

THE MORAL AGENCY OF ANIMALS:  
RESPONSIBLE IN PRACTICE

REBECCA RING

A THESIS SUBMITTED TO  
THE FACULTY OF GRADUATE STUDIES  
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF  
MASTER OF ARTS

GRADUATE PROGRAM IN PHILOSOPHY  
YORK UNIVERSITY  
TORONTO, ONTARIO

DECEMBER 2014

© Rebecca Ring, 2014

## **Abstract**

Mark Rowlands argues some non-human animals can be moral subjects that can act for moral reasons, but cannot be moral agents because they lack sufficient understanding for responsibility. I argue Rowlands' mere moral subjects are responding to, not acting for, moral reasons. Action for moral reasons is necessarily normative and the actor must be able to track the moral reason. I argue Rowlands' conflation of moral agency and moral autonomy results in falsely denying responsibility to animals. Moral autonomy is an ideal to which some humans can aim. Responsibility is not contingent on this ability, but on the cognitive and volitional capacities of the individual and her normative social practices. Some animals can be moral agents in virtue of their normative social practices that involve harm to others and sharing resources. Moral agency and responsibility can be ascribed to some animals in terms of their intentional agency within such practices.

## **Dedication**

*In memory of Anne Mary Christensen, my best friend and kindred spirit. Your beautiful, brilliant mind and spirit, and your unfailing faith in me, will always be my inspiration. This one – the first one - is for you soul sister, with all my love and gratitude.*

## **Acknowledgements**

I am most grateful for all the support, patience, encouragement and invaluable advice of my supervisor, Dr. Kristin Andrews, and my second reader, Dr. Robert Myers. I am also grateful for the equally invaluable input of Dr. Alice MacLachlan and Dr. Joni Sasaki, and for the inspiring and thought provoking discussion of my thesis topic with fellow students, especially Imola Ilyes, Michelle Ciurria, Josh Mugg and Èric Arnau. And I could not have done any of this without the loving support of my partner, Rich Lee, and my parents, Bob and Carol Ring. I would be remiss if I did not express my undying gratitude for the wonderful non-human beings in my life, past and present, who have given me so much love, joy and food for thought. Wondering what goes on in those furry heads is how it all starts. The list could go on, as many others have supported, helped and inspired me along the way, so to all of you, thank you.

## Contents

Abstract .....	ii
Dedication.....	iii
Acknowledgements .....	iv
Introduction .....	1
Section 1: Moral subjects and moral agents .....	5
1.1 Tracking moral reasons .....	6
1.2 Moral reasons and normative force.....	11
1.3 Moral reasons and moral responsibility.....	14
Section 2: Moral agency and Moral autonomy .....	21
2.1 A metacognitive account of moral agency .....	22
2.2 Rowlands' first condition for moral agency.....	23
2.3: Rowlands' second condition for moral agency .....	37
2.4. Rowlands' third and fourth conditions for moral agency.....	49
Section 3: Moral agency in practice .....	69
3.1 A non-metacognitive account of moral agency.....	70
3.1.1 Action for a reason - Intentional action .....	71
3.1.2 Acting for a reason – normative desires.....	74
3.2 The moral practice .....	78
3.2.1: Setting the limits .....	83
3.2.2: Mere convention or moral norm?.....	85
Conclusion .....	87
Bibliography.....	88

## The moral agency of animals: responsible in practice

### Introduction

Can non-human animals be moral agents? Mark Rowlands (2012) has argued that the moral status of non-human animals ('animals' from here on) goes beyond that of moral patient, whereby they have their own interests and those interests should be respected in our dealings with them. But they are not quite moral agents according to him, as the hallmark of moral agency is moral responsibility and accountability, and that involves a level of understanding unique to 'normal' adult humans. Rather, Rowlands claims that some animals are moral *subjects* that can act for moral reasons, and those reasons are emotions with morally laden intentional content that are responses to morally salient features of situations. Positing an externalist consequentialist conception of morality, he argues that to be a moral subject, one need only respond correctly to the morally salient features of a situation, and animals can do so based on emotional experiences of those features. The response can be evaluated according to normative facts, which a moral subject need not be capable of entertaining. He claims that being a moral subject is not sufficient for moral responsibility, and moral responsibility is necessary for moral agency.

Rowlands denies that animals can be morally responsible because this requires a level of understanding that he claims requires metacognitive capacities which all non-human animals lack. He lays out four conditions that a moral subject must meet in order to be considered a moral agent: (1) an agent must make fine grained distinctions among the features of moral facts; (2) know the moral facts; (3) know the principles upon which those facts are based; and (4) be able to examine these principles. Rowlands allows that these conditions are required for moral autonomy, but insists that moral responsibility requires moral autonomy.

Rowlands distinction between a moral subject and a moral agent rests on two important claims: (1) moral subjects can act for moral reasons; and (2) moral autonomy is necessary for moral responsibility. I will object to these two claims throughout this paper. In Section 1, I will argue that acting for a moral reason entails the possibility of ascriptions of moral responsibility. If I am right, then Rowlands' moral subjects are either not acting for moral reasons, or they are subject to ascriptions of moral responsibility.

In Section 2, I analyze each of Rowlands' four conditions for moral agency, which he claims require metacognitive abilities. I will argue that the first two conditions could be met without metacognition. A non-metacognitive actor could make distinctions among states that are sufficiently fine-grained for his social context and form of life. If his social context has normative moral features (a moral practice), then social competence in that context allows the possibility of ascriptions of moral responsibility. A non-metacognitive actor could know the moral facts. If he can have intentional states that have evaluative content which is directed at the moral facts of situations, and he is socially competent in his moral practice, then he can have propositional content that means "M is wrong". If I am right, then some animals can meet Rowlands' first two conditions for moral agency.

I will argue that Rowlands' last two conditions for moral agency are not required for moral responsibility. They might be required for moral autonomy, but I will argue that moral autonomy is not necessary for moral responsibility, and that moral agency and moral autonomy are distinct concepts. The implication of conflating them is that too many normal adult humans will be inappropriately absolved of moral responsibility. I will argue that Conditions 3 and 4 reflect the regression that is inherent in moral autonomy - an ideal that can be aimed at, but never fully realized due to inescapable contingencies of one's form of life.

In Section 3, I will return to the question of what it means to act for a moral reason. I will provide an account of intentional agency that counters Rowlands' claim that an intentional agent can act for a reason that they could never possibly grasp. I will counter Rowlands' concept of normativity and object to it on the grounds that normative reasons must be action guiding, and that requires that an agent can grasp the normativity of the reason. I will argue that moral action – intentional action for a moral reason – does not require metacognitive abilities, and some animals can be intentional agents. I will then provide an account of moral responsibility. I will argue that ascriptions of moral responsibility are necessarily contained within contextual moral practices. Some animals can be responsible moral agents within their moral practice. A non-human animal cannot be a full moral agent in a human moral practice, and a human cannot be a full moral agent in a non-human practice. If moral responsibility is thus contained, then non-human animals that can act for moral reasons can be moral agents.

My *prima facie* concept of moral agency is in keeping with Rowlands' concept of moral subjecthood, in that it involves emotions with propositional content, allows an objective consequentialist account of normativity, external reasons and moral facts. In other words, my account allows that there are facts in the world that make a moral agent's actions right or wrong, and the normativity guiding such actions is not dependent on anything in the agent's motivational set. And it allows that information about such things need not be intellectualized, but can be carried and conveyed via affect. However, it is sufficient for moral agency that one can act for moral reasons and is responsible for those actions. A moral agent knows how to act according to a tacit code of conduct, or moral practice, and can govern his behavior by it (Gert, 2012). The guiding code is moral in that it involves avoiding and preventing harm to others and sharing resources, which include food, territory, and reproductive resources. This might be



considered a descriptive account of morality, but that does not preclude moral responsibility. I will argue in Section 1 that moral responsibility is necessary to morality, whether it is a descriptive or normative account. A normative account would move us into the area of moral autonomy because the focus is on how the content of a moral code should be determined. I maintain that the skills necessary to make such assessments are not necessary for moral responsibility. It is sufficient for moral responsibility that such a code exists; that its content is morally normative in the sense of exerting normative force on the individuals, society or group members involved; and that such individuals have sufficient cognitive capacities to know what kinds of actions their moral code prohibits, requires, discourages, encourages, and allows, and must have sufficient volitional ability to use the moral code as a guide for their behavior. Those lacking these characteristics are not subject to moral judgment (Gert, 2012).

My *prima facie* concept of moral autonomy is in keeping with Rowlands' account of moral agency except that I deny that moral autonomy is required for moral responsibility. A morally autonomous agent acts on motives, reasons, or values that are her own (Stoljar, 2014). To be autonomous is to be one's own person, to be directed by considerations, desires, conditions, and characteristics that are not simply imposed externally upon one, but are part of what can somehow be considered one's authentic self (Christman, 2011). In other words, an autonomous moral agent must be able to try to scrutinize and set himself apart from the contingencies of his existence - of place, culture, education, intelligence, personality, social norms (including moral norms), social relations, etc. We humans are necessarily fallible and vulnerable to our contingencies, so attaining perfect moral autonomy is an ideal that one can aim for, but likely never reach. The outcome of my argument is that non-human animals that can act for moral reasons can be moral agents.

## Section 1: Moral subjects and moral agents

Can animals be moral? Rowlands argues that some animals can be moral subjects in that they can act for moral reasons. He constructs the concept of a moral subject and argues that it is distinct from that of a moral agent. He claims that a moral subject can act for moral reasons, but cannot be held responsible, whereas a moral agent is morally responsible. A moral agent is a moral subject that can be held responsible. On my view, Rowlands' argument makes great strides toward showing that some animals can act for moral reasons, but falls short of making the case in two important ways. First, I will argue that responsibility is essential to acting for a moral reason. Responsibility must be attributable to an actor, if his reason for action is a moral one. Morality *qua* morality requires responsibility because actions can only be deemed as moral in a social context that involves accountability. I will argue that his concept of a moral subject is incoherent because responsibility is a necessary component of the moral quality of the subject's action.

