant thought that conceptual thought distinguished the higher from the lower faculties, success at RTMS tasks provides evidence for an empirical marker for morality (the higher faculties) in animals.

One might object that Kant's account of nonconceptual cognition is too impoverished, and that a more sophisticated contemporary account of nonconceptual cognition will be able to explain success at RMTS tasks. I think that this objection underestimates the relative sophistication of Kant's account of nonconceptual cognition. Many contemporary Kantians characterize animals as completely lacking any cognitive capabilities (e.g. Brandom, 1994),

meaning that they will likely be unable to even explain the IMTS tasks without invoking conceptual capabilities. However, perhaps a perceptual rule along the lines of 'seek X' could be used to explain success at RMTS tasks, where X is a nonconceptual memory or representation of the sample card (Pepperberg, 2021). This sort of explanation seems sufficient for IMTS tasks (e.g. 'seek X' where X is a perceptual representation of a square), but unless X is a representation of same or different, it is unclear how this explanation is supposed to work.

Remember that in many of the RMTS tasks, the correct card shares no physical features with the sample card, instead all they share is the relation of similarity or difference. Same and different are abstract relations, like on or under, which means that they "do not have a bounded, identifiable and clearly perceivable referent" (Borghi *et al.*, 2017: 263). There is no object that one can point to and say "see, that is what difference is." Furthermore, there is no limitation to the types of objects that can fall under the concept, as long as they share the relation of difference. The fact that same and different are relations means that they have an inferential structure built into them. For example, one cannot answer questions involving the concept on, like "What is on the shelf?", without some further ability to classify objects as objects, and understand those objects as sharing relations. Additionally, basic abstract relations often come with a conceptual contrary (e.g. same vs. different, on vs. under, etc.). Same/different transfer tasks where Clark's nutcrackers have a choice between identifying a set of pictures as same or different have shown that subjects have similar levels of success on both classifications, at the very least, suggesting that these capabilities are intertwined (Magnotti *et al.*, 2015; see also Hochmann, 2021: 137). These sorts of holistic or semantic web-like features, in addition to relative stimulus independence, are the very ones that many theorists take to be distinguishing marks of the conceptual (Newen and Bartels, 2007). In other words, the only way to make sense

of a representation of same or different in a way that it could play a role in a rule like 'seek X' where X is difference, is to give it features that distinguish conceptual representations from merely perceptual ones. The abstractness of same and different means that they have a level of independence from immediate perceptual cues and their relationality means that they only make sense as part of a semantic web.[9]

While I have focused on the details of a particular set of experiments in this section, it is worth noting that there is a much wider set of results in animal behavioural science that Kant's account of obscure representations will have trouble dealing with. For example, further evidence that some nonhuman animals are capable of comparing representations to identify what differentiates them comes from Pepperberg's (2021) work with African grey parrots, especially her subject Alex. If nonconceptual representations are obscure, then animals that lack conceptual capabilities should not be able to succeed at tasks that require them to identify what makes those objects the same or different (van den Berg, 2018: 8). After being trained to apply labels to objects based on their colour, shape, or matter (e.g. wood), Alex was trained to answer questions about what makes two objects the same or different when presented with a pair of them. Alex was then tested on objects for which he lacked the words for their colours, shapes or matter. If asked "what's different?" for a pink plastic flamingo and a pink plastic elephant, he could

[9] As an anonymous reviewer has stressed, there are well-known arguments against attributing conceptual or propositional content to animals. These arguments claim that a mismatch between our semantic web and whatever sort of web an animal has means that any attribution will misfire. For example, it seems odd to attribute the concept squirrel to a dog, given that the dog does not understand that a squirrel is a mammal. However, for abstract relational concepts, it is not clear what further sorts of beliefs would cause this semantic mismatch. At the very least, the semantic web seems to require a conceptual contrary, but as mentioned above, there is suggestive evidence that animals can meet this standard. Furthermore, as Newen and Starzak (2022: 15) have pointed out, the same sorts of problems arise when it comes to other humans, and there is even a level of indeterminacy when it comes to attributing mental content to ourselves, given the difficulties that individuals have in linguistically representing their core beliefs. This means that either we should be skeptical about all attributions of conceptual content, whether to humans or animals, or something has gone wrong with these arguments. I suspect it is the latter, but such an argument falls outside the scope of this paper.

respond "shape" (Pepperberg, 2021: 148). In his first set of test trials, Alex was already achieving 85% accuracy. These results seem to demonstrate that Alex was not only capable of comparing representations, but that he was conscious enough of those representations that he was able to isolate what part of the representation distinguished it from another.

