Forthcoming in *Biology and Philosophy*

Animal Morality: What is The Debate About?

Simon Fitzpatrick

sfitzpatrick@jcu.edu

Abstract

Empirical studies of the social lives of non-human primates, cetaceans, and other social animals have prompted scientists and philosophers to debate the question of whether morality and moral cognition exists in non-human animals. Some researchers have argued that morality does exist in several animal species, others that these species may possess various evolutionary building blocks or precursors to morality, but not quite the genuine article, while some have argued that nothing remotely resembling morality can be found in any non-human species. However, these different positions on animal morality generally appear to be motivated more by different conceptions of how the term "morality" is to be defined than on empirical disagreements about animal social behaviour and psychology. After delving deeper into the goals and methodologies of various of the protagonists, I argue that, despite appearances, there are actually two importantly distinct debates over animal morality going on, corresponding to two quite different ways of thinking about what it is to define "morality", "moral cognition", and associated notions. Several apparent skirmishes in the literature are thus cases of researchers simply talking past each other. I then focus on what I take to be the core debate over animal morality, which is concerned with understanding the nature and phylogenetic distribution of morality conceived as a psychological natural kind. I argue that this debate is in fact largely terminological and non-substantive. Finally, I reflect on how this core debate might best be re-framed.

1. Introduction

In recent years, there has been much interest in whether *morality* exists in some non-human animals (henceforth, "animals"), or, put differently, whether some animals possess a *moral psychology*: whether they possess the requisite psychological capacities to engage in some form of moral cognition and action—for instance, make judgments of moral approval or disapproval about others' behaviour, internalize and enforce moral rules or norms, and act for moral reasons (e.g., act punitively towards another individual because of a moral evaluation of that individual's behaviour).

Such questions have been prompted by a burgeoning empirical literature on the remarkably complex and intricate social lives, particularly of our closest primate

relatives, but also of other social mammals like elephants, domestic dogs, wolves, whales, dolphins, and rats, and even some non-mammalian species, such as ravens. For example, chimpanzees appear to engage in third-party policing of behaviour, which seems to indicate the existence and enforcement of norms of conduct within their communities (de Waal, 1996, 2014; Rudolf von Rohr, et al., 2012). Special place is typically accorded to infants, for instance, such that aggression towards them is met with loud protests and active intervention on the part of uninvolved bystanders (Rudolf von Rohr, et al., 2011, 2015). Many other social mammals also appear to enforce various behavioural norms. For instance, many species of primate, along with dogs, wolves, and dolphins, engage in elaborate play rituals and appear to punish individuals that break the rules governing such interactions, such as ostracizing animals that play too aggressively (Flack and de Waal, 2004; Bekoff and Pierce, 2009). There has also been work that purports to indicate other-directed emotional capacities like sympathy and empathy that have long been thought to be important in human moral cognition and motivation (see Bekoff and Pierce, 2009; Andrews and Gruen, 2014 for reviews). In a famous study, rhesus monkeys refused to press a lever to receive food (even in to the point of near starvation), when they discovered that this would result in another monkey receiving an electronic shock (Wechkin et al., 1964). Though this result could be explained in a variety ways (e.g., the monkeys merely avoided doing something that caused an aversive stimulus), a not unreasonable interpretation is that the monkeys recognized and wished to avoid causing distress in others, suggesting some degree of sympathetic concern. Similar pro-social helping behaviours suggestive of empathy and sympathy have been documented in several species, including rats (Bartal et al., 2011; Sato et al., 2015), and chimpanzees, who have been shown to direct consoling behaviours towards losers after fights (de Waal, 1996; Fraser and Aureli, 2008) and display physiological signs of emotional arousal in response to images of violence or other chimpanzees displaying fearful or distressed facial expressions (reviewed by Rudolf von Rohr et al., 2011).

In light of such research, various scientists and philosophers have proposed accounts of the moral capacities of animals, their similarities and differences to those of human beings, and of the evolution of morality more generally.

At most generous end of the spectrum are researchers like Bekoff and Pierce (2009), Rowlands (2012; 2017), Musschenga (2013), Andrews and Gruen (2014), and Monsó (2015), all of whom argue that at least a core subset of the psychological capacities that underlie human morality are far from uniquely human, but are rather things that we share with many social animals. Whatever differences may exist between their moral psychologies and moral systems and those of humans—none of these authors deny that there are such differences—should be seen as differences in the *extent* and *sophistication* of moral capacity. Researchers like de Waal (1996, 2006a; Flack and de Waal, 2000) are rather less generous, however, arguing that what we find in animals, particularly other primates, is *proto*-morality: various psychological "building blocks" or "evolutionary precursors" to morality, but not the fully-fledged article. They argue that while there is important evolutionary continuity here, a crucial evolutionary change occurred uniquely in the human lineage, giving rise to genuine morality (see also Joyce, 2006; Kitcher, 2006, 2011; Rudolf von Rohr et al., 2011; Boehm, 2012; Haidt, 2012; Suddendorf, 2013; Prinz, 2014). At the other end of the spectrum are researchers like Korsgaard (2006) and Ayala (2010), who deny, albeit for different reasons, that anything remotely resembling morality or a moral psychology can be found in animals. Even de Waal's claim that some species possess "building blocks" of morality goes too far, amounting to a comparison between apples and oranges, so different are the capacities of non-humans from what is required to possess the genuine article. These authors thus regard the capacity for moral cognition as representing "a break with our animal past" (Korsgaard, 2006, p104).

One of the interesting things about this debate is that there has been relatively little disagreement about the empirical data. Though much of the relevant research is controversial, largely for methodological reasons, since much of it consists of anecdotal reports of animal behaviour, and because there has been some inconsistency between the results of field and lab-based studies of pro-social

behaviour (see de Waal, 2006a; Bekoff and Pierce, 2009; Rudolf von Rohr et al., 2011; Tomasello, 2016), it is not the data itself that has been the primary focus of this debate. Nor, indeed, has there been much disagreement about what specific psychological capacities can be inferred from this data.[1] Rather, the disagreement has mostly been about the *standard* that "genuine" or "proto"-moral creatures must live up to—not what psychological capacities particular species actually possess, but what capacities they must have in order for us to describe them as having morality or proto-morality. Indeed, even those who generally fall into the same camp on the question of whether animals have morality or a moral psychology endorse different definitions of what it is to have such a thing, or to possess "precursors" or "building blocks" of morality. The question I want to press in this paper is the meta-philosophical one: what counts as getting this standard or definition *right*?

Though, as we will see, it isn't easy to keep descriptive and normative issues apart, to be clear, the debate is ostensibly about what it is to have morality or a moral psychology in the *descriptive* rather than the *normative* senses of "morality" and "moral". Normative definitions of these terms are tied to some account of what are the correct or ideally rational moral beliefs, attitudes, actions, and so forth—this is the sense in which philosophers might talk about the "demands" or "requirements" of morality. Purely descriptive definitions, however, are meant to be independent of such normative claims about what morality requires (Gert, 2016). For instance, a neo-Nazi may be regarded as having a "morality" or a "moral psychology" in the descriptive senses of these terms, insofar as she/he possesses psychological capacities that enable the holding of various beliefs and attitudes about moral issues. However, a neo-Nazi might not be regarded as having a "morality" in the normative sense, insofar as we may want to regard she/he as possessing false or irrational beliefs/attitudes, or as behaving in a morally incorrect

[1] One important area of disagreement concerns the type of empathetic capacity present in various species. This is linked with disagreement about the putative link between the type of empathy taken to be important for morality and mind-reading, and disagreement about the mind-reading capacities of animals (see fn.3 for further discussion). Some researchers have also disputed whether social norms can actually exist in animals with limited mind-reading and social learning capacities (see Andrews, 2009; Tomasello, 2016).

way. Failure to appreciate this distinction has led to some unfortunate episodes in the debate over animal morality. For instance, some researchers have taken the question of whether morality or moral cognition exists in animals to amount to the question of whether they behave in ways that we might regard as morally *praiseworthy* (see, for instance, Jensen et al., 2007 on whether chimpanzees have a sense of fairness). However, as several commentators have pointed out (e.g., Joyce, 2006; de Waal, 2006b; Bekoff and Pierce, 2009), whether or not animals behave in ways that we might judge to be right or good according to a particular normative standard is as irrelevant to the question of whether they are capable of moral cognition or action as the repulsiveness of National Socialism is to the question of whether it constitutes a moral system in the descriptive sense of "moral".

Of course, the meta-philosophical question posed above is not unique to the debate over animal morality. Many of the descriptive accounts of what it is to have a morality or moral psychology that have been offered in this context are inspired by various of the main traditions in moral philosophy (in particular, sentimentalism and Kantianism), each of which can be regarded as offering different definitions of these and other related notions—including what it is to engage in moral reasoning or moral judgment, be a moral agent, and of the primary concerns or subject matter of morality more generally. Indeed, it is an under-appreciated feature of moral philosophy the extent to which these different traditions assume quite different conceptions of the *target* of moral theory. With respect to the bounds or subject matter of morality, there is also currently a vigorous debate in cognitive science concerned with the nature of human moral psychology. Haidt (2012) has argued that much of the field has adopted what he refers to as a "liberal" conception of the moral domain, focused on issues of harm and fairness, ignoring more "conservative" concerns, such as purity, respect, and group-loyalty, meaning that many important aspects of human moral psychology have largely gone unstudied. Haidt argues that this is partly due to the influence of the work of Turiel and colleagues (Turiel, 1983), who have offered a psychological account of the putative difference between genuine "moral" judgments and so-called "conventional" normative judgments (e.g., normative judgments about matters of etiquette and taste), according to which

moral judgments concern issues of harm and fairness and display a characteristic psychological profile quite different to that of conventional normative judgments—for instance, they are typically regarded as universal, authority-independent, more serious, and tend to be justified by appeal to notions of harm, rights, and justice. This account of the moral/conventional distinction has come in for much criticism, including from Haidt, who has argued that it illegitimately places many "conservative" concerns outside of the moral domain. In each of these instances, the assumption is that there is a *correct* account of the relevant concepts (*morality*, *moral domain*, *moral judgment, moral norm, moral agency*, etc.) to be had. This clearly gives rise to the question of how we are tell when we have in fact locked on to the correct account of any of these notions.

My inspiration for asking this meta-philosophical question comes from Stich and colleagues (Nado et al., 2009; Stich, 2009), who have pressed it in relation to much recent work in human moral psychology, particularly the Turiel tradition and the putative moral/conventional distinction. They regard Turiel and colleagues as attempting to articulate moral judgment as a psychological *natural kind* (defined by the characteristic subject matter and psychological profile described above). This contrasts with the standard approach of philosophers towards defining such notions, which typically involves some form of *conceptual analysis*. Ultimately, Stich and colleagues argue on the basis of some empirical work (e.g., Kelly et al., 2007; Fessler et al., 2015) that the Turiel account fails to pick out a really existing natural kind and that there is no good reason to believe in the existence of a psychologically distinct sub-class of normative judgments that we can regard as genuinely "moral" as opposed to merely "conventional".

After describing the main contours of the current debate over animal morality (Section 2), I will utilize Stich and colleagues' distinction between conceptual analysis and natural kind approaches to defining "morality" and argue that we can find representatives of both types of approach in the current literature (Section 3). After delving deeper into the goals and methodologies of these two approaches, we will see that, despite appearances, there are actually *two* importantly distinct debates over animal morality going on. This, of course, implies

that several of the apparent skirmishes in the current literature are actually cases of researchers simply talking past each other. I will then focus on what I take to be the core debate that has been going on, which is concerned with understanding the nature and phylogenetic distribution of morality conceived as a psychological natural kind (Section 4). I will argue that this debate is in fact largely terminological and non-substantive. Finally, I will reflect on how this core debate might best be re-framed (Section 5). I will argue in favour of a more fine-grained approach that asks not whether animals possess a "moral" or "proto-moral" psychology, but whether they possess certain more tightly defined psychological mechanisms.