Second, I will argue that the Rowlands' account of normativity results in the behavior of a moral subject being more appropriately considered caused behavior than action for a reason. He provides an objective consequentialist account of normativity, wherein the normativity of the moral fact is external to the subject's motivational set, and the motivations of a moral subject are assessable as correct or incorrect in light of normative facts which can be inaccessible to the subject's reason. On his view, one can act for a moral reason that one can never grasp. I will argue that the implication of this view is that such a subject cannot be considered as acting for that inaccessible moral reason. It cannot be counted among his reasons for action, and can only be counted as a cause of his behavior. If it is categorically impossible for a particular external reason for action to be a motivating reason for a particular intentional agent, then it is better

understood as a cause in that context. For Rowlands, it is enough that the action can be assessed on the basis of its appropriateness as a response to the morally salient features of situations, but I will argue this can only establish that a moral subject might be sensitive to these features, and can respond to them as causes. I will argue that the normative facts about these features can be a moral reason for action only if they are accessible to the actor's practical reason.

### 1.1 Tracking moral reasons

If we assume that animals do not have the metacognitive abilities that humans have, then how can an animal act for moral reasons?<sup>1</sup> Rowlands argues that emotions can be moral reasons for action in that they can have intentional content that is of a moral character. He argues that some animals have emotions with identifiable morally laden intentional content and they sometimes base their actions on these emotions, which he claims are sufficient but not necessary conditions for moral action. Morally laden emotions are marked by concern for the welfare or fortunes of another, such as empathy, sympathy, compassion, anger, jealousy, etc. There is plenty of evidence that makes a *prima facie* case that animals are motivated to act by emotions with moral content. For example, Rowlands cites a study by Russell Church in which rats refused to press a lever that delivered food where pressing the lever also meant delivering electric shocks to other rats. What is at issue is how such evidence should be interpreted.

Rowlands' goal is to show there is no conceptual problem with interpreting the evidence positively – that there are emotions with identifiable moral content, even if the subject cannot entertain the content, and that there is no principled reason to preclude their instantiation in animals. First, he argues that animals can have intentional mental states *tout court*. To act for

---

<sup>1</sup> I am not committed to denying metacognition to non-human animals. I am not convinced that none are capable of it to any degree. However, I will assume in this paper that they are not capable of metacognition. Otherwise, the question would take us too far afield.

moral reasons, an animal must be capable of intentional mental states with prescriptive and descriptive content. This content is action guiding in that it forms the basis for an animal's judgments and evaluations, as opposed to stimuli merely causing his behavior. Rowlands claims that when the intentional state has an emotional character and regards the welfare or fortunes of others, then the animal can be said to act for moral reasons. I will argue later on that this is not sufficient for acting for moral reasons. For now, let us focus on the problem at hand: can some non-human animals have mental states with propositional content?

The problem in attributing intentional states to animals revolves around the identity of the content possessed by those states. Donald Davidson and Stephen Stich claim a creature must be capable of attribution holism in order for him to have an intentional state. To have a single intentional attitude, one must possess a network of related beliefs. Further, truth values of beliefs can only be determined if their intensionality can be determined, which they claim is only possible with human language. How then are we to explain the behavior of such individuals? How do we explain a dog's barking up a tree, into which she chased a squirrel that escaped and is no longer in the tree? Does the dog believe *that* the squirrel is in the tree? Rowlands says that the alternative to ascribing intentional states is to explain the dog's behavior in terms of bare causal contact with the world and describe the content of the state that drives the behavior in demonstrative terms. He argues that this is implausible because it assumes that the dog has no sortal for the objects in her environment, that she has no way of representing the world, and we have no way of explaining her behavior. He sees the problem as being that we do not understand how animals represent their world and which sortals they use to categorize objects.

Rowlands' response to the Davidson-Stich argument lays the groundwork for his claim that an animal can act for reasons she herself cannot entertain, which is important to his

externalist consequentialist conception of morality and to his claim that acting for moral reasons does not require self-scrutiny of one's motivations and actions. His argument supports the idea that one can act for reasons without the metacognitive content *that* one is acting for reasons. He identifies *de dicto* ascriptions of content as the target of the Davidson-Stich argument and says they are *prima facie* problematic because they are sensitive to the way in which the object of belief is represented. This is not true for *de re* ascriptions (the dog believes *of* the squirrel that is in the tree), but the behavior is still not explained as intentional action because *de re* ascriptions do not reflect the intensionality of intentional ascriptions. To justify a *de re* ascription of intentionality, we "need to be in possession of specifically *de dicto* information concerning the mode of presentation of that with which [an individual] is in causal contact" (Rowlands, 2012, p.53). To explain behavior, a *de re* ascription needs a *de dicto* counterpart to provide the intensionality of intentional content. This presents the epistemological problem of whether or not it is possible for us to identify the content of the states we ascribe *de dicto* to animals, and the ontological problem of whether there exists content that is possessed by the intentional states of animals that can be correctly employed in such attributions.

Setting aside the epistemological problem, Rowlands presents a solution to the ontological problem. Humans can explain animal behavior through the use of *de dicto* content (*p*) that tracks whatever content the animal possesses (*p\**):

(*Tracking*): Proposition *p* tracks proposition *p\** iff the truth of *p* guarantees the truth of *p\** in virtue of the fact that there is a reliable asymmetric connection between the concepts expressed by the term occupying the subject position in *p* and the concept expressed by the term occupying the subject position in *p\** (p.58).

Proposition tracking needs to be understood in terms of context. It may be difficult to determine how a dog represents a squirrel that ran up a tree, but as long as the human-context bound

concept employed in  $p$  is reliably, asymmetrically connected to a plausible dog-context bound concept employed in  $p^*$ , then the truth of  $p$  guarantees the truth of  $p^*$ . The dog's behavior can be explained by way of content that she need not be capable of entertaining. This content (e.g. the external moral reason) guarantees the truth of the content that she does entertain (e.g. her motivating reason).

Why is the relationship between the truth conditions of  $p$  and  $p^*$  asymmetric? Rowlands says that there may be more than one plausible candidate for  $p^*$  and anyone of them may be true if  $p$  is true. For example, if  $p$  is "the squirrel is up the tree", then  $p^*$  could be "the chaseable thing is up there" or "the chaseable thing that smells tasty is up there", etc. The relationship is asymmetric because  $p$  can be false and  $p^*$  can be true if, for example, it was a chipmunk that ran up the tree. The implication is that  $p^*$  need not involve the fine-grained distinctions equal to those found in human-context bound propositions. The truth of the human-context bound proposition guarantees the truth of the dog-context bound proposition, but not vice versa, so we need not know the precise content of  $p^*$  (which the dog entertains) and the dog need not be capable of entertaining the content of  $p$ .

Rowlands applies the proposition-tracking method for ascribing intentional states to ascribing morally laden emotional states. He characterizes the latter as the combination of the factual and morally evaluative. Emotions, he claims, involve both factual cognitive content (that a shark is swimming toward me), and evaluative content (that a shark should be feared and avoided). The evaluative component in morally laden emotions "takes on a specifically moral hue" – (that the shark is swimming toward you and I should somehow intervene). Rowlands sees the evaluative component as problematic in ascribing morally laden emotions to animals because of the question of if, and how, they can understand the "ought" of moral obligation. His solution

is to employ the proposition tracking strategy with its distinction between tracking the truth of a proposition and entertaining a true proposition. He modifies the strategy by making a misguided or misplaced emotion analogous to a false belief. An emotion is misguided if it is grounded in a false factual assertion, or misplaced if it is based on an erroneous assumption of entitlement. His concept of a morally laden emotion is:

An emotion, *E*, is *morally laden* if and only if (1) it is an emotion in the intentional, content-involving, sense, (2) there exists a proposition, *p*, which expresses a moral claim, and (3) if *E* is not misguided or misplaced, then *p* is true. (p.69)

The ‘truth’ of *E* tracks the truth of *p*, if there is a reliable asymmetric connection between *E* and *p*. In this way, an animal need not entertain or understand the “ought” of a moral obligation (*p*) and still act for moral reasons (*E*). For example, suppose that Smith is upset because his plans to blow up the CN Tower were foiled. His dog Rufus perceives that Smith is upset, is upset at this sight, and desires to mitigate this, so tries to comfort him. According to the tracking scheme, (*E*) Rufus is sad that Smith is sad; (*p*) It is right that Rufus should comfort the suffering Smith. In this case, *E* is misplaced so *p* is false. Even if Rufus fails to meet the criteria for moral agency, Rowlands would have us count his action as moral because it tracks the moral claim, albeit incorrectly. In other words, Rufus’ emotional response to the morally salient features of the situation is incorrect.

Rowlands claims his account of moral subjecthood does not rely on any one moral theory. What is important for him is that moral motivation can take an emotional form of experience, and since animals have emotions, they can be motivated to act for moral reasons. On his view, some animals possess a normative sensitivity to (some) morally salient features of their situations, where this is grounded in the operations of a reliable mechanism or module. He

construes this sensitivity as normative in that it consists in emotions that have intentional content directed at morally salient features of situations.