Further evidence for animals' abilities to use abstract concepts, compare representations, or be conscious of complex representations can be found in work beyond evaluating representations of sameness or difference. For example, Hosokawa et al. (2018) argue that Japanese monkeys are able to succeed at group reversal tasks by forming functional categories for equivalence.[10] Group reversal tasks require a subject to initially respond to images where a response to some leads to an appetitive reward (e.g. fruit juice) and a response to others leads to an aversive stimulus (e.g. saline). After the subject has learned which images lead to the rewards or the punishments, the relation is switched so that the previously punished responses are rewarded and vice versa. Hosokawa et al. (2018) argue that the ability of Japanese monkeys to rapidly adapt to these reversals show that they have formed equivalence classes for each set of stimuli. In addition, Tsutsui et al. (2016) have used neuroscientific evidence to show that the category level representations of Japanese monkeys that are used to succeed at these tasks are coded independently of representations for individual stimuli. This kind of chunking of simple or individual representations into category level representations would seem to require the sorts of capabilities that Kant would exclusively attribute to creatures with conceptual capabilities and hence the higher faculties. [11]

[10] Success at these types of tasks has also been demonstrated by pigeons (Vaughan, 1988), dolphins (von Fersen and Delius, 2000), sea lions (Kastak, Schusterman and Kastak, 2001), and chimpanzees (Tomonaga, 1999).

[11] I would like to thank an anonymous reviewer for emphasizing the importance of discussing the extent of the experimental results that Kant's cognitive picture of animals will have trouble dealing with without attributing conceptual capabilities and for suggesting a far greater wealth of potential examples than I have room to fully discuss here.

For Kant, the ability to think conceptually means that one can form principles and then make decisions about what principles to act on.[12] Kant thinks that this ability uniquely provides the power of character to beings with the higher faculties. In his empirical psychology, Kant defines character as "nothing other than that which is peculiar to the higher capacities" (2012: 25:227) which is defined as "that property of the will by which the subject binds himself to definite practical principles" (2006: 7:292). In the higher faculties, character is what grounds the connection between cognition and the faculties of feeling and desire, similarly to the way that the instincts and inclinations ground those connections for the lower faculties (Frierson, 2014: 72). It is this ability to choose one's principles that provides the empirical spontaneity that Kant regards as a marker for morality. For Kant, character is something that is acquired and cultivated (Frierson, 2014: 77–78), and the level of acquisition or cultivation can be demonstrated through the firmness with which one holds their principles (Kant, 2012: 25:1175). Character is not only involved in sticking by moral principles but can be involved in any decision that involves sticking by one's practical principles, such as sticking by 'early to bed' or 'eat healthy' even when tempted to do otherwise (Frierson 2014: 74). This means that character can come in degrees and that behavioural tasks that involve self-control, whether they are explicitly moral or

[12] One might worry that the attribution of concepts does not justify the attribution of propositional content. Given Kant's commitment to the priority of the propositional (1781: A68/B93), this objection would not make much sense to him. However, for those that give concepts priority, the rest of this section can also run as a potential argument for attributing propositional attitudes to animals in addition to attributing concepts. While it might not seem immediately obvious that the experiments discussed below require propositional content, Newen and Starzak (2022: 17–18) have argued that temporal decisions, similar to the ones involved in delayed self-gratification tasks, require a level of structure and information integration that justifies the attribution of beliefs. In addition, Newen and Bartels (2007: 298) provide an argument for thinking that similar sorts of structures justify the attribution of propositional content.

not, can demonstrate that one has character or at the very least demonstrate the types of capacities that underly character.

One of the ways in which self-control has been tested in both humans and animals is through delayed self-gratification tasks. These tasks assess if a subject is capable of not taking a smaller reward if they know that waiting will lead to a larger one. For example, chimpanzees have been tested on accumulation tasks where the longer they wait to take a food reward, the more food accumulates. Beran and Evans (2006) found that all of their chimpanzee subjects were able to wait for multiple minutes before taking the reward. In studies where the chimpanzees had access to items that they could use to distract themselves, they used similar self-distraction techniques as those used by human children. Chimpanzees that had access to these items were able to delay gratification longer and accumulate a larger reward than those that did not (Evans and Beran, 2007). Self-distraction techniques were observed less in cases where the subjects did not have to self-impose delay maintenance, such as when the accumulating food was out of reach, suggesting that the "chimpanzees apparently knew when they needed to do something to boost or support their self-control" (Beran, 2015: 354). These results suggest that chimpanzees are capable of making choices that require some measure of self-control.

One might reasonably object that these results are not sufficient to demonstrate character, but they will need an argument for why self-control in one case (going to bed early, even when invited out by friends) counts as character whereas as it does not in other cases (withholding action in order receive a greater reward). One tempting way to make this argument might be on the basis of whether a principle is involved or not, but two ways of making this argument have already been cut off. First, if the argument is that principles require concepts, I have already provided reasons for thinking that a Kantian perspective should acknowledge that at least some

animals have conceptual capabilities. Second, if the argument is about specifying a particular practical principle, I have already pointed out that Kant is skeptical of our ability to identify such principles for other humans or even ourselves from a theoretical perspective. That said, I think it is reasonable to want something a bit more sophisticated here. For that reason, I think these results may be better read as showing the sorts of capacities that can be developed into character or the very early stages of character that humans are able to cultivate further. However, if this much is conceded, then I think that my central claim that a Kantian perspective both cannot rule out the possibility of moral animals and that it provides ways to interpret and shape future investigations should be accepted.[13]

Now, one might reasonably object that while both conceptual capabilities and character are part of Kant's account of the higher faculties, there is nothing necessarily moral about them.[14] And while Kant takes moral capabilities to emerge from the higher faculties, since we are discussing his empirical psychology, they cannot be necessary and sufficient for them. Therefore, it is at least possible that one could have the higher faculties while lacking moral capabilities. For that reason, I now turn to Kant's empirical account of moral psychology and argue that there is at least preliminary evidence that merits further investigation into whether some animals have empirical markers extracted from that account.