2. Moral animals?

In their book, *Wild Justice*, Bekoff and Pierce define "morality" as:

> …a suite of inter-related behaviours that cultivate and regulate the complex interactions within social groups. These behaviours relate to well-being and harm. And norms of right and wrong attach to many of them. (2009, p7)

Bekoff and Pierce adopt the view—common to many evolutionary theories of morality—that morality evolved to facilitate and improve levels of co-operation in the small-scale communities that our ancestors lived. The idea is that codes of conduct that regulate individual behaviour, inhibit selfishness, discourage free riding, reduce intra-group violence, and increase group cohesiveness make co-operative endeavours easier and more effective, and were thus likely adaptive for our ancestors, who depended on co-operation with others for survival and successful reproduction. Similar fitness benefits may have accrued from them having a basic level of concern for the interests of others in their group.[2] However, Bekoff and Pierce see no reason to think that morality evolved only recently in the human lineage, since the ancestors of many other animals plausibly also lived in rich social ecologies that involved co-operative endeavours like hunting, defence against

[2] Though they do appear open to the possibility of group selection playing a role in the evolution of some aspects of morality, as they define it, Bekoff and Pierce lean towards the view that the evolution of mechanisms that produce pro-social behaviours can be explained without necessarily having to invoke selection at the level of groups (see also, Joyce, 2006; de Waal, 2006a).

predators, care for infants, grooming, play, and so forth, and thus plausibly also needed the "social glue" that morality is taken to provide.

Bekoff and Pierce (2009, p8) argue that the empirical evidence for morality in animals comes in three clusters: the *co-operation* cluster, which includes putative instances of "altruism, reciprocity, trust, punishment and revenge" in many species; the *empathy* cluster, which includes various other-directed behaviours suggestive of "sympathy, compassion, caring, helping, grieving, and consoling"; and the *justice* cluster, which includes behaviours suggestive of "a sense of fair play, sharing, a desire for equity, expectations about what one deserves and how one ought to be treated, indignation, retribution, and spite". Their claim is thus that many social animals possess a variety of psychological capacities—including other-directed emotional capacities like sympathy and empathy,[3] pro-social and altruistic motivation, and a primitive sense of "right and wrong" tied to various social norms[4]

[3] Though the terms "sympathy" and "empathy" are sometimes used interchangeably, Bekoff and Pierce recognize a distinction between empathy as a type of emotional mimicry (feeling *what* another is feeling) and sympathy as having an emotion on behalf on another (feeling *for* the other) (see also Prinz, 2011). They also take empathy to come in various degrees of complexity, ranging from low-level *emotional contagion*, where an emotion is triggered in an individual as result of merely observing a behavioural cue from another (such as a distressed or fearful facial expression), to *cognitive empathy*, where the individual is able to fully adopt the emotional perspective of another and understand the reasons for it (e.g., understanding that another individual is fearful and what has caused this). The latter requires a rich mind-reading capacity, while lower levels of empathy needn't require any ability to represent others' mental states. Sympathy is similarly taken to come in varying degrees of complexity, reflecting the extent to which individuals are able to put themselves in another's situation.

Following de Waal (2006a), Bekoff and Pierce regard cognitive empathy as the type of empathy most relevant to morality, since it involves genuine recognition and understanding of another's emotional state, and are willing to attribute full-blown cognitive empathy to several species (de Waal restricts this capacity to apes). Others are much more sceptical about cognitive empathy in animals, largely because of doubts about their mind-reading capacities. Andrews and Gruen (2014; see also Gruen, 2015) provide an account of empathy and its putative connection with morality in apes that tries to carve some space between emotional contagion and full cognitive empathy. Monsó (2015) argues that even emotional contagion can be viewed in moral terms; hence, the debate over animal morality can be fully separated from the debate over animal mind-reading.

[4] Bekoff and Pierce take these social norms to exist in the form of implicit expectations about appropriate and inappropriate behaviour: animals respond to norm violating behaviour with protests (e.g., "waa" barks in chimpanzees), or with punitive behaviours of their own (e.g., refusing to play with animals that have played too roughly), but needn't have any conscious or reflective understanding of the relevant norm itself. Much of human thinking about social norms has been claimed to be like this (e.g., Nichols, 2004; Sripada and Stich, 2006; Haidt, 2012). In many, perhaps most, cases, human social norms are unconsciously internalized early in development, and all the individual typically has conscious access to are the agonistic emotional states (like anger) that

that exist within their communities—that make them worthy of being regarded as moral beings, insofar as these psychological traits are plausibly homologues or analogues to those that underlie central aspects of human morality.[5] This is not to deny that human morality and moral cognition and chimp or wolf morality and moral cognition are different in important ways. For instance, they claim that the content of morality is importantly "species-relative", so the moral norms of chimp communities are likely quite different to those of wolf or human communities, and that different species may have more sophisticated moral capacities than others. But, at a general level, the capacity to *possess* morality is something that we share with many other mammals, including, they argue, "bonobos, chimpanzees, elephants, wolves, hyenas, dolphins, whales, and rats", and potentially even with some non-mammalian social animals like ravens (2009, p83).

Bekoff and Pierce build much of their account on the work of de Waal, a key pioneer of the contemporary study of the rich emotional and social lives of non-human primates. However, de Waal himself isn't prepared to go as far as Bekoff and Pierce. Instead, de Waal (1996, 2006a; Flack and de Waal, 2000) sees himself as modernizing the position of Darwin in *The Descent of Man*:

---

accompany their observing norm violating behaviour and the intrinsic motivation to punish norm violators.

[5] Though Bekoff and Pierce tend to talk about "patterns" and "clusters" of "moral behaviours", their focus is really on the internal psychological *mechanisms* that drive these behaviours. It is the possession of these mechanisms that make animals moral beings, on their view, not the behaviours *per se* (Musshenga, 2013). For instance, they emphasize the following "threshold requirements" for being a moral animal:

> [A] level of complexity in social organization, including established norms of behaviour to which attach strong emotional and cognitive cues about right and wrong; a certain level of neural complexity that serves as a foundation for moral emotions and for decision making based on perceptions about the past and the future; relatively advanced cognitive capacities (a good memory, for example); and a high level of behavioural flexibility (2009, p83).

Moreover, when discussing instances of pro-social and altruistic behaviour, they emphasize that merely acting to help another individual at cost to oneself is insufficient for the behaviour count as moral behaviour. What matters is the underlying *motivation*—i.e., whether the behaviour is the product of a desire to help that is itself other-regarding. Hence, when they talk about altruism as an instance of moral behaviour, what they mean is *psychological* altruism, not just so-called "biological" altruism, which is defined exclusively in terms of reproductive fitness, without reference to underlying motivation.

> Any animal whatever, endowed with well-marked social instincts, the parental and filial affections being here included, would inevitably acquire a moral sense or conscience, as soon as its mental powers had become as well, or nearly as well developed, as in man (Darwin, 1871, p68-69).

Darwin was sympathetic to the sentimentalist tradition of moral philosophers like David Hume and Adam Smith that rooted human moral cognition in sentiment, particularly our ability to empathize with others, and argued that our moral sentiments should be seen as an outgrowth of the pro-social instincts and emotional capacities of our non-human ancestors, which we share with many other species. In a similar vein, de Waal points to what he regards as the evolutionarily ancient "building blocks" of moral cognition—sympathy and empathy towards others, pro-social and altruistic motivation, and what he calls a primitive "sense of fairness" tied to social norms—which we share with other primates (apes, in particular). However, like Darwin, de Waal argues that there is a key difference between human morality and the sentiments and social norms of other animals. Darwin largely adopted the view of Hume and Smith that the possession of a true moral sense or "conscience" required not just capacities for empathy and sympathy, or the capacity to make judgments about others' behaviour, but also a special type of reflective capacity. For Hume, this was the ability to "perceiv[e] the duties and obligations of morality" (Hume, 1978, p468), and to abstract away from one's own situation to make judgments from a position of impartiality. For Darwin, it was the ability to self-consciously reflect on one's actions and motives, "and of approving or disapproving of them" (1871, p85). It was this capacity for critical self-reflection, which came with the evolution of increased "mental powers" in humans (Rowlands, 2012).[6] De Waal doesn't explicitly locate the difference in such a capacity for self-reflection, but rather in the scope and explicitness of human moral codes:

[6] In this respect, Darwin seems to have viewed the human moral sense as a by-product of the evolution of sophisticated reasoning capacities, rather than a specific psychological adaptation in its own right (Ayala, 2010). He also suggested that the development of human moral norms (i.e., the content of specific moral belief systems, rather than the psychological mechanisms that underlie the capacity to have such systems) was shaped by a process of cultural group selection:

> Instead of merely ameliorating relations around us, as apes do, we have explicit teachings about the value of the community and the precedence it takes, or ought to take, over individual interests. Humans go much further in all of this than the apes […] which is why we have moral systems and they do not. (2006a, p54)

The sentiments and social norms of non-human primates are too local and specific to interactions between individuals to count as being genuinely *moral*. De Waal argues that this widening of concern to the community as a whole, giving rise to genuine moral belief systems in the human lineage, was partly the product of the evolution of warfare. As communities became larger and engaged in greater and more deadly inter-group conflict, the harmony and cohesiveness of the community became even more important, leading to more explicit and more general rules governing behaviour. De Waal also places emphasis on the evolution of human language as a tool for regimenting and transmitting genuine moral norms and judgments. Albeit with some important differences in detail, Joyce (2006), Kitcher (2006, 2011), Rudolf von Rohr et al. (2011), Boehm (2012), Haidt (2012), and Suddendorf (2013), among others, have made similar claims to de Waal about the continuity yet distinctiveness of human morality and animal proto-morality.

In a widely cited commentary on de Waal's claims about the "building blocks" of morality being present in other species, Korsgaard (2006) argues that true moral cognition requires:

> [N]ormative self-government... a certain form of self-consciousness: namely, consciousness of the grounds on which you act *as grounds*... you have a certain reflective distance from the motive, and you are in a position to ask

---

> It must not be forgotten that although a high standard of morality gives but a slight or no advantage to each man and his children over the other men of the same tribe, yet that an advancement in the standard of morality and an increase in the number of well-endowed men will certainly give an immense advantage to one tribe over another (1871, p159).

Modern theorists of the evolution of morality disagree about the extent to which human moral capacities are themselves psychological adaptations or by-products of adaptations for other functions, and about the extent to which the specific content of human moral codes and judgments have been shaped by genetic rather purely cultural evolution (for a survey, see Machery and Mallon, 2010).