## **1.2 Moral reasons and normative force**

Moral reasons are by definition normative, so what gives these reasons normative force? Rowlands claims that normative grip is not internal to the subject, but is an external feature of situations. There are good and bad making features, and a moral subject is sensitive (emotionally or otherwise) to these objective features. His is an objective consequentialist account of morality. Normative value is not found in the subject's beliefs and desires, but rather in the consequences of his actions. He assumes that there are objective moral facts in that there are certain features of a situation that make it good or bad that are independent of the subject's states. These features can be evaluated using different moral theories and he does not want his argument to be dependent on any specific one. Rowlands' point is that the moral quality of a person's actions or motives is determined, at least in part, by factors external to the person's agency. Further, he claims that it does not matter to his argument whether or not those features are intrinsic bearers of value.

In the case of emotions as providing moral reasons for action, it is not the feelings in the subject that constitute the good/bad making features, but rather emotions are the form of detection of those features. In the case of empathy, the happiness or suffering of the other is the intentional object of the subject's emotional state. A moral subject can be emotionally sensitized to features that make a situation good or bad, but need not form general principles about what is good or bad. However, a moral subject need not be directed toward the moral content emotionally. The manner in which she is aware of the content does not determine the content



itself. The manner in which one is sensitive to good/bad making features of situations is not what is at stake, but rather the expressions of this sensitivity.

An emotionally directed subject can be normatively evaluated based on her emotional responses as they can be judged as correct or incorrect in certain situations. Her emotions *track* the good and bad-making (morally relevant) features of situations, and can succeed or fail to do so correctly. On Rowlands' view, the possibility of motivations as being in accordance or discordance with objective moral facts is what confers normative status on those motivations. Further, if a subject's sensitivity is significantly reliable, then it is not just normatively describable but actually normative. A minimal moral subject must possess: "(1) a sensitivity to the good- or bad-making features of situations, where (2) this sensitivity can be normatively assessed, and (3) is grounded in the operations of a reliable mechanism (a 'moral module') (p.230).

Rowlands claims that moral subject and moral agent are logically independent concepts: an individual counts as a moral subject to the extent his or her motivations satisfy the normativity requirement, and such an individual counts as a moral agent to the extent to which she understands her actions, their consequences, and how to evaluate them. He separates the question of motivation from the question of evaluation. He claims that the distinction between a moral subject and a moral agent follows from the distinction between the motivation of an individual's action and the evaluation of that action. He addresses the objection that when the motivation and evaluation are moral ones, then the distinction does not apply, because the subject must evaluate his own motivation in order for it to count as moral. He argues that this objection is based on a false premise that a motivation is moral only if the subject has control over it. He goes on to make a case that it cannot be the possibility of control over one's motivations that confers a

morally normative status on them (I will discuss this argument in Section 2.4). I think this argument works well, but it does not let Rowlands off the hook. He must show how a motivation could stand in relation to a moral reason such that the normative grip of morality is conferred.

Moral reasons are necessarily normative, thus requiring an evaluative component. To avoid ascribing responsibility to a moral subject, Rowlands construes two ways in which evaluation can be involved without such an ascription. Recall that a moral subject can act for moral reasons, where those reasons are morally laden emotions with intentional content. This intentional emotion has an evaluative component, so that the subject is not only sad, but sad that you are sad (intentional), and he should mitigate your sadness (evaluative). He now has the desire to mitigate your sadness and beliefs about how to do so, which result in his action. His motivating reason for action is therefore his desire to mitigate your sadness. His external reason for action is whatever it is that caused your sadness, of which he needn't be aware. His empathetic response to your sadness is correct if your sadness is not morally wrong, and if it is warranted. He may try to console you, unaware that your sadness is morally wrong because it is due to the failure of your evil plan. In this case, his morally laden emotion is misplaced. You may be sad because you believe that God hates you, in which case the emotional response of your compassionate friend is misguided. If he is a moral subject who lacks the sufficient capacity for understanding that is required for moral agency, then his emotional response and the actions he takes cannot be appraised in terms of responsibility on Rowlands' view, but they can be appraised in terms of correctness. If the response accords with the moral facts, then it is correct. It is only the justifying moral reason (external) that has the morally normative component of moral force or normative grip, whereas the motivating reason has merely a moral character (not

moral force) because it involves the welfare and fortunes of others, but no responsibility in its action. I will argue that this disqualifies the response from being a moral action.

### **1.3 Moral reasons and moral responsibility**

Rowlands' view so far raises two important questions: (1) Can one act for moral reasons and not be morally responsible, and (2) can one who is not a moral agent, and therefore not morally responsible, have a moral reason for action? I will argue that moral responsibility is a necessary component of moral action, and if an intentional agent acts for moral reasons, then the action is necessarily moral. Given that Rowlands is committed to the existence of external reasons, we can grant that the normativity of the moral reason may be external to the actor's motivational set for ascriptions of moral responsibility. The fact that an agent may have no reason from his motivational set to commit or refrain from an action does not mean that we cannot hold him responsible. In fact, this notion seems to fit how moral responsibility is ascribed in everyday judgments. The fact that a pedophile may have no reason based on his beliefs and desires to refrain from sexually assaulting a child does not release him from moral responsibility. Further, the fact that he has the bad moral luck to have such desires in the first place offers him no escape. There are morally salient features in this situation that are appraisable independently of the agent's motivational set.

Assuming that moral reasons for action can be external, is moral responsibility a necessary component of moral action? Properly understood, a moral action must be an intentional action. We can again consider the phenomenology of the way in which responsibility is ascribed in the everyday world. For example, if one is driving a car and accidentally hits and kills a dog that ran out in front at the last second, then it would be irrational to ascribe moral responsibility for the action. The person driving the car was doing so intentionally, but hitting the

dog was not an intentional action. Now if the driver was speeding, we might reasonably ascribe responsibility, because the intentional action includes speeding, which can be considered wrong because it may cause harm (above and beyond the accepted risk of driving a car according to all the rules). The moral appraisal in this case could reasonably be that the driver should not have been speeding because it can lead to harm, and harming the dog may have been the result of acting for this immoral reason. The driver in this case is not only responsible for the dog's death, but morally responsible (even if speeding was not the actual cause). In the former case, the driver's actions caused the dog's death, but only in the second case can we (reasonably) ascribe moral responsibility. Only the action of killing the dog in the second case can be said to be moral (and morally wrong), whereas in the first case, there is no moral quality to the action. It may be an unfortunate or unlucky action that killed the dog, and the dog's death may have caused pain for her and her loved ones (concerns the welfare of others), but it can only be a moral action if there is some reason for the action for which the agent can be held morally responsible (a moral reason, like speeding). I do not claim that responsibility can be ascribed whenever a creature acts or fails to act on such reasons. My argument is meant to show that, in order to act for moral reasons, ascriptions of responsibility are necessarily applicable. I oppose Rowlands' claim that a moral subject can act for moral reasons without the possibility of responsibility, because without the possibility of responsibility, the reasons can never be moral ones for that subject. And if they can never be reasons for that subject, then we are talking about causes.

If the agent intentionally committed an action that is "objectively wrong" that resulted in the dog's death, then the agent can be held morally responsible for that action and its results. In other words, for an intentional action to be moral, the intentional agent must be appraisable as being responsible for committing the moral component of the action. Otherwise, it makes no

sense to say that the agent hit the dog for a moral reason, and if there is no moral reason (internal or external), then there is no moral action. Further, if the agent did hit the dog for a moral reason (because he was speeding), it is reasonable that it makes no difference for moral responsibility whether or not the driver was aware of this fact (that he was speeding), or even aware that it is wrong to speed. All that matters is that he is capable of being aware that he is speeding and he is capable of being aware that this is wrong in virtue of intentionally driving the car. This makes sense of legislating a minimum age and level of cognitive function for driving.

It is this essential normativity of moral reasons – the moral ought – which seems to be missing from Rowlands' concept of the moral action of moral subjects. A moral reason for action entails that an intentional agent ought to commit or omit the act, and can be held morally responsible for the action or omission. The moral component of the reason is what makes its associated action praise or blameworthy. Perhaps it is possible for an animal to be sensitive to moral reasons without really acting on them, but if Rowlands wants to say that they really do act on them, then he may have to accept that they are responsible as well.

This brings us to the problem of external reasons, which are reasons that do not depend on anything in one's motivational set. Rowlands argues that external moral reasons can be "owned" by the moral subject, and he can therefore act for that reason, even if it is a reason that the subject can never possibly grasp. But this seems incoherent to me. If the subject is categorically incapable of grasping the actual external reason (Smith is sad because his evil plan failed), then it can never be the subject's actual reason for action. For example, there exists the external prudential reason to get a prostate exam because early detection of prostate cancer can prevent death. But this is not any reason, external or otherwise, for someone who does not have a prostate. The reason must be compatible with the appropriate system or it is no reason at all. A

moral reason for action can only reasonably be attributed to one capable of grasping that reason, because if it is impossible that the moral reason guided her action, then it is incoherent to attribute her action to something that could not possibly guide it. The benefactor who has an external moral reason to donate money (helping others), but whose actual motivation is solely to pay less taxes, can reasonably be morally praised for the action based on the external reason alone *because* it is a reason for which she *could* act. The consequentialist may not care that her motive was greed because the ends resulted in people being helped, but this does not change the condition that the reason needs to be theoretically ascribable to the intentional agent.

One may counter with the example of a cat who is allergic to chicken, which gives him a reason not to eat chicken, but this can never be an actual motivating reason for the cat (assuming the cat has no concept of allergies or allergens) but it is nevertheless a reason that applies to the cat. While it may be true that there is a reason why the cat should not eat chicken, the intentional action of eating things that are good/bad for you may not be in the cat's repertoire of intentional action, so consuming an allergen can never be an intentional action for the cat, but merely a consequence of a behavior (even if the chicken itself was eaten intentionally). He may learn that when he eats chicken he feels sick, so the external reason (allergy) can cause a desire (to avoid chicken), but now we are back in the space of causes.