---

[13] There are of course potentially alternative ways to interpret the experiments and, as a reviewer stressed, imagined rewards might be enough to explain the results without attributing practical principles to animals (as noted earlier, Kant attributes the imagination to animals). However, even if this possibility is conceded, the point still stands. The inductive nature of theoretical arguments means that any interpretation will always be underdetermined to some degree. Evaluating which interpretation is better will require further investigation and given that both alternatives are to some degree shaped by Kant's picture of minds and morality, there is a potential contribution for his work to make here. So, the central claim that the question is an open one and that Kantian ideas can productively shape answers to it is still in good standing.

[14] Though empirical work has demonstrated that rats will refuse to press a lever for food when pressing it will also shock a rat in a neighbouring cage (Church 1959). Similar results have also been found with pigeons and rhesus macaques, including a macaque that resisted shocking a fellow macaque for twelve days, despite lacking an alternative way to access food (Wechkin, Masserman and Terris, 1964).

## 4. Kant's Empirical Moral Psychology

Kant's empirical account of moral psychology has the same structure as his more general psychology (Frierson, 2014: 123). Morally motivated actions can be explained through the connections between the higher faculties of cognition, feeling, and desire. A moral action starts with a cognition of the moral law, which leads to a feeling of respect, and that feeling causes a desire "to act in accordance with the moral law" (Frierson, 2014: 123). For example, one might cognize the principle 'refrain from false promises' which leads one to feel respect for what makes that principle morally right, and that feeling triggers the desire to refrain from false promises. In cases where that desire leads to a morally right action, one will choose an action that refrains from false promises, even if another option is compelling.

One might reasonably worry that, even if animals have conceptual capabilities and/or character, this sort of mental process will still be beyond their capabilities. However, it is important to not interpret the above picture as claiming that individuals must use some explicit form of the categorical imperative in their ordinary moral reasoning processes. Kant frames his project as rooted in common-sense morality (1788: 5:9n), so identifying empirical markers for moral responsibility must start with "common-sense moral judgements about moral responsibility" (Frierson, 2014: 186). Ido Geiger has argued that "Kant's discussion of the FUL [Formula of Universal Law] assumes agents who have ordinary usually informal and implicit knowledge of the moral realm" (2010: 291). Even without access to an explicit form of the categorical imperative, these agents know "in most everyday situations…what their duties are and what actions fulfill and violate them" (2010: 281). Philosophical knowledge of the categorical imperative, such as Kant's formulations, make explicit "the contradiction between a temptation to violate our duty and a moral law we know well binds us" (Geiger, 2010: 286), and

in doing so, allows us to express our moral duties in an explicit form. In other words, Kant's practical philosophy already presupposes agents that know the moral law and what their duties are, but outside of the philosophical domain, that knowledge is implicit. However, even in these implicit forms, to have a duty is to understand that the moral law commands universally (Geiger, 2010: 281). While understanding an explicit formulation of the categorical imperative is clearly beyond the abilities of animals, and identifying an implicit grasp of the moral law and/or duties will be exceptionally difficult, we can ask: could animals have the sorts of basic moral capabilities that the categorical imperative is attempting to make explicit?

One possible avenue for answering this question comes from behavioural research on fairness. Geiger writes that the contradiction that the categorical imperative makes explicit is most obvious in "struggles against the inclination to make an exception for ourselves" in the face of a universal duty (2010: 291). Violating a duty demonstrates a kind of partiality to oneself that is unfair to the moral demands of others. I am not claiming that fairness and universality are necessarily the same thing, but that a sense of fairness could help demonstrate an individual's awareness of the universality of moral duties. In other words, an implicit understanding of the universality of the moral law is an empirical marker for moral responsibility, and a sense of fairness could provide evidence for the existence of that marker.

While behavioural scientists are generally wary of morally loaded terms like fairness, Brosnan and de Waal (2003) have shown that brown capuchin monkeys have an aversion to disadvantageous inequities. In their experiment, the subjects would receive unequal rewards (either a grape or a cucumber) for the same task. The monkey that received the less preferred cucumber while watching another monkey receive a grape would eventually refuse to do the task or would refuse the reward. These sorts of behaviours were not found if both subjects received

the less preferred reward. These acts of protest suggest that an unfair distribution of rewards can influence the desires and actions of the monkeys.