> yourself "but should I be moved in that way? Wanting that end inclines me to do that act, but does it really give me a reason to do that act? (2006, p113)

This might sound like Darwin's point about the ability to reflect on one's actions and motivations being distinctive of human moral psychology. However, in contrast to Darwin and de Waal, Korsgaard does not see this core feature of moral cognition as a part of a continuum that includes the proto-moral capacities of animals, but rather as representing a fundamental *discontinuity* in nature. It is thus a mistake to regard animals as even proto-moral beings. The reason for this is that Korsgaard adopts a largely Kantian conception of moral psychology, centred not on sentiment, but upon a capacity for rational deliberation about the normative justification for one's actions and judgments ("*ought* I perform this action?"; "is this the judgment that I *should* make in this situation?"). Though one can, in a loose sense, regard de Waal's "building blocks" as precursors to human morality, in so far as they were in place in our ancestors before they became moral beings, there is no sense in which these capacities can be regarded as continuous with the reflective capacity that constitutes the special ingredient in human moral psychology:

> [I]t is the proper use of this capacity—the ability to form and act on judgments of what we ought to do—that the essence of morality lies, not in altruism or the pursuit of the greater good. So I do not agree with de Waal… The difference here is not a mere matter of degree. (Korsgaard, 2006, p116-7).

Other advocates of discontinuity include Ayala (2010), who argues that the capacity for genuine moral judgment and agency requires very sophisticated reasoning capacities that, he argues, are plausibly absent in non-humans. These include: i) the ability to anticipate the consequences of one's actions for others, which requires the ability "to anticipate the future and to form mental images of realities not present or not yet in existence";[7] ii) the ability "to perceive certain

[7] There is strong evidence that many animals are capable of anticipating the future and predicting the likely outcomes of their actions (e.g., Clayton and Dickinson, 1998; Martin-Ordas et al., 2010). However, Ayala seems to have something more sophisticated than mere causal reasoning and anticipation in mind—something more like what is often referred to as "mental time travel", which

objects or deeds as more desirable than others", which requires a capacity for highly abstract thought; and iii) the ability to make reflective choices between different courses of action (2010, p9018-9019). Because he regards each of these capacities as necessary conditions for genuine moral cognition and agency, Ayala thus also denies any form of "incipient" morality in animals.

Bekoff and Pierce respond to those more sceptical about animal morality by acknowledging that there *are* significant cognitive differences between humans and animals. For instance, they accept de Waal's claim that only humans are able to explicitly formulate and teach moral norms via language, and Korsgaard's claim that animals likely lack the rich meta-cognitive capacities required for normative self-government. However, they argue that these are differences *within* the moral domain, not *between* moral humans and non-moral (or proto-moral) animals:

> We view each of these possibly unique capacities (language, judgment) as outer layers of the Russian doll, relatively late evolutionary additions to the suite of moral behaviours. And although each of these capacities may make human morality unique, they are all grounded in a much deeper, broader, and evolutionary more ancient layer of moral behaviours that we share with other animals. (2009, p141)

Similarly, Andrews and Gruen (2014) criticize the tendency of philosophers like Korsgaard to focus on "the most rarefied and linguistically mediated" aspects of human moral cognition and behaviour:

> Once we are able to look past the most salient examples of human morality, we find that moral behaviour and thought is a thread that runs through our daily activities, from the micro-ethics involved in coordinating daily behaviours like driving a car down a crowded street […] to the sharing of someone's joy in getting a new job or a paper published. If we ignore these sorts of moral actions, we are overintellectualizing human morality (2014, p194).

---

involves the ability to mentally project *oneself* backward or forward in time, and is widely held to be uniquely human, largely because it is thought to require a particularly rich form of self-consciousness (e.g., Suddendorf, 2013; though see Clayton and Dickson, 2010). The type of mental time travel he regards to be most important for morality also involves being able to project oneself into someone else's situation in time—for instance, being able to anticipate what their emotional state would be.

Such a less intellectualized conception of what it is to think and behave morally—for instance, Gruen's (2015) own "entangled empathy" account, which requires just that one has some understanding of another's situation and needs, and how to respond to their situation—is much friendlier to including the capacities for sympathy and empathy and the tight social bonds and relationships of apes, in particular, inside the moral domain.[8]

Another important contribution to this debate comes from Rowlands (2012). He criticizes Bekoff and Pierce for offering too expansive a definition of what it is to have a moral psychology, including the capacities underlying various helping behaviours and social norms, which need not, he argues, be seen in moral terms. Indeed, he suggests that they have essentially defined "morality" so broadly as to make the question of whether animals can be moral beings *uninteresting*. Instead, Rowlands argues that the real question is whether animals are capable of acting for moral reasons. He adopts a largely sentimentalist account of moral motivation, according to which one can act morally if one is moved by certain emotional states, such as compassion at the plight of an other, which may incline one to act so as to alleviate their suffering, or indignation at another's actions, which may incline one to behave punitively towards them.[9] He adopts an externalist theory of moral content, according to which particular emotional states represent moral properties if they bear appropriate causal relations to them. Crucially, Rowlands argues, a creature need not be aware of these relations in order for these emotional states to have moral content and constitute moral reasons for action.

Against those Kantians, Aristotelians, and sentimentalists (Rowlands includes Hume and Darwin here) that have claimed that genuine moral motivation

[8] Andrews and Gruen (2014) argue that this recognition and concern for others needn't require particularly rich mind-reading capacity. Hence, cognitive empathy needn't be necessary for *moral* empathy.

[9] Rowlands does not regard this as the *only* route to moral action. Hence, he departs from a strict sentimentalism by allowing for the possibility of moral action being produced by "cold" reasoning processes, without affective states having to play a necessary role. However, he thinks that such cognitive forms of moral motivation are probably unique to humans.

requires that one at least sometimes be consciously aware of one's reasons and be able to reflect on their normative force, Rowlands points to the case of Myshkin, a character based on the prince from Dostoyevsky's *The Idiot*. Myshkin experiences what seems like compassion for others and is thus compelled to act in ways that we would ordinarily regard as kind or compassionate, yet lacks the capacity to subject these feelings and actions to critical scrutiny. He is unable to consciously recognize these emotional states *as reasons* for action and unable to think about whether they are the *correct* ones to have in the circumstances. Since he lacks these capacities, Rowlands argues that Myshkin cannot be morally evaluated (praised or blamed) or held morally responsible for his actions, and thus should not be regarded as a fully-fledged *moral agent*. However, it is plausible to regard him as a *moral subject*, since he is surely motivated by emotions that track moral considerations—his feelings of compassion are caused by others' suffering, and he clearly acts in order to alleviate this suffering. Thus, Rowlands argues that Myshkin possesses a genuine moral psychology, but one that operates on a more "visceral" level than that of full moral agents. Even if he lacks the reflective capacity for full moral agency, he can still act for moral reasons. In so doing, Rowlands tries to diffuse various traditional philosophical arguments for denying that Myshkin's motivations can be genuinely *moral*. The result is that social animals that also lack these reflective capacities, but, like Myshkin, are capable of possessing other-directed emotional states (like sympathetic distress) that track moral considerations and play a causal role in their behaviour, may also be regarded as moral subjects.[10] Of course, the question of which species actually satisfy these conditions for moral subjecthood is, Rowlands emphasises, an empirical one, but he sees the work on animal emotion cited by Bekoff and Pierce and de Waal as providing at least a *prima facie* case for the existence of non-human moral subjects.[11]

---

[10] Monsó (2015) points out that Rowlands' externalist account of what it is to track moral considerations allows that animals that lack the capacity for full cognitive empathy may still possess, and be motivated by, moral emotions. Even emotions produced by emotional contagion can count as moral.

[11] For their part, Bekoff and Pierce (2009, p144-145) express scepticism about the traditional philosophical concept of moral agency and argue that its application to animals is "likely to promote

3. Two debates

The debate over animal morality has featured relatively little empirical disagreement about what animals and humans are able to do or not do, to the extent that one can, as I have, lay out the contours of the debate while saying little about the empirical data itself. Rather, the dispute has mostly been about where to draw the boundary of morality and moral cognition. However, though there has been much debate about different definitions of what it is to possess morality or a moral psychology, there has been next to no explicit discussion of what standard of correctness should be used for evaluating these rival definitions. Yet, the debate appears to make *no sense* unless we assume that there is such a standard. The protagonists clearly do not see themselves as merely offering stipulative accounts of morality* or morality†. The debate is taken to be substantive and *not* purely terminological.

One difficulty is that theorists can and have offered definitions of "morality" and "moral" at all sorts of different levels: in terms of behaviour, in terms of content (e.g., the characteristic subject matter of moral norms or judgments), in terms of form or character (e.g., the logical or psychological structure of moral judgments or moral norms), in terms of underlying processes or capacities (e.g., the possession of certain types of emotion or reasoning processes), and so forth. But, setting that complication aside for the moment, what, in general, is it to define "morality" or "moral"? As we will see, consideration of this question leads to the conclusion that there are actually *two* quite different debates over animal morality, which need to be carefully distinguished.

3.1 Conceptual analysis vs. natural kind approaches to defining "morality"

---

philosophical confusion and should ultimately be avoided". However, they do suggest that animal behaviour can be morally evaluated within the context of animal communities, such that the behaviour of a wolf towards a fellow wolf is morally evaluable, but "predatory behaviour of a wolf towards an elk is *amoral*".

Taking their cue from Taylor (1978), Stich and colleagues (Nado et al., 2009; Stich, 2009) distinguish between two different types of approach to defining "morality", which employ different criteria for what it is to get the definition right: *conceptual analysis* and *natural kind* approaches.[12]

Though conceptual analysis can take many forms, Stich and colleagues focus on a common version that is employed in many areas of contemporary philosophy. One begins with a common sense understanding of the concept to be analysed—for instance, the concept, *moral judgment*. This might, for instance, be based on certain commonly recognized instances or distinctions (e.g., that there is a distinction between judgments about canonical moral issues and judgments about matters of taste or etiquette). A philosophical analysis of the concept is then proposed, typically involving a set of necessary and sufficient conditions for application of the concept. This analysis is then tested against intuitions about whether or not the concept does actually apply in various actual or hypothetical cases. The analysis is taken to stand or fall depending on how well it matches up with these intuitions over a wide range of cases.[13]

While empirical data may be relevant to the conceptual analysis approach, in so far as it elicits intuitions about the application of the concept under analysis and can thus be used to test the adequacy of proposed analyses, the natural kind approach is, according to Stich and colleagues, much more of an empirical project. Natural kinds are real categories of thing that exist in nature, independently of our beliefs about them, and which can support inductive generalizations (Bird and Tobin, 2015). Here, one may begin with an intuitive or theoretically motivated conception of what, say, a moral judgment is. Like the conceptual analysis approach,

---

[12] Stich and colleagues also talk about a third type of approach: Oxford-style *linguistic analysis*. This would involve studying how people use moral terms in ordinary language. I won't discuss that sort of approach here, since I don't think that any of the protagonists to the debate over animal morality would see themselves as engaging in such a project.

[13] The introduction to Wallace and Walker (1970; cited by Stich and colleagues) provides a nice summary of various conceptual analyses of *moral rule* that can be extracted from the literature in moral philosophy and the problems that they face. Stich is a longstanding critic of conceptual analysis in philosophy, and thus Stich (2009) expresses much scepticism about this approach to defining "morality".

this might involve pointing to some intuitively clear instances and non-instances of the kind. However, one then conducts empirical investigation into the properties of the clear instances, seeking to articulate certain *essential* or typically *co-occurring* properties that individuate the kind, and which can then play a role in explaining the various generalizations that can be made about its instances. Crucially, this may lead one to revise the starting conception in various ways—for instance, deciding that what one might have previously regarded as an instance of the kind is really not (e.g., in the case of the natural kind, *water*, an articulation of the kind in terms of an essential property—possession of the chemical structure, H2O—implies that certain clear, odourless liquids are not really instances of water), or deciding that certain putative essential properties are not really essential to the kind (e.g., modern biology forces us to abandon the vitalist claim that living organisms are distinguished from non-living things by the possession of some intrinsic non-physical property). In this respect, the standard of correctness is empirical: how well does the proposed articulation of the kind match up with what we find in nature? Moreover, it is nature that ultimately has the last say about whether the putative kind is an actually existing natural kind at all (consider *phlogiston* and *caloric*), and the properties that something must possess in order to be an instance of the kind.