Rowlands claims moral reasons are objective moral facts, whose moral status is external to the subjective states of the subject. In other words, there is nothing derived from a moral subject's motivational set that confers moral status on the external reason. If that is the case, and the subject is to act for a moral reason, then the moral status of the action must derive from the external reason. There must be some connection between the action and the reason – the reason can justify the action. But how can we make sense of a justifying reason that, on Rowlands'

account, the moral subject *qua* intentional agent can never possibly have as a motivating reason? Looking at the allergic cat example we can see how sense can be made of this, but that example involves strictly *mind-independent* reasons. The cat and everyone else in the world could be mindless automatons and the cat-bot would still have a reason to avoid chicken (if reasons could exist). Moral reasons are different. We cannot make sense of them as being strictly mind-independent. In a world without moral agents, moral action is impossible. Moral reasons for action are mind-dependent in that they necessarily depend on the existence of intentional agents.<sup>2</sup> No intentional agency is required for the allergic cat to have a reason not to eat chicken (leaving aside the question of whether that could then be called a reason or only a cause). But Rowlands is committed to reasons, not just causes, so I cannot make sense of his claim that a moral subject can act for moral reasons, but not be morally responsible for these actions. If a moral subject acts for moral reasons, then she just is a moral agent.

To really act for an external moral reason, it must be possible that the reason is action guiding for the actor. To clarify my point, let's apply Rowlands' concept of a moral subject to a thermometer and see if it can act for a moral reason. If one's motivations track external moral reasons, and can therefore be characterized as moral, then we could ascribe moral motivations to thermometers. When we adopt Daniel Dennett's intentional stance, we can ascribe desires and beliefs to a thermometer in that it has the desire to maintain a constant temperature and believes that the environment either is or is not the same as that to which it is set to turn on or off, and acts accordingly (turn on or off) (Dennett, 1987,1998). The thermometer has the intentional states of desiring a certain temperature and beliefs about how to achieve it. If this thermometer is maintaining the temperature of an incubator keeping an infant alive, then we could say it has the

---

<sup>2</sup> I am not arguing here whether or not the value in the reason is itself mind-dependent, as that would take us too far afield. I only claim that acting for moral reasons entails intentional agency, which is mind-dependent.

external moral reason to adjust the temperature so as to help and not harm the infant. However, the thermometer can never have the baby's welfare among its reasons, or have reasons generally. Dennett points out that there is a difference between knowledge *in* a system and knowledge *for* a system. Knowledge about the temperature is in the system, but knowledge about the moral reason why it is operating can never be knowledge for the system. Only the humans that design, operate and maintain the thermometer can be motivated by this reason as it is their intentional agency at bottom. The external moral reason can only apply to those agents and not to the thermometer itself – it can be a reason *that* the thermometer behaves in a certain way, but not a reason *for* the thermometer to behave in this way. There is no practical reason or intentional agency on the thermometer's part – just cause and effect.

Rowlands characterizes his moral subjects as being real intentional agents that can engage in practical reason. This is entailed by his claim that a moral subject acts for reasons, rather than reflexively responding to causes: “Animals are moral subjects in that they can act for moral reasons. But reasons, unlike mere causes, are states with intentional content” (Rowlands, 2012,p.40). With Rowlands' proposition tracking scheme, we are to keep in mind the context, or form of life, of the subject in question when ascribing intentional content. I very much doubt that thermometers are among his candidates for moral subjects, but his view could qualify them as such. He must allow moral responsibility if he wants to claim that animals can act for moral reasons. If I am right, then the problem becomes how to make sense of ascribing moral responsibility to animals. The worry is that animals may be blamed/praised inappropriately. In Section 3, I propose a sketch of moral agency that can include animals without precluding responsibility or putting them in danger of moral persecution by humans. In this sketch, I look further into what might constitute intentional agency.



In the next section, I will present Rowlands' four conditions for moral agency and argue that his own arguments support the claim that non-human animals can meet the first two conditions. I will argue that the final two conditions are not required for moral responsibility, and that moral agency and moral autonomy are conceptually distinct. Rowlands conflates these concepts, and as a result, requires moral autonomy for moral responsibility. The implication is that too many normal adult humans would not qualify as moral agents.

## Section 2: Moral agency and Moral autonomy

Rowlands has argued that some non-human animals are moral subjects, in that they can act for moral reasons; and no non-human animals are moral agents, in that they cannot be held responsible for their actions or motivations. He claims that ascriptions of responsibility are required for moral agency, and such ascriptions require that the moral subject has the ability to understand her motivations and actions, their consequences, why they are right or wrong, and what makes them so. He lays out four conditions for understanding that would allow a moral subject to be held responsible and thereby be considered a moral agent. I will analyze these four conditions for moral agency and show that some animals can meet the first two conditions based on Rowlands' own accounts of intentional state ascription, practical wisdom and normative practices. I will show that the last two conditions should be rejected as being necessary for ascribing responsibility, and are therefore unnecessary for moral agency. Finally, I will argue that Rowlands' latter two conditions for moral agency are better applied to the concept of moral autonomy, which is distinct from the concept of moral agency. Rowlands claims that these two concepts are interchangeable, and therefore a moral agent must be morally autonomous. I will argue that moral agency requires moral responsibility, but not necessarily moral autonomy, if moral autonomy is understood as meeting all four of Rowlands' conditions for moral agency. I will provide *prima facie* accounts of moral agency and moral autonomy opposing those of Rowlands.

Rowlands argues that (1) Some animals act for moral reasons; and (2) Moral responsibility is sufficient for moral agency. Assuming (1) and (2), I argue that (3) Acting for moral reasons entails moral responsibility (argued in Section one); (4) Moral agency does not require moral autonomy; therefore (5) Some animals are moral agents, if moral agency is

understood as acting for moral reasons and being morally responsible for those actions. Because he denies (3) and (4), Rowlands denies that some non-human animals can be moral agents.

## **2.1 A metacognitive account of moral agency**

In Rowlands' view, moral agency requires freedom of action, and freedom of action requires metacognition. To hold an agent morally responsible, Rowlands assumes there must be a reasonable degree of choice or freedom of action. He characterizes this freedom as one's possessing choice mechanisms that translate beliefs about alternatives and desires into plans of actions to realize those desires, and those mechanisms are responsive to one's reasons. He claims that some animals can have such choice mechanisms but not be moral agents because freedom of action depends on one's understanding of one's motivations and actions, and such understanding requires the capacity for metacognition that animals lack. Although some animals can make choices, none can be held accountable because they cannot sufficiently understand what they are doing or the consequences involved. On Rowlands' view, sufficient understanding involves metacognitive abilities that are beyond the scope of non-human animals, so he commits to denying the status of moral agency to all of them. Such metacognitive abilities can include: ascribing mental states to others and to yourself; evaluating those states in relation to your beliefs and desires, including your normative beliefs; being able to represent your reasons for making particular evaluations; and further representing and evaluating your reasons for those reasons. I will argue that the implication of making these abilities necessary for moral responsibility is that it becomes too intellectually sophisticated for most normal adult humans to qualify. I will say more about this in Section 2.4, where I argue that moral agency and moral autonomy are distinct concepts, and their conflation would result in absolving too many normal adult humans of moral responsibility.

Rowlands lays out four conditions that a moral subject must meet in order to be considered a moral agent and insists that meeting these conditions requires metacognition. I will look at each condition in turn to see if it is plausible that they are not necessary for moral agency or if they could be met without metacognition.

## 2.2 Rowlands' first condition for moral agency

**Condition (1): A moral agent has the ability to make qualitative distinctions between states to which one is sensitized (e.g. nervous anticipation versus anxiety). Such distinctions require scrutiny of the state, including its phenomenological profile and its causes.**

The idea in Condition (1) is that to be morally responsible, one must be able to determine if states of suffering/happiness are justifiable regarding desert and intensity, and be able to consider local versus global interests. A moral agent must be able to determine if someone's suffering or happiness is deserved in order to decide whether or not we should have sorrow at their suffering or joy at their happiness. Perhaps the person deserves the opposite – should we find joy in his suffering, if for example he was shot while executing innocent people? Taken even further, should the state execute such murderers, and if so, should we be concerned about their suffering? Not a very enlightened view perhaps, but a moral agent must be able to make such determinations on Rowlands' view. A moral agent must be able to determine if the intensity of one's suffering or happiness is justifiable. For example, a person might have a terrible fear of clowns and go into absolute hysterics at the mention of a circus, and perhaps even behave badly in his panic. This person's suffering might be intense, but is this intensity warranted? In this case, moral responsibility would require that an agent understand the complexities of phobias to justify the intensity of the suffering involved, and any ensuing behaviors, such as being abusive to clowns. A moral agent must be able to consider local versus global interests, and determine if one's suffering is justified by the gains it might bring. For example, one must judge whether the

suffering induced by chemotherapy as a cancer treatment is warranted to protect a person's global interests in having a longer life. Should chemotherapy be forced on someone? What about someone who cannot understand global interests, such as a young child or a cat? On Rowlands' view, moral responsibility is contingent on the ability to ask and effectively answer such questions.

Rowlands claims that scrutinizing states of suffering/happiness requires metacognition, and since no non-human animals have such capacities, they cannot meet Condition (1). However, he also claims that the practical wisdom required for a moral subject to "get things right" is context specific, and some animals are quite capable within their contexts. He further argues that some animals can learn the normative practices of their community, thereby becoming inducted into the practice, and such practices can include moral norms. I will argue that the degree of scrutiny of the states and causes of suffering/happiness required for moral responsibility is relative to the context of the actor.