One might reasonably object that the monkeys are protesting when they are unfairly made the exception compared to an advantaged other, whereas the categorical imperative is about not unfairly making oneself the advantaged exception. There is less evidence for advantageous inequity aversion, but some experiments with rats, capuchin monkeys, and chimpanzees (see Oberliessen and Kalenscher, 2019) seem to suggest such an aversion. These experiments are usually set up so that a subject has to decide whether to give just themself a reward or to give both themself and another subject, who has also completed the task, a reward. There is also evidence that wild chimpanzees in the Taï Forest divide up meat after a successful hunt based on the level of participation, regardless of the existing social hierarchy (Boesch 2002). While there are certainly legitimate concerns about inferring too much from these studies, they do suggest that some animals have a sense or awareness of fairness. [15]

While my aim here is more to signal a potential empirical marker instead of making any strong claims about whether animals actually have that marker, any attempt to actually establish such a marker will have to deal with a number of prominent objections to attributing a sense of universality to animals other than humans. Even commentators as sympathetic as de Waal (2014) have argued that the distinction between the proto-moral practices of animals and the moral practices of humans are the abstraction and universality of the latter. I think this type of worry

[15] Lucas Thorpe (2018) has also suggested ways in which prelinguistic infants and nonhuman primates could have the psychological capabilities connected to the second and third formulations of the categorical imperative. He withholds attributing them to nonhuman primates because of experimental results that seem to show they lack shared intentionality (Moll and Tomasello, 2007). However, there are models of shared intentionality that are less cognitively demanding and can conceivably be attributed to nonhuman primates, such as bonobos (e.g. Papadopoulos, 2021). I think these arguments could support the case I am making here, but for reasons of space, I will have to leave that discussion for another time.

can potentially be met on two levels. First, it is not clear how universally oriented behaviour would need to be in order to demonstrate an implicit grasp of the moral law. Humans are often xenophobic, nepotistic, and fail in both minor and horrifying ways to universally treat other humans as morally relevant others. So, if we are to regard animal behaviours as too partial to count as properly moral, then we need to make sure we are not doing so just because we are holding it up against an anthrofabulated human standard. Recognizing the moral law does not entail always acting in accordance with it, especially if that recognition is implicit or underdeveloped. Second, if human morality does turn out to have greater universality, this does not necessarily rule out the claim that animals could have some implicit sense of that universality. We might regard some human ability or condition, such as language, living in larger communities, or the development of moral philosophy, to be what sharpens that implicit sense or brings it into the forefront that then allows it to be developed or made more explicit. I do not think that either of these claims in their current condition completely avoid the above objections, but both highlight the need for more scientific and philosophical work before coming down on one side or the other.

The final empirical marker I will draw from Kant's empirical moral psychology comes from Kant's account of the role of feeling in moral motivation. Kant (in)famously claimed that "if the determination of the will takes place…by means of a feeling…then the action [lacks] morality" (1788: 5:71). The point of this claim is not to deny that the faculty of feeling plays a role in moral motivation from a theoretical perspective; instead, the point is to deny that feelings can play a justificatory role in moral decision making from a practical perspective. If I give money to a charity because I know that it will give me pleasure, then the reason for my action is

not moral. Though, if I give to charity because it is the morally right thing to do, one can still analyze my action theoretically as involving a desire that is caused by a feeling.

According to Frierson, Kant's claim that moral actions cannot be determined by feelings can also be interpreted psychologically, and that this interpretation provides a way to empirically distinguish moral motivation from non-moral motivation (2014: 151). Pleasure can play more than one role in the motivation of action. Frierson refers to the role of pleasure in Kant's standard account of action as antecedent practical pleasures (APPs) (2014: 151). When an APP plays a role in a morally motivated action, a cognition of the moral law leads to a feeling of pleasure which motivates a desire. This contrasts with cases where a cognition of anticipated pleasure (CAP) motivates action. A CAP represents the pleasure "to oneself in cognition as being the subject result of an object, action, or state of affairs…" (Frierson 2014: 152). When the starting point of an action comes from a CAP, then the action is nonmoral. To put this in the language of markers, an APP is a reason for thinking that an action is morally motivated, whereas if the action is the result of a CAP, then we have a reason for thinking that the action is not morally motivated.

If this distinction can be used when analyzing animal behaviour, then it provides an advantage to Kantian inspired investigations of animal morality. Distinguishing between actions motivated by moral norms and actions motivated by non-moral social norms has led to skepticism about whether such a distinction can even be made (e.g. Andrews, 2013: 187–189). However, Kant's distinction between APPs and CAPs shows that the difference between morally motivated actions and non-morally motivated actions is not about whether the norm being followed is moral or social. In both APPs and CAPs, one can have a cognition of the moral law. Furthermore, social norms, such as driving on the right side of the road, can play a role in

morally motivated action, if one is following them for moral reasons (e.g. one can drive on the right side of the road because of the social or legal norm, but one can also drive on the right side of the road because to do otherwise would immorally put the lives of others at risk). Kant's distinction shows that one has a reason to think that an action is morally motivated if the pleasure motivating a desire comes from a cognition of the moral law instead of a cognition of the pleasure that an act will bring.