Stich and colleagues view the Turiel account of the moral/conventional distinction as an example of this latter approach (see also Kumar, 2015). Turiel and colleagues have marshalled an impressive amount of cross-cultural and developmental evidence, which is claimed to show that neuro-typical humans (both pan-culturally and at a fairly early age) respond to violations of prototypical moral norms quite differently to violations of prototypical conventional norms. This characteristic psychological profile (universality, authority-independence, greater seriousness) is also claimed to go along with a characteristic subject matter—issues of harm and fairness—which play a distinctive role in the justifications that people tend to offer for their judgments. These psychological properties can therefore be read, Stich and colleagues suggest, as supposedly constituting a *nomological cluster*—a set of typically co-occurring properties that are meant to individuate a

distinct psychological kind: *moral judgment*.[14] Hence, though Turiel and colleagues were initially inspired by the philosophical literature on moral judgment, rather than attempting to specify the content of the *concept* of moral judgment, they are best interpreted as doing something more akin to what physicists sought to do in providing an empirical account of the nature of heat. Crucially, as is the case with heat, the ultimate outcome of such an approach may bear little relationship to prior folk concepts or armchair philosophical analyses. For instance, on Turiel and colleagues' account, a judgment that has the psychological profile of a conventional judgment wouldn't count as a moral judgment, irrespective of what intuition or one's favoured philosophical account of moral judgment might say.

As we will see, what I want to call the *conceptual approach* to animal morality is rather more complex than the picture that Stich and colleagues paint of traditional conceptual analysis in philosophy—in particular, the methodology isn't just one of testing proposed analyses against intuitions about the application of the relevant concept. Nonetheless, Stich and colleagues' distinction does help us to isolate two quite different approaches to the question of whether morality exists in animals.

3.2 Conceptual vs. natural kind approaches to animal morality

Though it is also possible to read his account of moral motivation as describing a natural kind (see Section 4), Rowlands is quite explicit that he sees his project as one of "conceptual analysis and clarification" (2012, p33), and though there is more to his case than just an appeal to intuition, much of his discussion of Myshkin and

[14] Stich and colleagues suggest that, if Turiel and colleagues are right, then moral judgments would constitute something like a *homeostatic property cluster* (HPC) kind (Boyd, 1999). HPC kinds are individuated by clusters of typically co-occurring properties, where this clustering can be explained in terms of a shared underlying casual (homeostatic) mechanism. In this instance, the homeostatic mechanism would presumably be the particular psychological processes that underlie moral as opposed to conventional judgments. Crucially, unlike on classical essentialist accounts of natural kinds, members of HPC kinds needn't share sets of properties that are both necessary and sufficient for kind membership, which is why the HPC account has become popular as an account of biological and psychological kinds, which tend to exhibit significant internal variability, but nonetheless display stable clusterings of properties—in virtue, for instance, in the case of biological species, of a shared evolutionary history.

the distinction between moral subjects and moral agents clearly fits with the traditional methodology of conceptual analysis that Stich and colleagues describe.

Rowlands' main claim is that there is a distinction to be drawn between the concepts of moral motivation and moral responsibility (or evaluability). His preliminary argument for this distinction is that analyses that equate the two via some form of reflection condition don't match up with what intuition seems to tell us about Myshkin. Intuitively, Myshkin does act for moral reasons, even though he lacks the capacity for critical moral reflection on his motivations and actions. However, because he lacks this reflective capacity, intuitively, Myshkin ought not be regarded as worthy of praise or blame. Hence, we have a reason to at least entertain the possibility of a distinction between moral subjects (who act for moral reasons, but need not be morally responsible) and full-blown moral agents (who act for moral reasons *and* are morally responsible/evaluable for their actions).[15] This opens the door for animals without rich reflective to potentially act for moral reasons.

I want to stress that this isn't the only argument that Rowlands gives for the claim that creatures without rich reflective capacities can act for moral reasons. However, at least at the outset, much hangs on our intuitions about whether the concepts of moral motivation and moral responsibility apply to Myshkin. In so far as intuition suggests that the former but not the latter apply, that provides preliminary support for Rowlands' externalist analysis of what it is to act for moral reasons. Most of the rest of the book is concerned with developing this analysis and rebutting various Kantian and Aristotelian arguments for resisting the intuition that Myshkin's motivations are genuinely *moral*. The use of the Myshkin case also assumes that sceptics about animal morality, like Korsgaard, are engaged in the same project, and Rowlands (2017) suggests such sceptics are often inclined towards invoking reflection conditions in analyses of moral motivation because of

[15] Rowlands (2017) makes the same sort of argument in the case of the notorious real life 10-year-old killers of Jamie Bulger. Intuition suggests that 10-year olds lack full moral responsibility, but also that these boys were motivated by (bad) moral reasons—for instance, they reported planning on killing a child that day.

the intuition that non-human animals cannot be moral beings—for instance, that they are simply prisoners of their desires and that such creatures cannot act for truly moral reasons. The Myshkin case and the arguments he builds around it are thus designed to trump that intuition and undermine Korsgaard's claim that normative self-government is necessary for a creature to be able to act for truly moral reasons. In addition, Rowlands' criticism of Bekoff and Pierce also seems to presuppose this type of approach. His claim is essentially that their expansive definition of what it is to be a moral creature misses the core component of the concept—acting for moral reasons. He takes this to be demonstrated by various examples of behaviour sufficient for a creature to be regarded as a moral being under their account, but which are not intuitively *moral* behaviours at all (see Rowlands, 2012, p25-32).

However, Bekoff and Pierce do not actually seem to be engaged in that type of classic philosophical project. They don't seem to be interested in our *concepts* of moral motivation or moral agency. Rather, they explicitly regard morality as a *biological* phenomenon, and seem to reject the idea that it is something that can be defined at a conceptual level.[16] Hence, the definition they offer of "morality" isn't, I suggest, meant to be an articulation of the concept of morality. Rather, it is meant to pick out a natural kind. Like Turiel and colleagues on moral judgment, Bekoff and Pierce seem to see content as important, claiming that moral behaviours concern "harm and well-being". However, their principal focus is on the evolutionary *function* of morality. They take morality *qua* natural kind to be a cluster of psychological capacities (and associated behaviours) that can be grouped together according to a common proper function, which is to facilitate and increase levels of co-operation within groups of social animals. At least in its broad outlines, this evolutionary-functional account is common to many theories of the evolution of

[16] I don't mean to imply that Rowlands thinks that the capacities that underlie what it is to be a moral being cannot be understood in biological terms. He does hold these capacities to be a product of evolution by natural selection. Korsgaard appears similarly open to evolutionary explanations for normative self-government. The issue is about how we are to determine which evolved capacities are *moral* capacities.

morality (e.g., Sober and Wilson, 1997; Joyce, 2006; Kitcher, 2011; Haidt, 2012; Greene, 2013), which see it, in one way or another, as a psychological adaptation for group living. The difference is that Bekoff and Pierce are more expansive in terms of what capacities they claim to share this common evolutionary function, and are thus more liberal about what types of psychology a creature can possess in order to be classified as having a moral psychology.[17]

In his arguments against what he refers to as the *veneer theory*, which regards morality as an artificial overlay on our biological nature, de Waal (2006a) also regards morality as a biological phenomenon to be understood in functional terms. He adopts much the same position as Bekoff and Pierce when it comes to grouping together capacities that he regards to be "building-blocks" of morality—other-directed emotions, pro-social and altruistic motivation, primitive social norms. Each of these represent psychological adaptations to the demands of social life, sharing broadly the same function of improving levels of co-operation and group cohesiveness. However, de Waal restricts the natural kind, *morality*, to human beings by including only those capacities that allow the scope of social norms to be widened to the community as a whole and for them to be externalized in language, the idea being that these facilitate even better levels of co-operation and group cohesiveness than one finds in other primates. He thus regards morality to be a specific adaptation to greater co-operative demands placed on our hominin ancestors. The result is that we have two related but distinct natural psychological kinds: the cluster of *proto-moral* capacities shared with other primates (and potentially other social mammals), which constitute (in Bekoff and Pierce's words) a social glue, and the cluster of *moral* capacities unique to humans, which constitute a *better* social glue.

[17] It is worth noting that this appeal to function requires a different conception of natural kinds to the HPC account that Stich and colleagues appeal to when making sense of the claims of Turiel and colleagues, since that account has difficultly in accommodating function (rather than clustering of properties in virtue of an underlying causal mechanism) as a criterion for kind membership (Ereshefsky and Reydon, 2015). Other theories of natural kinds are friendlier to such functional kinds (e.g., Ereshefsky and Reydon, 2015; Slater, 2015).

It seems, then, that Rowlands and Korsgaard are actually engaged in quite a different debate to Bekoff and Pierce and de Waal. The former are engaged in a disagreement over whether particular moral *concepts* can be applied to the psychological states and behaviours of various animals. The issue is about correctly categorizing animal behaviour and psychology either as falling inside or outside the moral domain, but drawing the boundaries of this domain is a matter of investigating the content of these concepts. The latter group of researchers, however—and here, I think, we can also include most of the other authors cited in Section 2—seem to be pursuing quite a different project: understanding the nature and distribution of morality as a biological phenomenon—a psychological *natural kind*. This is an empirical rather than conceptual project, and may, like investigations into other natural kinds, lead to definitions of "morality" that depart significantly from the content of pre-existing folk concepts or philosophical analyses. As Bekoff and Pierce put it:

> We want to detach the word *morality* from some of its moorings, allowing us to rethink what it is in light of a huge pile of research from various fields that speaks to the phenomenon. (2009, p12).

In this respect, the two groups have largely been talking past each other. For instance, when Korsgaard says that de Waal's "building blocks" have little to do with morality because the "essence of morality" is normative self-government, "not altruism or the pursuit of the greater good" (2006, p116), she is making what she takes to be a conceptual point—it is basically a category mistake to regard instances of altruism or sympathetic concern as having anything to do with morality. However, the capacities de Waal cites aren't meant to be building blocks of human morality in a conceptual sense, but in an evolutionary one, and they are to be regarded as continuous with the capacities that underlie human moral psychology because of their functional, not conceptual, similarity.

In response to Korsgaard, de Waal says something that is quite illuminating in this context:

> To neglect the common ground with other primates, and to deny the evolutionary roots of human morality, would be like arriving at the top of a tower to declare that the rest of the building is irrelevant, that the precious concept of "tower" ought to be reserved for its summit. While making for good academic fights, semantics are mostly a waste of time. Are animals moral? Let us simply conclude that they occupy several floors of the tower of morality. Rejection of even this modest proposal can only result in an impoverished view of the structure as a whole. (2006b, p181).

Here, de Waal seems to recognize that Korsgaard's project is more of a conceptual one than his, which is to understand the evolutionary roots of morality conceived as a natural kind. Yet, his response to her simply misses the point: for Korsgaard, de Waal's "tower of morality" isn't a tower of *morality* at all, and to regard the other floors below the summit as having anything to do with morality in her sense of the term is to change the subject. Rowlands makes a similar observation about Bekoff and Pierce's (2009, p140-1) earlier quoted "Russian doll" response to Korsgaard:

> [T]heir response seems curiously off-target… Korsgaard claims that the ability to reflect on or form judgments about what we ought to do is the *essence* of morality. Any behaviour that is not subject to this sort of normative self-reflection is not moral behaviour… appearances notwithstanding. (Rowlands, 2012, p111)

However, Rowlands himself also seems to misunderstand what Bekoff and Pierce and de Waal are doing: from their perspective it makes perfect sense to talk of a "tower" or "Russian doll" of morality if "morality" is defined in evolutionary-functional terms. These researchers are thus clearly talking at cross-purposes.