Rowlands' argument that pertains to Condition (1) is:

- (1) Moral subjects (including animals) act for moral reasons, which include emotions regarding the welfare or fortunes of another (e.g. empathy with another's suffering/happiness).
- (2) A moral subject can be sensitive to morally salient features of situations which can include the emotional states of others that indicate suffering/happiness, and their potential causes.
- (3) Practical wisdom, or phronesis, is context dependent, varying from one social group (context) to another, due in part to the cognitive capacities of the individuals.
- (4) Some animals have practical wisdom.
- (5) Some animals belong to normative practices.
- (6) Morally salient features of situations are context specific, as not all morally salient features are possible in all contexts.
- (7) A moral agent must make sufficiently fine-grained qualitative distinctions among states and features to which one is sensitized.
- (8) Metacognitive abilities are required to make sufficiently fine-grained qualitative distinctions among the states and features to which one is sensitized.
- (9) No non-human animals have been proven to have metacognitive abilities.

- (10) Since no non-human animals have metacognitive abilities, no non-human animals can meet Rowlands' first condition for moral agency.

I have two objections to this argument:

(a) If premises 1 to 6 are true, then some non-human animals can make sufficiently fine-grained qualitative distinctions among states and features to which one is sensitized (premise 7). And if some non-human animals can meet the required degree of sufficiency, then this challenges premises 8 and 10. I will argue that context specificity can extend to making qualitative distinctions, so some animals can make sufficient distinctions within their contexts. (b) I challenge premise 8. I will argue that metacognition is not required to make sufficiently fine-grained qualitative distinctions among the states to which one is sensitized.

*(a): Context specificity can extend to making qualitative distinctions, so some animals can make sufficient distinctions within their contexts.*

Premises 1-6 of Rowlands' argument pertaining to Condition (1) are supported by his account of normative behavior in non-human animals in that normativity is context-specific. He argues that normativity is essential to the concept of morality, so acting for a moral reason requires some form of normativity. It is essential that his moral subject gets things right often enough to indicate that he knows how to respond in the right way to the morally salient features of situations. The moral action of moral subjects is not merely caused behavior. It is not by accident that a moral subject keeps getting things right, rather he knows how he should behave. He attributes normativity to animals in terms of their capacities for context-specific phronesis, and induction into normative social practices. A normative social practice is context-specific by definition.

Rowlands identifies the ability to belong to a normative practice with the ability to adjust one's use of a sign in accordance with the norms of that practice, and argues that some non-

human animals have these abilities. Language and some other kinds of communication require such normative practices. He uses vervet monkeys and dogs as examples of non-human species that can use signs normatively. The monkeys use distinct alarm calls to signal distinct threats, and the dogs use a play bow to signal that ensuing behavior is not intended to be serious. He says it is implausible to suppose that animals do not belong to normative communicatory practices. Communication is essential to the coherence of social groups and members are socialized to use the communication practices correctly. There are normative standards governing the correct use of signs, which are normatively assessed and sometimes upheld through sanction. For example, juvenile vervet monkeys often show less discrimination and make errors with their alarm calls. As they mature, they learn to better discriminate and make fewer errors, due in part to being corrected by older monkeys. Some animal communication shows that some animals can be part of a practice and act for reasons. If we extend these phenomena to include moral reasons and practices (e.g. those involving an injured group member), then it is plausible that such animals can be socialized to make qualitative distinctions among the states of others within their context-specific social practices.

In his discussion of phronesis, or practical wisdom, Rowlands argues that animals can “‘keep getting things right’, even in the absence of the sort of practical wisdom that is required for a human to do so” (Rowlands, 2012. p.139). Since Rowlands claims that phronesis is group relative, the sort of practical wisdom required for a human to get things right must be the human sort. I take this to mean a more intellectually sophisticated sort of practical wisdom that utilizes metacognitive abilities. But since phronesis is group relative, what counts as phronesis varies from one context to another, from one social group to another, due in part to variations in cognitive capacities. In other words, practical wisdom is context-specific. The sorts of morally

salient features that an animal can respond to in a situation (e.g. the emotional states of others), are just the sorts of features that are possible in her situation or context. For example, deception cannot be a morally salient feature in a context where deception is impossible. If a social group is made up of individuals incapable of deception, then it cannot be a feature of moral situations in that social context. If context-specific phronesis helps its possessor negotiate and navigate her way through her particular social world, then it seems implausible that she could do so successfully without making qualitative distinctions based on phenomenological profiles and causes within her context.

States and features of situations are identified by their qualities. It is implausible that one can be sensitive to states or features, but not to the distinctive qualities that distinguish them from one another. This is not to say that we must be able to sense or perceive any and all qualities in order to make any distinctions, but Rowlands describes non-metacognitive moral subjects as having a binary sensitivity to states and situations: they are positively or negatively sensitized to happiness/suffering, good/bad. If this were the case, their social competence would be limited to such a degree that the complexities characteristic of social interactions in many animal groups would be inexplicable. If one is sensitized to some states and features, then it follows that one is sensitized to some of the qualitative aspects of those states and features. If sensitivity to states and features is context-specific, then sensitivity to the qualitative aspects of those states and features is also context-specific. It stands to reason that non-metacognitive moral subjects can make qualitative distinctions among the context-specific states and morally salient features of situations to which they are sensitized.

For example, Rowlands' dog can distinguish between different consequences of the same behavior directed at different individuals. He distinguishes between a human adult and child



regarding the amount of pressure he exerts in a play-bite which suggests that he understands that a given bite-pressure can result in different amounts (and perhaps kinds) of suffering in different situations. Rowlands could respond that this is a quantitative distinction (amount of suffering as a function of amount of bite pressure and the size of the individual bitten). But I argue that the dog is making qualitative distinctions because adjusting their behavior to the capacities of playmates would require judgments other than those of quantity; it requires assessing the qualities of the playmate. For example, a dog that is playing with another dog that is smaller than the child may play rougher with that dog, despite the size of objects, thereby making qualitative distinctions between a human child and small dog. He may not even engage in such play with a fearful person, thereby judging attitudes about engaging in play. This is not to say that he must attribute mental states to his potential playmate. He need only gauge how that playmate is behaving. Another example is that a dog can distinguish between a play-bite and a warning-bite delivered with the same inhibited force. If he can make a qualitative distinction between a bite delivered as a threat and one delivered as play, this is largely due to learning how to interpret the signs of the practice to which he belongs (e.g. play-bow preceding the bite) and thereby making a normative practical judgment. The dog learns to make qualitative distinctions that make sense in the context of his form of life – it is this form of life that Rowlands claims is fundamental to induction into a moral practice. Anxiety may be within the dog's context, but perhaps not nervous anticipation. He might not make a qualitative distinction between these kinds of suffering. It is plausible that both grief and fear are within his context, so he might make a qualitative distinction here. The extent to which some animals could make qualitative distinctions between states and features to which they are sensitized, is the extent to which they

can assess the phenomenological profile and causes of these states and features within their group-relative contexts.

Rowlands objects to the idea of group-relative (species or context -relative) moral agency, so he does not intend that his conditions for moral agency be construed as group or context-relative. His objection to such a view is that it depends on the claim that what is moral action for one may not be moral action for another. Rowlands claims that when humans and animals are acting on the basis of morally-laden emotions (acting for moral reasons), they are doing the same thing and this type of action is not species relative. When a wolf and a human empathize and try to mitigate the suffering of another, they are doing the same thing. But how we express morally-laden emotions is most certainly context-relative (species-relative and no doubt culturally-relative) as are their eliciting causes, so they do differ in character. Although, the wolf and I may both act for the moral reason of mitigating the suffering of an injured pup, I would probably stroke her, while the wolf would probably lick her. I might also pick her up to take her to a vet, which would exponentially increase her local suffering, but hopefully increase her global happiness (not everyone would agree this is the morally right action), whereas the wolf may pick her up to take her to a quiet safe place. If these types of action can differ in character and still be the same thing on Rowlands' account (acting for a moral reason), then it is consistent that when humans and animals make context-specific qualitative distinctions, which may differ in character, they are doing the same thing. When a human or a dog is deciding whether an aggressive action of another is hostile or playful, they are doing the same thing – that is, they are both making qualitative distinctions about the phenomenological profile and cause of the state of the other.

Condition (1) requires that a moral agent must make qualitative distinctions between the states of others, including their causes and phenomenological profile. Some animals can meet this condition according to Rowlands' own arguments. His arguments for context-specificity allow that qualitative distinctions are context-specific. His arguments that moral action is identical across species (despite variation) allow that qualitative distinction is identical across species (despite variation). Since his arguments support premises 1-6, then it follows that his arguments support the claim that a moral subject, including some animals, can make sufficiently fine-grained qualitative distinctions, as per premise 7.

Rowlands could reply that moral responsibility, therefore moral agency, requires a degree of fine-grained distinctions that requires metacognition, of which only humans are capable. Condition (1) could require that a moral agent ask the sorts of questions and make the sorts of distinctions that require theory of mind. He could insist that a moral agent must be able to make abstract causal connections that require the cognitive abilities for abstract reasoning to a degree to which only humans are capable. I accept that having theory of mind and sophisticated abstract reasoning abilities would allow for greater moral autonomy, but I deny that moral autonomy is necessary for moral responsibility. Rowlands requires moral autonomy for moral responsibility, and insisting on finer and finer grained distinctions leads us into the sphere of moral autonomy. The problem with making responsibility dependent on moral autonomy is that too many normal adult humans might fail to qualify due to the inescapable contingencies that affect their capacity for moral autonomy. I will develop this argument in Section 2.4. Further, I will argue in Section 3 that responsibility must be conferred within a practice, so the degree of fine-grained distinctions of which some humans are capable only apply to human moral practices. I see how this could be question begging as it stands, but I hope that my further arguments will help to

answer the question of whether or not non-human animals can be morally responsible. The point to take away here, is that some non-human animals can make qualitative distinctions among the states to which they are sensitized and some of their causes, within the context of their moral practice. The question of how fine-grained these need be for moral responsibility depends on what one demands for moral responsibility, and that is the on-going question herein.