I concede that making this sort of distinction will be difficult when it comes to analyzing the behaviour of animals. However, empirical work on the role of empathy in rat behaviour shows that it is not impossible. Experiments have shown that rats will help other rats, especially when the other rat is distressed. For instance, rats are more likely to release a cage mate that is drowning in water than one that is not distressed. Debunking arguments have been made that "the animals...may have been simply experiencing the other's pain as an aversive stimulus" (Monsó and Andrews, 2022: 392). However, scientists have found that when two restrainers are placed in a rat's cage, one containing chocolate chips and the other another rat, the rat already in the cage will open both restrainers and share the food with the released rat (Bartal, Decety and Mason, 2011). This shows that their behaviour is not solely driven by an aversion to the distressed behaviour of others, and that rats are willing to act 'beneficently' even if it means making a sacrifice (i.e. sole possession of the food).

Some scientists have argued that "the helping rat was acting out of an egoistic desire for social contact" that was greater than the desire for food (Monsó and Andrews, 2022: 392). This possibility was controlled for by having the restrainer for the rat open into a second cage that was inaccessible to the unrestrained rat, meaning that there was no social reward for releasing the restrained rat (Bartal, Decety and Mason, 2011). These experiments found that the unrestrained

rat would still release the restrained rat. Further debunking arguments have been made, and further experiments have been done in response to those arguments. I will not go through all of them because my aim here is not to prove that rats are capable of empathy. Instead, my point is that the experimental responses to the debunking arguments can be interpreted from a Kantian perspective as ruling out a CAP interpretation of the behaviour. The debunking arguments can be read as pointing out that some anticipated pleasure, such as ending an aversive stimulus or creating an opportunity for social contact, is what is actually motivating the seemingly empathetic action. While there may still be reasons to think that the behaviour of the rats is not morally motivated (e.g. maybe they lack other moral markers, such as conceptual capabilities, or the APP is not triggered by the moral law, or there are further reasons for thinking that a CAP is involved instead of an APP), but the scientific debate above shows that it may be experimentally possible to distinguish between APPs and CAPs.[16] Therefore, a Kantian inspired investigation of animal morality could provide an empirical marker for what makes morally motivated behaviour distinct from other types of normative behaviour.

## 5. Conclusion

In this paper, I have argued that Kantians should regard the question of whether animals are capable of morality, as an open question. I have done that by arguing that Korsgaard's argument against animal morality does not work on Kantian grounds. Her argument interpreted

[16] An anonymous reviewer has suggested that reciprocal altruism may provide a more parsimonious explanation for these results. However, if reciprocal altruism is being proposed as an answer to questions about ontogeny, evolution, or survival value, then it is not clear that APPs and reciprocal altruism are competing explanations (whether APPs fit into the category of causation or an additional one beyond Tinbergen's four levels of analysis, such as private experience (Burghardt, 1997)). Instead, reciprocal altruism could explain the adaptive value of the behaviour, while APPs could explain the psychological causes of the behaviour. If reciprocal altruism is being proposed as a competing causal hypothesis, then selecting between the two hypotheses should not be done on the basis of parsimony, but on further experimental results. My point here is not to argue that the rats necessarily have an APP, instead my interpretation is meant to show that investigating the existence of APPs is not an in principle impossible task.

from a two-perspective view relies on an equivocation of the term ‘rationality’. Kant’s distinction between the noumenal and phenomenal means that no phenomenal trait can be necessary or sufficient for morality. Morality relies on freedom, which only makes sense from a practical perspective. In addition, a two-worlds interpretation is not sufficient for saving the validity of KAAAM because there is no way to guarantee that the noumenal object referred to in both premises is the same object. This means that if we are looking for indications that one is capable of morality, the best we can look for is an empirical marker. I have argued that Kant’s discussion of the higher faculties and his empirical account of moral motivation provide markers for morality that can be identified from a theoretical perspective. Given Kant’s account of conceptual capabilities, there is evidence that at least some animals have the higher faculties, and given his account of character, there is suggestive evidence that merits further investigation into whether some animals have the sorts of capabilities we would expect to underly such a power. Furthermore, there is suggestive evidence that animals can be motivated by fairness and APPs, providing further potential avenues for a Kantian investigation into animal morality. This final marker provides a Kantian account an advantage over other investigations of animal morality because it provides a potential way to distinguish morally motivated behaviour from other forms of normative behaviour.

**Acknowledgements:**

I would like to thank audiences at the Dalhousie Philosophy Colloquium, the Reconstructing Reason Graduate Conference at University of Toronto, the Canadian Philosophical Association’s meeting at the University of Alberta, and the Dartmouth College Work in Progress Group for their helpful feedback. I would especially like to thank Katherine Crone for her insightful

commentary and her patience with my last-minute changes. Special thanks to two anonymous reviewers at *Ergo*, Greg Scherkoske, Andrew Fenton, the Halifax Animal Studies Group, and the Dalhousie Philosophy Graduate Student Writing Group. Most of all, this paper would not have been possible without the countless hours of discussion and debate on Kant and Kantian philosophy with Andrew Lopez. Finally, I would like to thank Sissi for keeping me company during many hours of revisions.

## References

Andrews, K. (2013) 'Ape Autonomy? Social Norms and Moral Agency in Other Species', in K. Petrus and M. Wild (eds) *Animal Minds & Animal Ethics: Connecting Two Seperate Fields*. transcript, pp. 173–196.