3.3 "Morality" and normativity

So far, I have tried to show that these two groups of researchers, who are ostensibly engaged in the same debate over whether animals possess morality or a moral psychology, actually adopt quite different methodologies when it comes to determining what it is for a creature to possess such a thing. One group (Korsgaard and Rowlands) largely sees this as a conceptual project, while the other (de Waal, Bekoff and Pierce, and most of the other authors cited in Section 2) seem to conceive

of the project as one of uncovering the nature of a psychological natural kind.[18] Part of the reason for this methodological difference when it comes to determining the answer is that the two groups also interpret the *question* of what it is to possess a morality or a moral psychology quite differently.

It is very important to understand why philosophers like Korsgaard regard a capacity for normative self-government to be the essence of morality. As Rowlands points out, the key idea—which seems to be common to many philosophical traditions, not just the Kantian tradition to which Korsgaard owes her allegiance—is that this kind of reflective capacity is necessary in order for a creature have *control* over its motivations, and for the creature to exercise *autonomy*. Such control/autonomy is seen as necessary in order for both motivations and actions to count as genuinely moral:

> A motivation can count as moral when it is morally normative. And a motivation can be morally normative only when its subject has control over it. Control consists in the ability to critically reflect on or scrutinize one's motivations (a claim endorsed by both Kant and Aristotle), and this may be a function of the practical wisdom that allows one to grasp the morally salient features of a situation… There can be no (moral) normativity without control. That is why animals cannot be moral subjects: they cannot control their motivations and so those motivations have no normative status. (Rowlands, 2012, p122)

This passage illustrates a key point: for philosophers like Korsgaard and Rowlands, descriptive claims about moral psychology—e.g., what constitutes moral judgment, moral reasoning, or moral motivation—are crucially bound up with high-level meta-

[18] It would be too strong to say that Korsgaard and Rowlands regard the debate over animal morality as *entirely* conceptual. For instance, Korsgaard would be forced to abandon her view if empirical research established that normative self-government was in fact beyond the psychological capacity of human beings or that our reflective capacities never played a role in our putatively moral behaviour. The same would hold for Rowlands if cognitive science established that other-directed emotions play no motivational role in human or animal behaviour. Hence, both would accept that empirical research could show that the conditions of their respective analyses of what it is to be a moral creature fail to be met by prototypically moral creatures (i.e., human beings) and thus should be abandoned or modified. Both also regard it to be an empirical question as to which species actually turn out to be moral creatures, given whatever analysis is finally accepted. However, that is quite different from seeing empirical research as the primary tool for determining what morality *is*, which is the position of the natural kind approach.

ethical questions about the nature of normativity and moral value. Morality is here regarded as inherently normative: it makes normative claims on us. To have a moral psychology is to, in some sense, to live within the world of such normative claims. Rowlands and Korsgaard's primary concern in considering whether animals can be said to have a moral psychology is thus whether they possess psychological states that can be viewed as having what Rowlands calls *normative grip*—whether they have the right kind of normative status or "oughtness" to them. Crucially, this notion of normative status or oughtness isn't a purely psychological one. It is not simply about the causal role that particular psychological states (such as other-directed emotions) play in driving the behaviour of the creature; it is about whether these states have appropriate sensitivity to "the morally salient features" of the situation. In this respect, a creature can only possess a moral psychology if its psychological states stand in an appropriate metaphysical relationship to *moral properties*—in particular, whether these properties (however they are to be understood) can be seen as making normative claims on the creature.[19] Thus, when Rowlands talks about a creature acting for moral reasons, this is a claim about the normative status of the psychological states that motivate the creature's actions. What this means is that the answer to the question of what counts as a genuinely moral psychology and whether animals can have such a psychology depends on the account that we give of the nature of normativity and moral properties, and hence of the grounds for morality. In this sense, moral psychology is fully enmeshed in the classic foundational questions of meta-ethics.

Korsgaard (1996) has a rich and complex account of normativity and of what it is for morality to make normative claims on us as human beings, the full details of which are beyond the scope this paper. However, the basic idea is that the normativity of morality stems from "the reflective structure of human consciousness": ''The source of the normativity of moral claims must be found in the agent's own will'' (1996, p19); ''Autonomy is the source of obligation, and in

[19] It is important to note that standing in this relationship shouldn't require that one must have the *correct* moral beliefs or attitudes, otherwise this might rule out the possibility of someone like the neo-Nazi possessing a moral psychology.

particular of our ability to obligate ourselves'' (1996, p91). Since reflection is the source of moral normativity, creatures without such reflective capacity simply cannot live in the world of the normative, the world of values, or "the kingdom of ends". Hence, nothing they do or think can have *anything* to do with morality, except insofar as they make moral claims on us *qua* creatures with the right type of reflective capacity.

All of this lies in the background of Korsgaard's response to de Waal. This is why normative self-government is the essence of morality and animals lack even the building blocks of morality. This is also why it insufficient for critics of Korsgaard (e.g., Musschenga, 2013; Andrews, 2015) to respond to her denial of animal morality by citing the empirical work in cognitive science on the apparent rarity of reflection in everyday "moral" cognition in humans (e.g., Haidt, 2001; Mikhail, 2011). For one thing, Korsgaard can emphasize that neuro-typical humans still have this reflective capacity, even if it is rarely and imperfectly deployed, so these humans can still live within the world of the normative, while animals cannot. But, most importantly, she can claim that this work tells us nothing about normativity and hence about *moral* psychology, as she conceives of it.

Rowlands' goal is to undermine such intellectualist accounts of what it is for a creature's psychological states to have normative status. He argues that the notion of control or autonomy invoked by Kantians like Korsgaard is elusive and that adding in a capacity for reflection, by itself, fails to explain how particular psychological states can gain normative status. This opens the door for an externalist (rather than internalist intellectualist) account of what it is to have normative sensitivity to morally salient features of a situation. Add to this a broadly consequentialist account of moral properties, and we have an account of moral normativity that allows that the motivational states of unreflective animals and creatures like Myshkin—e.g., other-directed emotions, which track others' suffering or well-being—can be regarded as genuinely *moral* reasons for action.

This shows why the kind of conceptual analysis project Korsgaard and Rowland's are pursing is rather more complex than the picture of conceptual analysis we get from Stich and colleagues. Defining what counts as genuinely moral

cognition and motivation isn't just about comparing proposed analyses with intuitions about the application of the concepts. It is also crucially bound up with a variety of background theoretical concerns about the nature of normativity and moral properties and what it is for one's psychological states to track, or be appropriately sensitive to, such things—Rowlands depends on a broadly consequentialist theory (and, arguably, a form of moral realism [Monsó, 2015]), while Korsgaard adopts a form of Kantian constructivism. This means that the resultant analyses can be revisionary with respect to ordinary folk intuitions about the application of these concepts, in so far as this is necessary to meet the relevant background theoretical constraints (this is why the Myshkin case can't do all the work for Rowlands).

However, most importantly, this also shows why the two groups interpret the question of what it is to possess a moral psychology and whether animals have such a psychology quite differently. Despite their disagreements, Rowlands and Korsgaard agree that this is a question fundamentally bound up with philosophical theorizing about the nature of moral normativity. The issue is whether animals can have psychological states with normative status—whether they can live within the world of values. However, de Waal and Bekoff and Pierce do not seem to be concerned with such high-level philosophical issues about the nature of moral properties and what it is to be psychologically sensitive to the normative. Though they do make approving references here and there to the sentimentalist tradition (de Waal makes several references to Adam Smith, for instance) and to other philosophical theories of morality, neither of them offer, or claim to presume, any account of normativity or moral value. They *are* concerned with whether animals make normative evaluations in a purely psychological sense—for instance, judgments of disapproval about others' behaviour—but this isn't sufficient for normativity in the sense at stake for Korsgaard and Rowlands, since one has to show that such judgments make normative claims on the animal (i.e., that they bear the appropriate relationship to moral properties, however that relationship and the properties themselves are to be understood), rather than just playing a causal role in driving behaviour. Indeed, their project would still seem to make sense even if

one were to deny that there is such a thing as moral normativity or moral value in the sense of interest to meta-ethicists, or that humans possess psychological states with normative status. The issue, for them, is about uncovering the psychological mechanisms that underlie morality, conceived as a psychological natural kind—i.e., the extension of what we *ordinarily* regard as moral thinking or motivation—whether or not this matches up with what comes out of a meta-ethical account of what moral normativity is, and whether any animals possess instances of this kind.

Hence, what we have here are two very different conceptions of what it is to do moral psychology. The approach of de Waal and Bekoff and Pierce falls broadly in line with the interdisciplinary field that now gets called "empirical moral psychology", exemplified by researchers like Nichols (2004), Mikhail (2011), Haidt (2012), Greene (2013), Prinz (2014), and many others. For such empirical moral psychologists, the discovery, for instance, that normative reflection is rare in most actual human beings is a compelling reason to downplay its role in human moral cognition. Whatever normative or meta-ethical implications might be taken to follow from the answers that are given, questions about what counts as moral judgment, moral reasoning, and so forth, are seen as empirical questions that can, at least at the outset, be bracketed off from, say, an account of moral value. However, for Korsgaard and Rowlands' brand of moral psychology, how one answers such questions *is* crucially philosophically loaded, and it is impossible to understand the motivations for their respective accounts of what it is to have a moral psychology without understanding their background assumptions about the nature of moral value.

I don't want to downplay the interest or the importance of the meta-ethical issues that form the background to Rowlands and Korsgaard's contributions. Rowlands, for instance, thinks that the question of whether there is an adequate account of normativity that can potentially be applied to the psychological states of animals is particularly important for thinking about our ethical responsibilities towards them. If it turns out that some animals are capable of doing good *qua* moral subjects, then, Rowlands (2012, p250-254) argues, this should incline us towards the view that these creatures are worthy of moral respect. Although they can't be

praised (or blamed) for what they do (since they lack full moral agency), they are capable of making the world a *better* place, and are thus deserving of a type of respect that goes beyond simply admiring their aesthetic or other properties.[20] That said, I do think that there is something fishy about defining moral cognition and motivation in such a way that it corresponds to one's preferred account of moral normativity and moral value. The idea seems to be that we can save the normative force of morality if we define it like *this*, hence anything that doesn't fit with this account of normativity can't really be morality.[21] This savours somewhat of stacking the deck, and certainly runs against the view of many contemporary meta-ethicists—including, for instance, Joyce (2006) and Kitcher (2011)—that the best way to approach the classic foundational questions of meta-ethics is to start from an empirically informed account of the nature and evolution of moral cognition, rather than have one's account of the nature of moral cognition depend on prior answers to these questions.

Nonetheless, what I think is curiously missing from the discussions of Rowlands and Korsgaard (given how careful they are in other respects), along with various commentaries from others on their putative engagements with de Waal and Bekoff and Pierce (e.g., Musschenga, 2013; Andrews, 2015), is awareness of the fact

[20] Although claims about the supposedly unique place of human beings in the world of the normative have often been used to justify our using animals for food and other purposes, Korsgaard (2006, p119) actually views the implications of what she believes to be our normative uniqueness quite differently: "As beings who are capable of doing what we ought and holding ourselves responsible for what we do, and as beings who are capable of caring about what we *are* and not just about what we can *get* for ourselves, we are under a strong obligation to treat the other animals decently, even at cost to ourselves".