*(b): Metacognition is not required to make qualitative distinctions.*

Objection (b) challenges premise (8) of Rowlands' argument: Metacognitive abilities are required to make fine-grained qualitative distinctions between the states to which one is sensitized. Rowlands maintains that metacognition is not required for a moral subject to distinguish between the states to which she is positively or negatively sensitized, despite being able to make normative practical judgments regarding states and morally salient features of situations (explicated above). However, without metacognition, it is not clear to him how a moral subject could distinguish among kinds within general categories of suffering or happiness, and scrutinize the states and their causes. In other words, metacognition allows one to explain the behavior and situations of others and self. If we are going to choose a moral action by making judgments about desert, intensity, local/global interests, or justification for states of suffering/happiness, then we must be able to seek and construct adequate explanations. For example, the ability to make intellectually sophisticated fine-grained distinctions about the mental states of others is crucial in our legal system for seeking and constructing explanations that determine the level of criminal responsibility one might have. Understanding that a person feared for his life, and why he might have feared for his life, and that he may have held a false belief, and that his false belief could be justified, can determine whether that person is responsible for murder or manslaughter. Although ascriptions of legal responsibility and moral

responsibility are certainly not identical, the analogy illustrates how importantly the ability to explain and justify behavior is related to making fine-grained qualitative distinctions between the states to which one is sensitized.

The implication of the metacognition requirement is that theory of mind is necessary for the level of explanation required for moral agency. To responsibly explain or understand the behavior of others, we must be able to think about the mental states of self and other, including intentional attitudes (beliefs, desires, and, on Rowlands' view, emotions with intentional content). Having theory of mind allows us to understand that intentional attitudes drive behavior, and that intentional agents act for reasons. This is the force of Rowlands' first condition – without theory of mind, we cannot explain or justify behavior, which is central to moral agency, and there is currently no scientific proof that any non-human animals have this capacity.

There are two problems raised here: Is theory of mind required for explanation seeking and construction? Is theory of mind required for the level of explanatory sophistication required for moral agency? We can solve the first problem by considering Kristin Andrews' pluralistic account of folk psychology (Andrews, 2012). She argues that theory of mind is only one of several routes humans take in explaining behavior. The standard view of folk psychology is that one attributes specific mental content and attitudes to the target, then uses a folk psychological theory (or mental simulation or both), and finally generates a prediction or explanation.

Propositional attitudes are central to the standard view because they are considered to be the cause of all intentional behavior. Andrews offers an alternative – although we may sometimes attribute beliefs and desires to the target when we predict or explain behavior, it is not necessary to do so. In our folk psychological practices of predicting, explaining and justifying behavior, “we more often think about other people in terms of their emotions, personality, moods, past

behavior, and experiences as well as our expectations about what the individual should do given her role in society, her group memberships, and her cultural norms (Andrews, 2012, p.5). We come to expect certain behaviors in certain contexts, and need not attribute beliefs and desires to predict that the professor will show up for class at 10:00 am, for example. If she fails to meet our expectation, then we might seek to explain the anomalous behavior in a variety of ways that need not involve attributing propositional attitudes (e.g. she always takes the bus, so perhaps her bus is late).

Learning social norms entails the possibility of violating such norms, and those who punish such violations implicitly recognize that the norm has been violated. In Rowlands' discussion of moral practice, he argues that some animals can belong to normative practices, taking vervet monkeys as an example. While learning the communicatory norms of their practice, young monkeys make mistakes or violate the norms of communication, and are corrected with sanctions. The older monkeys who punish such violations implicitly recognize that the norm has been violated. Such violations could lead to explanation seeking that is not dependent on metacognition. Andrews' cites another example of a normative practice wherein capuchin monkeys insert fingers into each other's nostrils or mouths, or suck on each other's tails, which might be a ritualistic bonding practice that fosters trust in that it involves the risk of harm and violation would result in injury and distress. The practice requires coordinating their bodies to touch in ways that involve risk, taking turns and role switching, which Andrews says suggests that the players are following well-developed rules. If we assume that monkeys are not capable of the metacognition required to attribute representational beliefs to others, then the development of such norms suggests that they can develop expectations of normal behavior. Further, some animals respond to alarming or disruptive situations in ways that suggest they are

seeking to know why. Chimpanzees will run toward a screaming cohort to see what the trouble is. At least behaviorally, they are seeking to know why. Andrews argues that this may be explanation seeking, as opposed to information seeking, based on the agent having “a puzzled affective state or ... conflict between the situation and the appropriate behavior from the animal’s social repertoire”, which can drive the explanation seeking (p.242).

Does Rowlands’ concept of moral subject allow that they might seek explanations? Andrews describes three features of satisfactory folk psychology explanations: (1) they are constructed in response to an affective tension, such as a state of fear, puzzlement, etc. about a person or behavior, and this tension drives the explanation seeking behavior; (2) explanations resolve the tension that drives the behavior, promoting a feeling of satisfaction; (3) they are believed by the explanation seeker. Given that a moral subject’s empathy is not merely that of emotional contagion, and she responds to the good/bad making features of situations, and is disposed to promote/mitigate the good/bad, it seems that these features could describe what a moral subject might do. Smith is screaming; Rufus perceives this as suffering; becomes distressed and is driven to reduce Smith’s suffering (satisfying feature (1)). Rufus is a moral subject so he gets it right a significant amount of the time. To get it right, not only must there be a reliable connection between the perception and the emotion, but his disposition to promote/mitigate the good/bad allows that he might be driven to discover connections between the affective state of the other and the good/bad making features of the situation (satisfying features (2) and (3)).

Andrews’ account of explanation does not require metacognition, and explanation could be accomplished by one with emotional sensitivity and the capacity to recognize social norms. Rowlands argues that animals can have these properties, so his arguments allow that they can

seek and construct explanations on Andrews' account. He need not deny that a moral subject can seek and construct explanations without metacognition, which allows that one might make adequately fine-grained qualitative distinctions among the states of others and their eliciting causes. This brings us to the problem of what might constitute an explanation adequate for moral agency.

Is theory of mind required for the level of explanatory sophistication required for moral agency? Rowlands might insist that moral responsibility, therefore moral agency, requires the sophisticated level of scrutiny and explanation that is apparently unique to humans and dependent on our metacognitive abilities. Perhaps we must imagine what another might think; what their motives might be, based on their beliefs and desires; and how they may have been coerced by the beliefs and desires of a third party. Further, we need to be able to follow a causal chain and consider remote causes. Otherwise, how could we blame the person who hires the contract-killer? This sophisticated level of scrutiny and explanation may very well be required for full responsibility within a human moral practice because of what is possible within our contexts, due in part to metacognition. But non-human animals do not have guns for hire. A lack of metacognition and having cognitive abilities that differ from those of humans do not preclude all other animals from being able to scrutinize states, situations and their causes to a degree sufficient for ascribing responsibility. Scrutiny and explanation need only be as sophisticated as the context demands. It may be sufficient to recognize and sanction norm violations related to moral features such as harm.

Rowlands could reply that the practices in which animals engage are too simple to involve ascriptions of responsibility. In Section 3, I will argue that some non-human animal social practices involve differing levels of tolerance for norm violations depending on the age of



the individuals. Infants and juveniles, with their less-than-fully developed cognitive capacities, are not held to the same standard of normative behavior as are older individuals. The fact that the application of normative sanctions accords with levels of competence indicates that there might be some concept of responsibility, and if this is the case, then the application of normative sanctions may sometimes actually be ascriptions of responsibility. I will further develop this line of reasoning in Section 3.

Andrews' conceptualizes folk psychology as constituting a social competence in identifying, predicting, explaining, justifying, normalizing and coordinating behavior. Scrutinizing the phenomenological profile and eliciting causes of the emotional states of others is one way of attempting to explain their behavior. Social competence involves explanation, and those explanations are sought and constructed in specific contexts. Any distinctions among emotional states (or other factors) need only be as fine-grained and sophisticated as the context demands for social competence, and if one's social competence requires moral competence (Rowlands claims that some animals can belong to a moral practice), then achieving such competence is sufficient for moral responsibility within a moral practice.

The idea that social competence is context specific is consistent with Rowlands argument that intentional attitudes are context-specific and anchored to the actor's form of life (explicated in Section 1). He argues that moral sensitivity is context specific: there is no need to detect exaggeration of one's suffering in contexts and forms of life where exaggeration is impossible. Desert is also context specific. Andrews cites evidence that chimpanzees become more impatient with those who are unable to give them food than with those who are unwilling to give them food. This seriously challenges Rowlands' claim that it is in virtue of his metacognitive abilities that a moral agent can make the fine-grained qualitative distinctions required in Condition (1).

The different responses to the caregivers' situations suggest that chimpanzees can make qualitative distinctions within the category "withholding food", are able to "explain why" a caregiver is withholding food, judge whether or not the behavior is justified, and act accordingly. If it is true that no non-human animals are capable of metacognition, then these chimpanzees must be using non-metacognitive abilities to make these judgments.

Rowlands' arguments support the claim that some animals can meet his first condition for moral agency: intentional attitudes can be ascribed to animals; animals can act for moral reasons; animals can belong to normative practices; and practical wisdom is group relative. I have argued that Condition (1) can be met without metacognition. If I am right, then non-human animals that can act for moral reasons (moral subjects) can meet Rowlands' first condition for moral agency.