Bartal, I.B.-A., Decety, J. and Mason, P. (2011) 'Empathy and Pro-Social Behavior in Rats', *Science*, 334(6061), pp. 1427–1430. Available at: https://doi.org/10.1126/science.1210789.

Beran, M. (2019) 'Animal metacognition: A decade of progress, problems, and the development of new prospects.', *Animal Behavior and Cognition*, 6(4), pp. 223–229. Available at: https://doi.org/10.26451/abc.06.04.01.2019.

Beran, M.J. (2015) 'Chimpanzee Cognitive Control', *Current Directions in Psychological Science*, 24(5), pp. 352–357. Available at: https://doi.org/10.1177/0963721415593897.

Beran, M.J. and Evans, T.A. (2006) 'Maintenance of delay of gratification by four chimpanzees (Pan troglodytes): The effects of delayed reward visibility, experimenter presence, and extended delay intervals', *Behavioural Processes*, 73(3), pp. 315–324. Available at: https://doi.org/10.1016/j.beproc.2006.07.005.

van den Berg, H. (2018) 'A blooming and buzzing confusion: Buffon, Reimarus, and Kant on animal cognition', *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 72, pp. 1–9. Available at: https://doi.org/10.1016/j.shpsc.2018.10.002.

Boesch, C. (2002) 'Cooperative hunting roles among taï chimpanzees', *Human Nature*, 13(1), pp. 27–46. Available at: https://doi.org/10.1007/s12110-002-1013-6.

Borghi, A. *et al.* (2017) 'The Challenge of Abstract Concepts', *Psychological Bulletin*, 143(3), pp. 263–292. Available at: https://doi.org/10.1037/bul0000089.

Brandom, R. (1994) *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press.

Brosnan, S.F. and de Waal, F.B.M. (2003) ‘Monkeys reject unequal pay’, *Nature*, 425(6955), pp. 297–299. Available at: https://doi.org/10.1038/nature01963.

Burghardt, G.M. (1997) ‘Amending Tinbergen: A fifth aim for ethology’, in *Anthropomorphism, anecdotes, and animals*. Albany, NY, US: State University of New York Press (SUNY series in philosophy and biology), pp. 254–276.

Church, R.M. (1959) ‘Emotional reactions of rats to the pain of others’, *Journal of Comparative and Physiological Psychology*, 52(2), pp. 132–134. Available at: https://doi.org/10.1037/h0043531.

Clarke, S. and Beck, J. (2021) ‘The number sense represents (rational) numbers’, *Behavioral and Brain Sciences*, 44. Available at: https://doi.org/10.1017/S0140525X21000571.

Darwin, C. (1871) *The Descent of Man, and Selection in Relation to Sex*. London: John Murray.

Evans, T.A. and Beran, M.J. (2007) ‘Chimpanzees use self-distraction to cope with impulsivity’, *Biology Letters*, 3(6), pp. 599–602. Available at: https://doi.org/10.1098/rsbl.2007.0399.

von Fersen, L. and Delius, J.D. (2000) ‘Acquired equivalences between auditory stimuli in dolphins (Tursiops truncatus)’, *Animal Cognition*, 3(2), pp. 79–83. Available at: https://doi.org/10.1007/s100710000063.

Fisher, N. (2017) ‘Kant On Animal Minds’, *Ergo, an Open Access Journal of Philosophy*, 4. Available at: https://doi.org/10.3998/ergo.12405314.0004.015.

Frege, G. (1950) *The Foundations of Arithmetic*. Translated by J.L. Austin. Oxford: Blackwell.

Frierson, P.R. (2010) ‘Two Standpoints and the Problem of Moral Anthropology’, in B.J.B. Lipscomb and J.K. Krueger (eds) *Kant’s Moral Metaphysics: God, Freedom, and Immortality*. Germany: De Gruyter, pp. 83–110.

Frierson, P.R. (2014) *Kant’s Empirical Psychology*. Cambridge: Cambridge University Press.

Geiger, I. (2010) ‘What is the Use of the Universal Law Formula of the Categorical Imperative?’, *British Journal for the History of Philosophy*, 18(2), pp. 271–295. Available at: https://doi.org/10.1080/09608781003643568.

Grüne, S. (2009) *Blinde Anschauung: Die Rolle von Begriffen in Kants Theorie sinnlicher Synthesis*. Frankfurt: Vittorio Klostermann.

Grüne, S. (2014) ‘Reply to Colin McLear’, *Critique*. Available at: https://virtualcritique.wordpress.com/2014/08/20/reply-to-colin-mclear/ (Accessed: 29 June 2023).

Hochmann, J.-R. (2021) ‘Asymmetry in the complexity of same and different representations’, *Current Opinion in Behavioral Sciences*, 37, pp. 133–139. Available at: https://doi.org/10.1016/j.cobeha.2020.12.003.

Hosokawa, T. *et al.* (2018) 'Behavioral evidence for the use of functional categories during group reversal task performance in monkeys', *Scientific Reports*, 8, p. 15878. Available at: https://doi.org/10.1038/s41598-018-33349-3.