[21] That kind of strategy seems to be particularly central to the Kantian tradition. As I read the *Groundwork*, Kant starts from the presupposition that morality can't be universal and rationally compelling unless we view it as springing from the dictates of reason and thus based on *a priori* rather than *a posteriori* foundations. This leads to an account of what it is to engage in moral cognition and action. Hence, epistemic and metaphysical concerns about the grounds of morality drive Kant's account of what moral cognition and action *are*. Similarly, Korsgaard's (1996) response to what she calls "the normative problem" and her concerns about traditional forms of metaphysical moral realism, drive her account of what normative thinking consists in. Rowlands is less guilty of this, regarding his sentimentalist account of moral motivation as at least partly motivated by empirical evidence about the role of emotions in human moral behaviour. After raising problems for reflection-based accounts of normativity, Rowlands (2012, chapter 9) takes seriously the possibility that normativity might be an illusion, but he still seems to think that understanding the nature of moral motivation and whether animals can act for moral reasons requires that we have an account of moral normativity in place.

that, in focusing on issues of normativity, they are pursing a *completely* different sort of project from the one that de Waal and Bekoff and Pierce are pursuing. Rowlands, in particular, sees himself as correcting the logical flaws in the arguments of researchers like Bekoff and Pierce when they claim that animals are moral beings. For instance, following on from his above quoted critique of Bekoff and Pierce's response to Korsgaard, Rowlands says this:

> In invoking the Russian doll analogy, Bekoff and Pierce have, in effect, issued an invitation: why don't you think of morality in this way? Korsgaard and Kant would likely respond with a firm "No thanks." To have any impact, the offer needs to be strengthened into something more like an offer that cannot be refused. (2012, p111-12).

But, Bekoff and Pierce shouldn't really be seen as being in genuine dialogue with Kantian conceptions of morality. That would be to confuse two related, but nonetheless distinct, sets of issues: issues about the normative status (in the metaphysical sense) of the psychological states of these animals—whether they live in the world of values, as Kantians conceive of it, for instance—and issues solely about whether these states represent instances of a natural kind, irrespective of their normative status. It is no failing of Bekoff and Pierce that their account of morality isn't going to be accepted by Kantians that want something completely different from an account of morality than what they claim to provide.

We've seen that the putative engagements between researchers like de Waal and Bekoff and Pierce and researchers like Korsgaard and Rowlands over whether animals possess a moral psychology are actually not genuine engagements at all; these researchers have simply been talking past each other and in more ways than one. This is one respect in which the current debate over animal morality is unproductive. I now want to focus on what I take to be the *core* debate over animal morality: the one that de Waal, Bekoff and Pierce, and most of the others cited in Section 2 are engaged in, which is concerned with the nature and distribution of morality conceived as a natural kind, but which isn't directly concerned with the meta-ethical issues that pre-occupy Rowlands and Korsgaard. Rowlands and

Korsgaard's accounts can be reconstructed in this light. However, I want to argue that this debate, as it stands, is in fact largely terminological and non-substantive.

4. The core debate: merely terminological?

The notion of psychological natural kinds is fundamental to standard conceptions of the subject matter and methodology of modern cognitive science (e.g., Fodor, 1974; Griffiths, 1997; Machery, 2009).[22] Andrews (2015) provides a nice description of this methodology—what she calls the *calibration approach*—in the study of animal cognition:

> [W]e start with a theory about the nature of some mental property, then we use that theory to make a considered judgment about whether some animal has that property, and use that judgment to empirically investigate the property. The results of that investigation may cause us to tweak our theory, our considered judgment, or both. (Andrews, 2015, p22)

She provides several examples of the approach at work, and shows how it can help us to move forward, both in terms of understanding the real nature of particular mental properties and determining how widely shared these properties are in the animal kingdom.

One of the implications of this way of thinking about the methodology of cognitive science is that the discovery that some putative psychological term fails to pick out a clear natural kind seems to constitute a reason to eliminate it from cognitive scientific discourse (Griffiths, 1997; Machery, 2009). For example, there is general agreement amongst cognitive scientists that "intelligence" and "memory"

[22] The notion of natural kinds is, of course, itself a contested one, with many different accounts of what natural kinds are (see Bird and Tobin, 2015; Slater, 2015)—so many, in fact, that Hacking (2007, p238) doubts whether there is "a precise [or] vague class of classifications that may usefully be called the class of natural kinds". There is also controversy about what account of natural kinds is best for cognitive science. As noted earlier, the HPC account (Boyd, 1999) is popular, but faces its problems (see, e.g., Ereshefsky and Reydon, 2015). I don't want to commit to any particular account of natural kinds or psychological natural kinds, but I will take it for granted that there do exist psychological natural kinds that can, at least partly, be understood in terms of their function. Hence, I will not take the more radical approach to critiquing the core debate over animal morality, which would be to challenge the very legitimacy of talking about natural kinds in general or specifically in cognitive science.

are not legitimate theoretical terms because neither term refers to a category of thing that has the hallmarks of a natural kind. The view is that there is no such thing as general intelligence, only more fine-grained types of cognitive capacity sub-served by a wide variety of different neurological mechanisms. Similarly, memory turns out not to be a unified psychological category, rather there are many different types of memory (semantic memory, declarative memory, and so forth), also sub-served by quite different neurological mechanisms.

In this vein, Stich and colleagues argue that, despite the impressive empirical evidence that Turiel and colleagues have marshalled, their account of moral judgment actually fails to describe a really existing psychological natural kind, since one can find many instances where prototypical moral judgments don't display the characteristic psychological profile they regard as typical of the kind (e.g., they aren't viewed by experimental participants as authority-independent or universal), and instances where judgments concerning putatively conventional matters (e.g., violations of norms that don't obviously concern issues of harm and fairness) display the "moral" rather than "conventional" profile (Kelly et al., 2007; Fessler et al., 2015). Hence, the Turiel account of moral judgment *qua* psychological natural kind fails to accurately carve nature at its joints: there isn't enough of a nominological clustering here to say that we have a genuine natural kind. Stich and colleagues go so far as to suggest that the term "moral judgment", when used to refer to a distinct sub-set of normative or evaluative judgments, might, therefore, need to be eliminated from the vocabulary of cognitive science.[23]

However, despite the problems that Stich and colleagues have identified with the Turiel account of the moral/conventional distinction (for responses, see Sousa,

[23] Once again, the assumption seems to be that if Turiel and colleagues are right, moral judgment would constitute something like an HPC kind. Hence, Stich and colleagues' argument is that moral and conventional normative judgments don't display sufficiently stable clusterings of properties to constitute different psychological kinds. Kelly and Stich (2007) also argue that the psychological processes that underlie the two putative types of judgment are likely the same, which would threaten the idea of there being two different homeostatic mechanisms. Stich and colleagues do, however, regard the more general category of normative judgment as a genuine natural kind.

Sinnott-Armstrong and Wheatley (2014) make a different type of argument for a similar conclusion: the category, moral judgment, is dis-unified in a similar manner to memory, so fails to constitute a genuine natural kind.

2009; Kumar, 2015), there remains the possibility of other ways of thinking of morality as a psychological natural kind. In the case of the debate over animal morality, Andrews expresses a certain amount of optimism about using the calibration approach to eventually determine "the sort of capacity required to make the moral-looking behaviour into truly moral behaviour" (2015, p184). I am less optimistic than Andrews, however. To explain why, let us review some of the different accounts that have been proposed for defining morality *qua* psychological natural kind.

As we've seen, Bekoff and Pierce adopt a very broad view, lumping together capacities for empathy and sympathy and other potential mechanisms for producing various (psychologically) altruistic helping and consoling behaviours, capacities underlying the internalization and enforcement of social norms (like norms governing play and the treatment of infants in chimpanzees, say), and more cognitively sophisticated capacities for explicitly formulating and promulgating social norms, explicit normative reflection, and so forth, that may only be present in humans. Once again, their reason for lumping these capacities together into a single category is that they are assumed to share broadly the same evolutionary *function*: they each represent ways of facilitating social co-operation and cohesion—that is the purpose of morality, as Bekoff and Pierce conceive of it.

Musschenga (2013) also offers a functional definition. However, his is slightly more restrictive in that it explicitly builds in the idea that morality only truly exists when there are norms or rules. While empathetic or altruistic behaviours may be the product of internalized social norms, Bekoff and Pierce appear to allow for the possibility of there being moral creatures with capacities for empathy and altruism, but without social norms.[24] Musschenga's functional definition is as follows:

[24] Although Bekoff and Pierce (2009, p83) include "established norms of behaviour" under their threshold requirements for being a moral creature, their original definition of "morality" actually seems to leave open the possibility of there being moral behaviours in animals (e.g., instances of pro-social and altruistic behaviour) that aren't necessarily guided by psychologically internalized norms, since, when describing which behaviours are moral, it says only that "norms of right and wrong attach to many of them" (2009, p7).

> Morality cultivates and regulates social life within a group or community by providing rules (norms) which fortify natural tendencies that bind the members together—such as sympathy, (indirect) reciprocity, loyalty to the group and family, and so on—and counter natural tendencies that frustrate and undermine cooperation—such as selfishness, within-group violence and cheating. (2013, p102)

Musschenga stresses that norms don't have to be explicitly formulated or consciously understood to count as moral in this sense, but can be implicit—for instance, "humbly internalized" in the sense of Railton (2006).

As we've seen, de Waal labels most of the capacities identified by Bekoff and Pierce as "proto-moral" and restricts the term "moral" for those that enable humans to explicitly formulate and promulgate social norms that concern the community as a whole: proto-morality is a social glue, but *morality* (so understood) is a *better* social glue. Here, again, morality is to be understood in evolutionary-functional terms, but its adaptive function is defined more narrowly than on Musschenga and Bekoff and Pierce's accounts, hence "morality" picks out a narrower set of capacities. Kitcher (2006, 2011) offers a similarly narrowed functional definition of what he terms "the ethical project". Kitcher grants other apes with capacities for sympathy, empathy, and psychological altruism, but argues that these capacities are limited in important ways, and claims that a linguistically-mediated capacity for "normative guidance"—essentially, a capacity to internalize and enforce commands, particularly in reference to one's own behaviour—evolved in early hominins specifically to remediate the repeated "altruism failures" from which other apes, particularly chimpanzees, suffer. The function of morality is thus to overcome what Kitcher sees as the fragility and instability of the societies of our ape-like ancestors.[25]

Rudolf von Rohr et al. (2011) adopt a very similar position to de Waal and Kitcher in their taxonomy of social norms, restricting the term "moral" to norms that are "collectivized": norms that are publicly understood by each member of the

[25] Kitcher (2011) uses this descriptive definition to build a form of pragmatic ethical naturalism, where a notion of ethical truth (and with it a normative standard for assessing ethical propositions) is constructed from this conception of the adaptive function of morality (for discussion, see Joyce, 2014).

community *as* norms of the community, rather than just existing in the form of implicit or personal expectations about appropriate and inappropriate behaviour. They argue that this collectivization of norms requires the capacity for shared-intentionality (the ability to actively share one's mental states with others and to understand that others have, for instance, the same goals or beliefs as oneself), which is widely believed to be unique to humans, and, probably, language. Hence, while animals like chimpanzees possess what they call "proto-social norms", they lack genuine morality.