### **2.3: Rowlands' second condition for moral agency**

**Condition (2): A moral agent has the ability to have moral knowledge and know moral facts. A moral subject may be sensitized to the good- and bad-making features of situations that underlie his motivations, but only a moral agent knows that a motivation is a good or bad one. He must be able to reflect on the motivations, not to control them, but to deem them good or bad. He must be able to think "M is wrong".**

Rowlands' second condition for moral responsibility requires that a moral agent has the ability to deem a motivation morally right or wrong. I will show that animals that meet Rowlands' criteria for moral subjecthood can satisfy this condition. The idea in Condition (2) is that to be morally responsible, one must know the difference between right and wrong, and this requires metacognitive reflection and propositional mental content. The implication is that moral knowledge and knowing moral facts requires having concepts of morality and right/wrong. If one has these concepts, then one can have a belief of the form "I know that M is wrong". I interpret Condition (2) as requiring the ability to have beliefs about what is right or wrong, but not that such beliefs must be true. Rowlands does not require that a moral agent be infallible, so he seems

to use “believe” and “know” interchangeably here. The thrust of Condition (2) is that moral responsibility requires metacognitive moral reasoning - one must be able to represent to herself that a motivation is a morally good or bad one. Rowlands’ argument that pertains to this condition is:

- (1) A moral agent must be able to deem a motivation good or bad.
- (2) To deem a motivation good or bad, one must have the propositional content that “M is wrong”.
- (3) Reflective metacognition is required to have the propositional content that “M is wrong”.
- (4) Since no non-human animals are capable of metacognition, no non-human animals are moral agents.

My objections to Rowlands’ arguments for Condition (2) are: (a) Metacognition is not required to know (or deem) “M is wrong”. I will reject premise 3, and argue that affect can carry the information required to make such judgments and that they only require first-order states, thus leaving premise 4 unsupported. (b) If some animals are intentional agents, and can have context-anchored intentional attitudes about the morally good/bad making features of situations, then they can have the non-linguistic propositional content “that M is wrong”. Rowlands claims that some animals can be moral subjects, with intentional attitudes that have evaluative content which reflects their sensitivity to moral facts. If this is the case, then they can have the properties listed in premises 1 and 2. I will show how his arguments are inconsistent with the claim that no non-human animals can do so. I will also present a plausible explanation for how one might deem a motivation right or wrong without metacognition, based on the work of Peter Carruthers, thus challenging premise 3, leaving 4 unsupported.

*(a) Metacognition is not required to know (or deem) “M is wrong”*

Since a motivation consists of mental states (beliefs, desires), it seems that one must be able to think about those mental states at the meta-level to assess them. If we assume that all

non-human animals are incapable of metacognition, then how could some non-human animals deem a motivation right or wrong? There is a way to explain how one might judge motivations without having metacognitive abilities, which I will base on Peter Carruthers' (2008) model. He characterizes metacognition as a cognitive "executive" that oversees thought or problem solving. They are mental activities that occur at a higher meta level rather than those at a lower object level during cognitive processing. He looked at several scientific experiments which could be interpreted as showing that some non-human animals are capable of metacognition, in that they monitor their own uncertainty or confidence about the correctness of their own response to a test (e.g. determining which line is longer, and placing a 'bet'). Carruthers claims that the data can be entirely accounted for in first-order terms, appealing to states and processes that are world-directed rather than self-directed. He posits a cognitive "gate-keeping" mechanism that weighs the strengths of competing beliefs and desires in making judgments, whereby it is not necessary that one represents the fact that one has those judgments.

Humans tend to conceptualize and report some states, such as surprise or uncertainty, in metacognitive terms. Carruthers warns against assuming that such metacognitive categorizations are metacognitive states, thereby pushing the metacognitive character downward into the state. While the categorization is metacognitive, the phenomenon itself can be entirely first-order. Consider the mental state of surprise: a state that is caused when a perception gives rise to conflict with a prior belief or expectation. Some consider it to be necessarily metacognitive (e.g. Donald Davidson, 1982). Surprise may appear to be metacognitive in nature because it involves coming to believe that one of your beliefs is false, and this presupposes metacognitive thoughts about your own belief states. However, Carruthers claims that all that is required is that some mechanism is sensitive to conflicts between the contents of one's occurrent judgments. When

conflict arises between a belief and a perception, there is a suite of reactions (heightened alertness, widening of the eyes, etc.) Detection of these reactions constitutes the feeling of surprise. Rather than a metacognitive comparison of one's beliefs before and after the surprising event, surprise is an affective state which requires that one has beliefs, not that one has thoughts about one's belief states. Any affective or emotional state carries information about the occurrence of the associated mental states. For example, fear carries information that a thought of danger has occurred (real or imagined) but the fear itself is not metacognitive – it is first order. It involves perceiving or imagining fearful stimuli (snake), belief that a snake might hurt me (whether learned or instinctive), and a desire to not be hurt.

Carruthers' first-order explanation for the data in the uncertainty monitoring experiments is as follows:

- (1) Beliefs and desires come in different strengths.
- (2) Different goals compete to control behavior.
- (3) Some animals have a "gatekeeping" mechanism that is sensitive to the strengths of beliefs and desires.
- (4) When beliefs and desires involved in contrary actions are too close to one another in strength, the gate-keeping mechanism causes an affective state (i.e. the feeling of uncertainty).
- (5) The affective state carries the information that drives behaviors (e.g. looking closer, getting a different viewpoint, etc.)
- (6) The resultant action is a function of strengths of beliefs and desires interacting with one another in accordance with plausible principles of practical reasoning.

The posited gate-keeping mechanism "compares" the competing beliefs and desires without the need to represent them. It is sensitive to one property of intentional mental states – their strength.

How does Carruthers' model apply to Rowlands' second condition for moral agency, whereby one must be able to deem a motivation right or wrong? His model provides a plausible explanation for how an animal might deem or think that a motivation is right or wrong without

metacognition. Beliefs and desires about the considered action or state-of-affairs (the motivations) cause other mental states (assuming mental causation) such as emotions. Affective states carry information about the motivation, in the same way that the feeling of surprise carries information about beliefs and perceptions. One motivation might feel bad or wrong (as fear or anxiety), another good or right (as contentment or relief). The action will depend on the strength of the interacting beliefs and desires. When they are too close in strength, the gatekeeping mechanism will cause the affective state of uncertainty. An animal uncertain about the right course of action might hesitate or seek more information. Perhaps she will test the action out in the group and the reactions of others will strengthen one set of beliefs and desires while weakening another. The more she learns about her beliefs and desires in these ways, the more she is able to deem them right or wrong, without ever having to represent them as such.

Consider an adult wolf who has learned the normative practices for sharing food in her pack. She may desire to steal food out of turn, but she anticipates that stealing food will incur physical or social sanctions (e.g. biting or shunning), so she feels fear or distress. Her emotion carries information about her motivation. In addition to desiring the food, she desires to be accepted, or not to be punished, etc., and she believes that if she refrains from stealing the food, she will be accepted and not be punished, which may feel like contentment or relief. One motivation feels bad or wrong, the other good or right, depending on the strength of the interacting beliefs and desires. She can judge her motivation to steal as right or wrong without having to represent this judgment, and she can act to achieve the 'right' goal.

To be capable of judging their motivations as right or wrong on Carruthers' model, it is sufficient that animals are capable of emotions, learning, memory, intentional attitudes and social relations. Rowlands ascribes these sufficient capacities to animals that can act for the reason of

morally laden emotions. If he is to deny that such animals can deem a motivation right or wrong, he must do so on other grounds. Rowlands could argue that what the wolf is doing does not constitute deeming a motivation right or wrong in the moral sense, but is simply deciding what is most appealing or prudent for her to do based on occurrent beliefs and desires. This takes us to my second objection.

*(b) If some animals are intentional agents, and can have context-anchored intentional attitudes about the morally good/bad making features of situations, then they can have the non-linguistic propositional content “that M is wrong”.*

Rowlands claims in Condition (2) that one must be able to represent to herself that a motivation is good or bad. In other words, deeming a motivation good or bad requires that one have the concept “motivation”. I can criticize this argument in the same way that Rowlands criticizes the Davidson-Stich argument against animals having beliefs. He takes an approach in Condition (2) not unlike the Davidson-Stich approach to belief, which stipulates that one must have the concept “belief” to have a belief. In other words, having a belief requires that one can represent beliefs as beliefs (discussed in Section 1), thus having beliefs requires having metacognitive abilities. Rowlands’ rejects the Davidson-Stich model of belief and makes a case that we can ascribe beliefs to animals that are assumed to lack metacognitive abilities. A dog that is sensitive to the fact that a squirrel ran up a tree, but is unaware the squirrel has escaped, acts accordingly. We can explain the dog’s barking up the tree by ascribing to her the context-anchored *de dicto* belief that a squirrel is up the tree when the truth of the human-bound proposition tracks the truth of the dog-bound proposition. I will now turn this argument on Rowlands’ claim that animals lacking metacognition cannot deem a motivation right or wrong.

We can combine Rowlands’ argument that some animals can have beliefs with Carruthers’ first-order explanation of uncertainty to show that some animals can make judgments

about their beliefs, deeming them true or false. How can we explain the behavior of a dog who stops barking up a tree, searches nearby areas, seemingly in squirrel-hunting mode, and periodically returns to bark up the tree with decreasing enthusiasm? Has this dog deemed her belief false or possibly false? Using Rowlands' model of intentional attitude ascription, we can ascribe to the dog the context-anchored belief that the squirrel might not be up the tree. Applying Carruthers' model of first-order surprise and uncertainty, the dog can deem or judge her former belief as false (or not certain) without representing any of her beliefs. She is acting on the affective states that the strength of her beliefs/desires cause, and on her concurrent beliefs about places where squirrels might hide. The memory of the squirrel running up the tree and what she has learned about squirrels plays a role in the dog's changing beliefs and actions. Further, the strength of her desire to pursue the squirrel will be influenced in part by the strength of her beliefs about the squirrel's location. If she believes she might not know where the squirrel is (uncertain), and she is not starving and so on, her desire to pursue it may wane, and she may decide that it is no longer worth acting on the motivation. This description is metacognitive, but the actual states and processes determining the dog's behavior need not be. Now let's turn to moral motivations.