Hume, D. (1739) *A Treatise of Human Nature*. Oxford: Oxford University Press.

Hutton, J. (2020) 'Kant, Animal Minds, and Conceptualism', *Canadian Journal of Philosophy*, 50(8), pp. 981–998. Available at: https://doi.org/10.1017/can.2020.50.

Kant, I. (1781) *Critique of Pure Reason*. Edited and translated by P. Guyer and A.W. Wood. Cambridge: Cambridge University Press.

Kant, I. (1785) 'Groundwork of the Metaphysics of Morals', in M.J. Gregor and A.W. Wood (eds), M.J. Gregor (tran.) *Practical Philosophy*. New York, New York: Cambridge University Press, pp. 37–108.

Kant, I. (1788) 'Critique of Practical Reason', in M.J. Gregor and A.W. Wood (eds), M.J. Gregor (tran.) *Practical Philosophy*. Cambridge: Cambridge University Press, pp. 133–272.

Kant, I. (1803) 'Lectures on Pedagogy', in G. Zöller and R.B. Louden (eds), R.B. Louden (tran.) *Anthropology, History, and Education*. Cambridge: Cambridge University Press, pp. 434–485.

Kant, I. (2000) *Critique of the Power of Judgment*. Edited by P. Guyer. Translated by P. Guyer and E. Matthews. Cambridge University Press.

Kant, I. (2006) *Anthropology from a Pragmatic Point of View*. Edited and translated by R.B. Louden. Cambridge University Press.

Kant, I. (2012) *Lectures on Anthroplogy*. Edited by A.W. Wood and R.B. Louden. Translated by R.R. Clewis et al. Cambridge: Cambridge University Press.

Kastak, C.R., Schusterman, R.J. and Kastak, D. (2001) 'Equivalence Classification by California Sea Lions Using Class-Specific Reinforcers', *Journal of the Experimental Analysis of Behavior*, 76(2), pp. 131–158. Available at: https://doi.org/10.1901/jeab.2001.76-131.

Korsgaard, C.M. (1996a) 'Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations', in *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press, pp. 188–221.

Korsgaard, C.M. (1996b) 'Morality as Freedom', in *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press, pp. 159–187.

Korsgaard, C.M. (2004) 'Fellow Creatures: Kantian Ethics and Our Duties to Animals', *Tanner Lectures on Human Values*, 24, pp. 77–110.

Korsgaard, C.M. (2006) 'Morality and the Distinctiveness of Human Action', in F. de Waal, S. Macedo, and J. Uber (eds) *Primates and Philosophers: How Morality Evolved*. Princeton, New Jersey: Princeton University Press, pp. 98–119.

Korsgaard, C.M. (2010) ‘Reflections on the Evolution of Morality’, *The Amherst Lecture in Philosophy*, 5, pp. 1–29.

Korsgaard, C.M. (2011) ‘Interacting with Animals: A Kantian Account’, in T.L. Beauchamp and R.G. Frey (eds) *The Oxford Handbook of Animal Ethics*. New York, New York: Oxford University Press, pp. 91–118.

Korsgaard, C.M. (2018a) *Fellow Creatures: Our Obligations to the Other Animals*. Oxford: Oxford University Press.

Korsgaard, C.M. (2018b) ‘Rationality’, in L. Gruen (ed.) *Critical Terms for Animal Studies*. Chicago, Illinois: University of Chicago Press, pp. 294–306.

Magnotti, J.F. *et al.* (2015) ‘Superior abstract-concept learning by Clark’s nutcrackers (Nucifraga columbiana)’, *Biology Letters*, 11(5), p. 20150148. Available at: https://doi.org/10.1098/rsbl.2015.0148.

McDowell, J. (1996) *Mind and World*. Harvard University Press.

Moll, H. and Tomasello, M. (2007) ‘Cooperation and human cognition: the Vygotskian intelligence hypothesis’, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480), pp. 639–648. Available at: https://doi.org/10.1098/rstb.2006.2000.

Monsó, S. and Andrews, K. (2022) ‘Animal Moral Psychologies’, in M. Vargas and J.M. Doris (eds) *The Oxford Handbook of Moral Psychology*. Oxford: Oxford University Press, pp. 388–420.

Nelson, E. (2020) ‘What Frege asked Alex the parrot: inferentialism, number concepts, and animal cognition’, *Philosophical Psychology*, 33(2), pp. 206–227. Available at: https://doi.org/10.1080/09515089.2019.1688777.

Newen, A. and Bartels, A. (2007) ‘Animal Minds and the Possession of Concepts’, *Philosophical Psychology*, 20(3), pp. 283–308. Available at: https://doi.org/10.1080/09515080701358096.

Newen, A. and Starzak, T. (2022) ‘How to ascribe beliefs to animals’, *Mind & Language*, 37(1), pp. 3–21. Available at: https://doi.org/10.1111/mila.12302.

Nisbett, R.E. and Wilson, T.D. (1977) ‘Telling more than we can know: Verbal reports on mental processes’, *Psychological Review*, 84(3), pp. 231–259. Available at: https://doi.org/10.1037/0033-295X.84.3.231.