Joyce (2006) and Prinz (2014) shift the focus from moral norms to the capacity for moral judgment as the essence of morality. Much like Turiel and colleagues, Joyce characterizes moral judgments as possessing a particular kind of psychological clout and inescapability to them. However, he also draws a fundamental link with concepts like *merit* and *desert*. He argues that other primates likely can make primitive normative judgments: for instance, judgments of disapproval about others' norm-violating behaviour. However, they can't make truly moral judgments because they plausibly can't judge, for instance, that a punitive response to another individual's behaviour is *merited* or *deserved*. Such concepts are, Joyce argues, likely beyond the reach of non-linguistic animals. Prinz identifies genuinely moral judgments with dispositions to feel certain complex self- and other-directed emotions like *guilt*, *shame*, and *contempt*: to judge something to be wrong is to be inclined (e.g., upon reflection) to feel such emotions towards the creature that did the thing in question. Such complex moral emotions, he suggests, are probably uniquely human.[26]

Others can be seen as requiring more cognitive sophistication than just the capacity to make moral judgments in either of the senses just described. As noted earlier, Hume can be read as holding that to have a genuinely moral psychology requires that one be capable of departing from one's private situation and adopting a position of impartiality, while Darwin held that only when a capacity for normative

[26] Prinz (2011) argues that empathy is neither necessary nor sufficient for moral judgment. Unlike Joyce, Prinz (2014) does not regard the capacity for moral judgment as a biological adaptation, but rather as a byproduct of other uniquely human adaptations.

self-reflection emerges do we have genuine morality (Rowlands, 2012). We can reconstruct Korsgaard's claims in a similar light, requiring a particularly sophisticated from of meta-cognition, as does Ayala's account.

In contrast, we can reconstruct Rowlands as claiming that it is a sufficient (but not a necessary) condition for morality *qua* natural kind to be instantiated when there is the capacity for behaviour motivated by other-directed emotions. Similarly, for Andrews and Gruen (2014), it is sufficient to demonstrate the existence of morality to show that members of a particular species are at least implicitly sensitive—e.g., via other-directed emotional capacities—to the needs and interests of others.

One can take issues with each of these accounts in various ways. However, the key question for our purposes is on what grounds should any of these proposals be preferred to any of the others as an articulation of the extension of the terms "morality" or "moral"? Once again, the assumption seems to be that this isn't a purely terminological dispute, but one of actual substance. Yet, when these different accounts are placed side by side, it is hard to see what justification there can be for regarding one of them as having any greater claim to the term "morality" than the others. Crucially, though each of these accounts makes substantive empirical claims about human and animal psychology and evolution that might turn out to be mistaken in various ways, the primary differences between them when it comes to categorizing various psychological capacities or mechanisms as inside or outside the moral domain don't appear to turn on such claims. For instance, Joyce and Bekoff and Pierce can agree that animals possess capacities for empathy and sympathy, pro-social and altruistic motivation, that they can make primitive normative judgments in accordance with social norms that exist in their communities, and that these traits broadly share a similar evolutionary function of improving levels of co-operation and social cohesion. They can also agree that only humans likely possess the concepts of merit and desert. Joyce and Bekoff and Pierce can also agree with Prinz that only humans are capable of feeling emotions like guilt and shame. All of them can potentially agree with de Waal, Rudolf von Rohr et al., and Kitcher that only in human communities do there exist explicitly shared norms of conduct, with

Hume and Darwin that only humans are capable of engaging in conscious normative reflection and adopting a position of impartiality, and with Korsgaard that only humans have the capacity for normative self-government. The disagreement between these researchers seems to concern not the nature of these capacities or which of them we share with other animals, but solely which grouping of them deserves a particular label. Hence, it is hard to see what empirical discovery could help us decide between them. In this respect, the situation appears to be quite unlike classic disagreements about the nature of natural kinds, such as whether heat is a type of fluid or molecular kinetic energy, where the disagreement is clearly resolvable via empirical investigation.

Interestingly, Joyce (2014) draws the comparison between accounts of the nature and evolution of moral judgment with discussions about the nature and evolution of language. Historically, linguists, particularly those in the Chomksian tradition, have tended to get upset if animal communication systems such as the honey bee waggle dance or primate alarm calls are referred to as "languages", since they like to reserve the term "language" for communication systems that have complex hierarchical and recursive syntactic structure. However, from the perspective of pre-theoretical classification, intuitions clearly appear to vary on whether these are languages or not. More recently, Chomskians (e.g., Hauser et al., 2002) have seemed to accept that there is no debate of substance to be had about whether the term "language" should be understood in a broad fashion to include such non-syntactically structured communication systems, or in a more restrictive way to refer solely to communication systems with such structure. As Joyce notes:

> There is no answer to the question of which idea captures what is "*really*" language; our vernacular concept of *language* is simply not so fine-grained as to license one answer while excluding the other. Faced with the query of whether vervet monkeys, say, have a language, the only sensible answer is "In one sense yes and in one sense no." (2014, p272)

Joyce continues, "The same may be true of morality. The vernacular notion of a moral judgment may simply be indeterminate in various respects, allowing of a variety of precisifications, with no particular one commanding acceptance" (ibid.).

That seems right to me. However, Joyce goes on to say that his account is an account of moral judgment "strictly construed", and that we can only countenance the possibility of chimpanzees possessing the capacity for moral judgment "very loosely construed". It is unclear why we should grant him this use of "strict" and "loose", since it implies different degrees of correctness of usage. What I think we should take away from this comparison is that it is fine for linguists to adopt the term "language" as a term of art and give it a precise definition for their own purposes that would exclude, say, vervet monkey alarm calls, but there is no genuine question as to whether this is "strict" or "loose" usage. In other words, "language" can be used to refer to (at least) two different putative natural kinds—the set of communication systems that would include human natural languages like English *and* vervet monkey alarm calls, or to the subset of such communication systems that utilize hierarchical recursive structure—but there is no question of which kind should be seen as having greater degree of ownership over the term—i.e., which constitutes *real* language. Similarly, it is fine for Joyce to use the term "moral judgment" to refer only to creatures that possess the concepts of merit and desert—excluding, for instance, the negative reaction of a chimpanzee towards a conspecific that attempts infanticide—but there is no question of whether this constitutes *real* moral judgment.

My claim, then, is that the core debate over animal morality *does* seem to be concerned with the nature and distribution of a variety of what may be genuine psychological natural kinds, namely capacities for empathy and sympathy and other potential capacities that may produce altruistic behaviour, capacities to internalize and enforce norms of conduct of various sorts, capacities for various types of normative judgment or evaluation, reflective normative reasoning, and so forth. There may also be groupings of these capacities that constitute genuine natural kinds—for instance, in virtue of sharing a common evolutionary function. Clearly, much work has to be done to understand the nature of the cognitive mechanisms that underlie these capacities, and when and why they might have evolved. However, the question of which of these, or which grouping of them, if any, constitutes "morality" or mere "proto-morality" is not a substantive one; it is merely

terminological. One may wish to appropriate the terms "morality", "moral", "proto-moral", to pick out some particular set of these capacities, but there is no reason to regard any such way of doing so as more or less correct. Like *language*, the pre-theoretical concept, *morality*, just isn't precise enough to give us a principled reason to endorse one over any of the others.

In places, Bekoff and Pierce suggest a pragmatic justification for their broad definition of "morality":

> The concept of animal morality encourages a unified research agenda. An exploration of moral behaviour in animals allows a number of seemingly distinct research agendas in ethology—research on animal emotions, animal cognition, and diverse behaviour patterns such as play, cooperation, altruism, fairness, and empathy—to coalesce into a coherent whole. (2009, p54)

They also claim that this shifts the focus to looking for cognitive similarities between humans and animals and that this can produce unexpected and important discoveries. But, of course, those that pursue cognitive differences between humans and animals as their guideline and wish to adopt more restrictive definitions of "morality" for that heuristic purpose could say much the same. In any case, surely what really matters is the nature of the underlying cognitive similarities and differences, and that we actively look for such things, not the terms we use to describe them. Whether or not we use terms like "morality" or "proto-morality" seems entirely beside the point.[27]

5. Concluding remarks

At this point, it might be suggested that Rowlands and Korsgaard actually have the right idea: if the pre-theoretical concept of morality is importantly indeterminate, such that it admits precisification in terms of a variety of different putative natural

[27] It might, of course, turn out that similar problems exist with respect to other terms in this debate. "Empathy", in particular, has also been the subject of terminological disagreement (e.g., what is *real* empathy? Is emotional contagion *really* empathy?), as has "social norm" (e.g., can there *really* be social norms without mindreading and language?). If that is the case, then similar conclusions should follow: what matters is the nature of the relevant psychological capacities and associated behaviours possessed by humans and animals, not the terms used to describe them.

kinds, it seems that it is only by appealing to some substantive moral theory that one can have a principled reason to pick out some set of psychological capacities as constitutive of *real* or *genuine* morality. Hence, what I have called the conceptual approach to animal morality might be seen as having an advantage over the natural kind approach: we do actually have to import heavy duty philosophical assumptions about the nature of moral value, normativity, and so forth, in order to determine whether morality *really* does exist in animals—that question cannot be rendered as a purely empirical one, if it is meant to be more than a matter of mere terminology.

However, while I think that there is much to admire, for instance, in Rowlands' (2012) externalist consequentialism and its application to animals, given the seeming intractability of the philosophical debates over such issues and the fact these are not waters in which they originally wanted to tread, I think that researchers like de Waal and Bekoff and Pierce would be well advised to continue to stay clear of the metaphysics of normativity. Hence, insofar as the core debate over animal morality is indeed meant to be both independent of substantive philosophical theorizing about the nature of moral value, and a debate of real substance, not mere terminological preference, it needs to be re-framed—there isn't anything of value to be had debating whether "morality" or "proto-morality" exists in animals.

As the discussion and the end of the previous section indicates, my suggestion is that researchers should adopt a more fine-grained taxonomic approach, focused on uncovering and delineating the particular psychological capacities and mechanisms that underlie the kinds of social behaviours that have been the subject of this debate. Of course, in this respect, I am not suggesting something that researchers haven't already been doing. However, such details about, for instance, the sorts of other-directed emotional capacities present in various species or the various types of social norms present in animal communities (and how they might be acquired and have evolved over time), have often been lost behind headline claims about "morality". The fact that research agendas have often been driven by prior conceptions of what counts as morality or moral behaviour has also meant that some social cognitive capacities have received more attention than

others. Hence, what I am suggesting is that these sorts of details ascend to the fore, and that they be separated from questions about "morality" or "proto-morality".

Finally, it might be objected that one important reason to keep the core debate focused on whether morality exists in animals concerns the potential ethical ramifications of the issue, given that some common arguments against the ethical considerability of animals and in defence of our using particular species for food, medical research, entertainment in zoos and circuses, and so forth, presume that morality is uniquely human and constitutes an ethically relevant difference between humans and other animals.[28] Musschenga (2013), for instance, argues that ascribing morality to some animal species would erase an important potential ethical difference maker between them and us, and should lead us to treat these species with more respect than might otherwise be the case.[29]

Again, however, it seems that what really matters when it comes to determining what bearing, if any, the core debate over animal morality might have on questions about the ethical status of animals, concerns the nature of the actual psychological similarities and differences between humans and animals, not what terms we use to describe them. For instance, when philosophers try to defend the ethical superiority of human beings by saying things like, "Normal human life involves moral tasks, and that is why we are more important than other beings in nature" (Machan, 2002, p10), it is not the use of the term "moral" that is really of significance, but what specific capacity is claimed to constitute the ethical difference maker, exactly what this capacity is taken to consist in, which species possess it, and whether its presence or absence can really support particular ethical conclusions. Thus, although the terms "moral" and "morality" might perhaps have more

[28] For instance, one sometimes hears it said that rights can only be extended to members of a moral community, or that only creatures that are themselves capable of participating in moral deliberation can have direct ethical status.

[29] As noted earlier, Rowlands also argues that attributing morality to particular species should lead us to treat them with greater respect. However, his argument is based not so much on the erosion of a putative psychological difference between them and us, but on the idea that these animals are capable (like some human beings) of being motivated by the good-making features of an action and of doing good. Again, this requires not just an account of the psychological states and capacities of these animals, but also an account of their metaphysical relationship to moral properties.

rhetorical effect in arguments for or against particular human practices towards animals, re-framing the core debate over animal morality in the way that I have suggested would not, I claim, deprive it of its potential ethical import. In fact, it would help to sharpen such discussions.

References

Andrews, K. (2009). Understanding norms without a theory of mind. *Inquiry*, 52, 433-448.

Andrews, K. (2015). *The Animal Mind*. New York: Routledge.

Andrews, K., and Gruen, L. (2014). Empathy in other apes. In H. Maibom (ed.), *Empathy and Morality*. New York: Oxford University Press.

Ayala, F.J. (2010). The difference of being human: morality. *Proceedings of the National Academy of Sciences*, 107, 9015-9022.

Bartal, I.B.A., Decety, J., and Mason, P. (2011). Empathy and pro-social behavior in rats. *Science*, 334 (6061), 1427-1430.

Bekoff, M., and Pierce, J. (2009). *Wild Justice: The Moral Lives of Animals*. Chicago: University of Chicago Press.

Bird, A., and Tobin, E. (2015). Natural kinds. In E. N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, URL = http://plato.stanford.edu/archives/spr2016/entries/natural-kinds/

Boehm, C. (2012). *Moral Origins: The Evolution of Virtue, Altruism, and Shame*. New York: Basic Books.

Boesch, C. (2002). Cooperative hunting roles among Taï chimpanzees. *Human Nature*, 13, 27-46.

Boyd, R. (1999). Homeostasis, species, and higher taxa. In R. A. Wilson (ed.), *Species: New Interdisciplinary Essays*. Cambridge, MA: MIT Press.

Clayton, N., and Dickinson, A. (1998). Episodic-like memory in during cache recovery by scrub jays. *Nature*, 395, 272-274.

Clayton, N., and Dickinson, A. (2010). Mental time travel: Can animals recall the past and plan for the future? In M.D. Breed and J. Moore (eds.), *The Encyclopedia of Animal Behaviour, Volume 2*. Oxford: Academic Press

de Waal, F.B.M. (1982). *Chimpanzee Politics: Power and Sex Among Apes*. London: Unwin.

de Waal, F.B.M. (1996). *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*. Cambridge: Harvard University Press.

de Waal, F.B.M. (2006a). Morally evolved. In J. Ober and S. Macedo (eds.), *Primates and Philosophers: How Morality Evolved*. Princeton: Princeton University Press.

de Waal, F.B.M. (2006b). The tower of morality. In J. Ober and S. Macedo (eds.), *Primates and Philosophers: How Morality Evolved*. Princeton: Princeton University Press.

de Waal, F.B.M. (2014). Natural normativity: The "is" and "ought" of animal behaviour. *Behaviour*, 151, 185-204.

Darwin, C. 1871. *The Descent of Man and Selection in Relation to Sex*. New York: Appleton and Company.

Ereshefsky, M., and Reydon, T. (2015). Scientific kinds. *Philosophical Studies,* 172, 969-86.

Fessler, D.M.T., Barrett, H.C., Kanovsky, M., Stich, S., Holbrook, C., Henrich, J., Bolyanatz, A. H., Gervais, M., Gurven, M., Kushnick, G., Pisor, A.C., von Rueden, C., and Laurence, S. (2015). Moral parochialism and contextual contingency across seven societies. *Proceedings of the Royal Society B,* 282, 20150907.

Flack, J. C., and de Waal, F.B.M. (2000). Any animal whatever: Darwinian building blocks of morality in monkeys and apes. In L. D. Katz (Ed.), *Evolutionary Origins of Morality: Cross-Disciplinary Perspectives*. Thorverton: Imprint Academic.

Flack, J.C., Jeannotte, L.A., and de Waal, F.B.M. (2004). Play signaling and the perception of social rules by juvenile chimpanzees (Pan troglodytes). *Journal of Comparative Psychology*, 118, 149–159.

Fodor, J. (1974). Special sciences, or the disunity of science as a working hypothesis. *Synthese*, 28, 97-115.

Fraser, O.N., and Aureli, F. (2008). Reconciliation, consolation and postconflict behavioral specificity in chimpanzees. *American Journal of Primatology*, 70, 1114–112.

Gert, B. (2016). The definition of morality. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, URL: http://plato.stanford.edu/entries/morality-definition/.

Greene, J. (2013). *Moral Tribes: Emotion, Reason, and The Gap Between Us and Them*. New York: Penguin.

Griffiths, P. (1997). *What Emotions Really Are: The Problem of Psychological Categories*. Chicago: University of Chicago Press.

Gruen, L. (2015). *Entangled Empathy: An Alternative Ethic for our Relationships with Animals*. New York: Lantern Books.

Hacking, I. (2007). Natural kinds: rosy dawn, scholastic twilight. *Royal Institute of Philosophy Supplement*, 82, 203-239.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review,* 108, 814-834.

Haidt, J. (2012). *The Righteous Mind*. New York: Vintage Books.

Hauser, M., Chomsky, N., Fitch, T.W. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298, 1569-1579.

Hume, D. (1978). *A Treatise of Human Nature*. Oxford: Oxford University Press.

Jensen, K., Call, J., and Tomasello, M. (2007). Chimpanzees are rational maximizers in an ultimatum game. *Science*, 318, 107-109.

Joyce, R. (2006). *The Evolution of Morality*. Cambridge, MA: MIT Press.

Joyce, R. (2014). The origins of moral judgment. *Behaviour*, 151, 261-278.

Kelly, D., and Stich, S. (2007). Two theories about the cognitive architecture underlying morality. In P. Carruthers, S. Stich, and S. Laurence (eds.), *The Innate Mind: Foundations and the Future*. New York: Oxford University Press.

Kelly, D., Stich, S. Haley, K., Eng, S., and Fessler, D.M.T. (2007). Harm, affect and the moral/conventional distinction. *Mind and Language*, 22, 117-131.

Kitcher, P. (2006). Ethics and evolution: How to get here from there. In J. Ober and S. Macedo (eds.), *Primates and Philosophers: How Morality Evolved*. Princeton: Princeton University Press.

Kitcher, P. (2011). *The Ethical Project*. Cambridge, MA: Harvard University Press.

Korsgaard, C. M. (1996). *The Sources of Normativity*. Cambridge: Cambridge

University Press.

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

Kumar, V. (2015). Moral judgment as a natural kind. *Philosophical Studies*, 172, 2887-2910.

Machan, T.R. (2002). Why human beings may use animals. *Journal of Value Inquiry*, 36, 9-14.

Machery, E. (2009). *Doing Without Concepts*. New York: Oxford University Press.

Machery, E., and Mallon, R. (2010). Evolution of morality. In J. M. Doris and the Moral Psychology Research Group (eds.), *The Moral Psychology Handbook*. Oxford: Oxford University Press.

Martin-Ordas, G., Haun, D., Colmenares, F., and Call, J. (2010). Keeping track of time: evidence for episodic-like memory in great apes. *Animal Cognition*, 13, 331-340.

Mikhail, J. (2011). *Elements of Moral Grammar: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*. New York: Cambridge University Press.

Monsó, S. (2015). Empathy and morality in behaviour readers. *Biology and Philosophy*, 30, 671-690.

Musschenga, B. (2013). Animal morality and human morality. In B. Musschenga and A. van Harskamp (eds.), *What Makes us Moral?: On The Capacities and Conditions for Being Moral*. Dordrecht: Springer.

Nado, J., Kelly, D., and Stich, S. (2009). Moral judgment. In J. Symons and P. Calvo (eds.), *The Routledge Companion to Philosophy of Psychology*. New York: Routledge.

Nichols, S. (2004). *Sentimental Rules: On The Natural Foundations of Moral Judgment*. Oxford: Oxford University Press.

Prinz, J. (2011). Is empathy necessary for morality? In P. Goldie and A. Coplan (eds.), *Perspectives on Empathy*.

Prinz, J. (2014). Where do morals come from?—A plea for a cultural approach. In M. Christen, M. Huppenbauer, C. Tanner, and C. van Schaik (eds.), *Empirically Informed Ethics*. New York: Springer.

Railton, P. (2006). Normative guidance. In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics, Volume 1*. Oxford: Oxford University Press.

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

Rowlands, M. (2017). Moral subjects. In K. Andrews and J. Beck (eds.), *Routledge Handbook of the Philosophy of Animal Minds*. New York: Routledge.

Rudolf von Rohr, C., Burkart, J.M., and van Schaik, C.P. (2011). Evolutionary precursors of social norms in chimpanzees: a new approach. *Biology and Philosophy*, 26, 1–30.

Rudolf von Rohr, C., Koski, S.E., Burkart, J.M., Caws, C., Fraser, O.N., Ziltener, A., and van Schaik, C. P. (2012). Impartial third-party interventions in captive chimpanzees: a reflection of community concern. *PLoS ONE*, 7, e32494.

Rudolf von Rohr, C., van Schaik, C.P., Kissling, A., and Burkart, J.M. (2015). Chimpanzees' bystander reactions to infanticide: An evolutionary precursor of social norms? *Human Nature*, 26, 143-160.

Sato, N., Tan, L., Tate, K., and Okada, M. (2015). Rats demonstrate helping behavior toward a soaked conspecific. *Animal Cognition*, 18, 1039-1047.

Sinnott-Armstrong, W., and Wheatley, T. (2014). Are moral judgments unified? *Philosophical Psychology*, 27, 451-474.

Slater, M. (2015). Natural kindness. *British Journal for the Philosophy of Science*, 66, 375-411.

Sober, E., and Wilson, D.S. (1998) *Unto Others: The Evolution and Psychology of Unselfish Behaviour*. Cambridge, MA: Harvard University Press.

Sousa, P. (2009). On testing the 'moral law'. *Mind and Language*, 24, 209-234.

Sripada, C., and Stich, S. (2006). A framework for the psychology of norms. In P. Carruthers, S. Laurence, and S. Stich (eds.), *The Innate Mind: Culture and Cognition*. New York: Oxford University Press.

Stich, S. (2009). Reply to Prinz. In D. Murphy and M. Bishop (eds.), *Stich and His Critics*. Chichester: Wiley-Blackwell.

Suddendorf, T. (2013). *The Gap: The Science of What Separates Us from Other Animals*. New York: Basic Books.

Taylor, P. (1978). On taking the moral point of view. *Midwest Studies in Philosophy*, 3: *Studies in Ethical Theory*, 35-61.

Tomasello, M. (2016). *A Natural History of Human Morality*. Cambridge, MA: Harvard University Press.

Turiel, E. (1983). *The Development of Social Knowledge.* Cambridge: Cambridge University Press.

Wallace, G. and Walker, A. (1970). *The Definition of Morality*. London: Methuen & Co.

Wechkin, S., Masserman, J.H., and Terris, W. (1964). Shock to a conspecific as an aversive stimulus. *Psychonomic Science*, 1, 47-48.