Oberliessen, L. and Kalenscher, T. (2019) ‘Social and Non-social Mechanisms of Inequity Aversion in Non-human Animals’, *Frontiers in Behavioral Neuroscience*, 13. Available at: https://www.frontiersin.org/articles/10.3389/fnbeh.2019.00133 (Accessed: 19 September 2022).

O’Neill, O. (1989) *Constructions of Reason: Explorations of Kant’s Practical Philosophy*. New York, New York: Cambridge University Press.

Papadopoulos, D. (2021) ‘Shared Intentionality in Nonhuman Great Apes: a Normative Model’, *Review of Philosophy and Psychology* [Preprint]. Available at: https://doi.org/10.1007/s13164-021-00594-x.

Pepperberg, I.M. (2021) ‘How do a pink plastic flamingo and a pink plastic elephant differ? Evidence for abstract representations of the relations same-different in a Grey parrot’, *Current Opinion in Behavioral Sciences*, 37, pp. 146–152. Available at: https://doi.org/10.1016/j.cobeha.2020.12.010.

Rosefeldt, T. (2006) ‘Kants Ich als Gegenstand’, *Deutsche Zeitschrift für Philosophie*, 54(2), pp. 277–294. Available at: https://doi.org/10.1524/dzph.2006.54.2.277.

Rosefeldt, T. (2022) ‘Being Realistic about Kant’s idealism’, in K. Schafer and N.F. Stang (eds) *The Sensible and Intelligible Worlds: New Essays on Kant’s Metaphysics and Epistemology*. Oxford: Oxford University Press, pp. 16–44.

Rosenberg, J.F. (1997) ‘Connectionism and Cognition’, in J. Haugeland (ed.) *Mind Design II: Philosophy, Psychology, Artificial Intelligence*. Revised and Enlarged. Cambridge, Massachussets: MIT Press, pp. 293–308.

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

Sellars, W. and Chisholm, R.M. (1958) ‘Appendix: Intentionality and the mental’, in H. Feigl, M. Scriven, and G. Maxwell (eds) *Concepts, Theories, and the Mind-Body Problem*. Minneapolis, Minnesota: University of Minnesota (Minnesota Studies in the Philosophy of Science), pp. 507–539. Available at: http://conservancy.umn.edu/handle/11299/184611 (Accessed: 16 December 2020).

Smirnova, A.A. *et al.* (2021) ‘How do crows and parrots come to spontaneously perceive relations-between-relations?’, *Current Opinion in Behavioral Sciences*, 37, pp. 109–117. Available at: https://doi.org/10.1016/j.cobeha.2020.11.009.

Thompson, R.K.R., Oden, D.L. and Boysen, S.T. (1997) ‘Language-naive chimpanzees (Pan troglodytes) judge relations between relations in a conceptual matching-to-sample task’, *Journal of Experimental Psychology: Animal Behavior Processes*, 23(1), pp. 31–43. Available at: https://doi.org/10.1037/0097-7403.23.1.31.

Thorpe, L. (2018) ‘Kant, Guyer, and Tomasello on the Capacity to Recognize the Humanity of Others’, in *Kant on Freedom and Spontaneity*. Cambridge University Press, pp. 107–136.

Tomonaga, M. (1999) ‘Establishing Functional Classes in a Chimpanzee (pan Troglodytes) with a Two-Item Sequential-Responding Procedure’, *Journal of the Experimental Analysis of Behavior*, 72(1), pp. 57–79. Available at: https://doi.org/10.1901/jeab.1999.72-57.

Tsutsui, K.-I. *et al.* (2016) ‘Representation of Functional Category in the Monkey Prefrontal Cortex and Its Rule-Dependent Use for Behavioral Selection’, *The Journal of Neuroscience*, 36(10), pp. 3038–3048. Available at: https://doi.org/10.1523/JNEUROSCI.2063-15.2016.

Vaughan, W. (1988) ‘Formation of equivalence sets in pigeons’, *Journal of Experimental Psychology: Animal Behavior Processes*, 14(1), pp. 36–42. Available at: https://doi.org/10.1037/0097-7403.14.1.36.

de Waal, F. (2006) *Primates and Philosophers: How Morality Evolved*. Princeton, New Jersey: Princeton University Press.

de Waal, F. (2014) *The Bonobo and the Atheist: In Search of Humanism Among the Primates*. New York, New York: W.W. Norton & Company.

de Waal, F.B.M. (2014) ‘Natural normativity: The “is” and “ought” of animal behavior’, *Behaviour*, 151(2–3), pp. 185–204. Available at: https://doi.org/10.1163/1568539X-00003146.

Walker, R.C.S. (2017) ‘Self and Selves’, in A. Gomes and A. Stephenson (eds) *Kant and the Philosophy of Mind: Perception, Reason, and the Self*. Oxford: Oxford University Press, pp. 204–220.

Wechkin, S., Masserman, J.H. and Terris, W. (1964) ‘Shock to a conspecific as an aversive stimulus’, *Psychonomic Science*, 1(1), pp. 47–48. Available at: https://doi.org/10.3758/BF03342783.