Can some animals know or deem their motivations as morally right or wrong? We can use Rowlands' proposition tracking framework and his criteria for moral subjecthood to ascribe to an animal the knowledge that a motivation is morally right or wrong. If it is true that a moral subject is sensitive to objective moral facts in the same way that a dog is sensitive to the fact that the squirrel ran up the tree, then on Rowlands' account, we need a *de dicto* ascription that "M is wrong" in order to explain her behavior. In other words, she can have the belief that "M is wrong" without representing beliefs and motivations as such. Recall that we need not know the



exact conceptual content of the animal's belief, but it needs to cohere with the human-bound concept. Where the intentional attitude in question is the morally-laden emotion of a non-human moral subject, then coherence demands that both the human-bound concept and the animal-bound concept are necessarily moral.

An example may help to clarify my argument. Imagine a wolf who drools, whimpers and looks longingly at the meal which the alpha wolf is eating, yet she refrains from trying to take it. To explain her behavior, we can ascribe the human-context anchored *de dicto* proposition "I am motivated to take food from the alpha wolf but that is wrong." It may simply be that she is thinking that she will get savagely attacked if she steals the food because that is what she learned by pushing limits in the past, or by witnessing others do so. It may be that she learned how to behave as a pup through more gentle corrections and by observing adults. No matter how she learned, the thought of stealing the food now causes an affective state – one which feels bad. The bad feeling is set in a complex situation that involves social relationships, individual desires, and beliefs about consequences of certain actions. Part of the wolf's social competence involves having appropriate affective states in certain kinds of situations – call them moral feelings. Thinking of stealing the food makes her feel bad and this can translate into thinking (deeming) that "M is wrong". This thought will factor into her practical reason to determine her action.

We need not know the actual content of the proposition that the wolf entertains (the wolf-bound concept). It only needs to be asymmetrically connected to the truth of the human-bound concept that "M is wrong". To explain the wolf's behavior, we can say that the wolf is deeming her motivations as good or bad in wolf-bound conceptual terms. She is making a judgment about how to behave in a given situation, and she learned how to make such a judgment within her moral practice. How she learned and what motivates her fidelity to the norm is not what is at

stake here. Rather, the ability to know a moral fact and deem a motivation as morally wrong is at stake. A human child may learn something is morally wrong through corporal punishment, but this does not amount to thereafter refraining from the wrong action as an adult in order to avoid a beating. Even if this is his motivation for restraint, this does not preclude him from thinking about it in first-order world-directed terms and deeming it morally bad because he knows the moral fact that “M is wrong”. The same can be said for the wolf.

Rowlands’ could object that while the content we ascribe to an animal might be described in moral terms with the human-bound concept that “M is wrong”, the animal’s actual content would not involve moral knowledge unless he has a concept of morality. I have two responses to this objection: (1) the asymmetrical connection between the truth of the human H-bound and animal A-bound propositions necessitates that the latter content can be reduced to the former; and, (2) if an animal has morally-laden emotions that are reasons for his intentional action, then the animal must have context-bound concepts of moral right/wrong.

Response (1): The asymmetrical connection between the truth of the human H-bound and animal A-bound propositions necessitates that the latter content can be reduced to the former. We are assuming that non-human animals lack the metacognitive capacity to represent mental states. Therefore, we must also assume that animals who are intentional agents have propositional attitudes that are first-order and world-directed toward others, things, objects, states-of-affairs and events, and not toward the minds of self and others. Hence, the propositional attitudes in question are toward objective facts. When Rowlands’ proposition tracking model is used to ascribe propositional attitudes to non-human animals, then it is necessarily restricted to world-directed propositions. No matter how a dog may conceptualize the thing and state-of-affairs about which he has beliefs, the content refers to the identical object, state-of-affairs, etc.,

that the human conceptual contents do. Recall the proposition tracking formula: If H-bound  $p$  is true, then A-bound  $p^*$  is true (and not vice versa). They refer to the same cognitive content. That the squirrel is up the tree is either true or false no matter how the object and state-of-affairs are conceptualized, and since the truth of  $p$  tracks the truth of  $p^*$ , then  $p^*$  can be reduced to  $p$ . The chase able thing up there is identical to the squirrel up the tree. Now recall that Rowlands is committed to objective moral facts. Hence, a moral claim has cognitive content. If this is the case, then H-bound proposition  $mp$  “stealing is morally wrong” tracks the truth of the A-bound proposition  $mp^*$ , however it is conceptualized, and the latter reduces to the former. It follows from Rowlands’ argument for intentional attitude ascription that an animal that can act for moral reasons must be able to have the propositional content that “M is wrong”, without representing their beliefs and desires – their motivations.

Rowlands could reply that although we know that  $mp^*$  is true whenever  $mp$  is true, this is insufficient to show that the former can be reduced to the latter. There is no guarantee that our context-bound contents are sufficiently alike to underwrite ascriptions of responsibility. Recall that Rowlands set out to make the conceptual case that some non-human animals can act for moral reasons, and does not try to make the empirical case that there exist some animals that can act for moral reasons. My argument that  $mp^*$  can be reduced to  $mp$  should be considered accordingly. The point to take away is, if Rowlands’ method for intentional ascription is to be of any use, then the content of the context-bound propositions have to come to the same thing in the world of facts, which includes moral facts. Perhaps an identity reduction is too ambitious, but the propositional content must be similar enough so that the observed behavior and the ascribed intention are logically related.

Response (2): If an animal has morally-laden emotions that are reasons for his intentional action, then that animal must have context-bound concepts of moral right/wrong. Rowlands could object that while it may be true that the cognitive content of an H-bound moral proposition might be identical to the A-bound moral proposition, it might not be true that there is a moral character to the proposition for the animal in question. The animal does not necessarily have a concept of moral right/wrong. The wolf may simply learn that if she steals, she will get attacked, and has no propositional content that “M is wrong”. But denying that some animals can have concepts of moral right/wrong is inconsistent with Rowlands’ concept of a moral subject. The moral subject is an intentional agent who can act for moral reasons, and metacognitive capacities are not required. They are simply sensitive to the moral features of situations. However, they can act for the reason of their morally-laden emotions which are about these features. Recall Rowlands’ conception of a morally laden emotion (discussed in Section 1): An emotion, *E*, is morally laden if and only if (1) it is an emotion in the intentional, content-involving, sense, (2) there exists a proposition, *p*, which expresses a moral claim, and (3) if *E* is not misguided or misplaced (grounded in a false factual assertion or based on an erroneous assumption of entitlement), then *p* is true. Further recall that the intentional content of these emotions must be other-regarding and that being a moral subject involves mitigating the bad and promoting the good. It follows that for an animal to be a moral subject, she must have context-bound moral concepts of what is good and what is bad. The state-of-affairs of your suffering is bad and that of your happiness is good (as empathically perceived but not metacognitively represented). If the content of a moral subject’s intentional attitude is morally-laden, then it follows that there is conceptual content that is moral in character. If this were not the case, then the content of the

intentional attitude (including emotions) cannot be considered morally-laden and such an intentional agent cannot be considered a moral subject.

Rowlands claims that a moral subject need not entertain, or even be capable of entertaining a moral proposition, and all that is required is that the morally-laden emotion tracks the true proposition. If this is true, then he can deny that a moral subject needs to have moral concepts. I argued in Section 1 that reasons that are beyond the grasp of the agent's form of life (context) cannot be considered reasons for the agent's action, but rather causes of their behavior. Rowlands is committed to the claim that moral subjects act for moral reasons and this action is not merely caused behavior. It is normative intentional action. I can accept that an agent might act for external reasons, in that the reasons might be outside of her motivational set, but I cannot accept that the reasons are outside of the cognitive capacities of her form of life. If the reasons for action are morally-laden emotions that have intentional content about moral facts or features of situations, then the moral character of these emotions, facts and features must be an element of the conceptual content of a moral subject. Non-human moral subjects may conceptualize moral feelings, facts and situations differently than do humans, but the essence of "moral" (whatever that may be) must be retained throughout, or these reasons can no longer be classed as moral. Therefore, non-human moral subjects must have context-bound concepts of moral right and wrong in guiding moral action. When using practical reason to determine a course of moral action, the agent must deem their motivation morally right or wrong.

Finally, Rowlands could deny that non-human animals can deem a motivation right or wrong because to do so requires a judgment independent of one's motivational set. Since moral reasons are external for Rowlands, a moral agent must be able to deem a motivation as right or wrong regardless of her motivating beliefs and desires, and non-human animals may be

incapable of this kind of impartial reasoning. But our wolf has learned the norms of her moral practice, which are arguably moral facts and external reasons. Rowlands considers morally laden emotions to be moral reasons, and that they are emotions which have intentional content about moral facts or morally salient features of situations (external reasons). For example, the moral fact here is that stealing is wrong. The pack's normative sanctions against stealing indicate this moral fact. According to her social practice, the wolf knows how it is "wrong" to steal food, no matter how badly she desires it, or even if she believes she will benefit from so doing. She knows how she should behave, and knows the consequences of her actions one way or the other, despite her motivations. She has morally laden emotions about this moral fact, which carry information about her motivation to steal. She can have a fully conceptual thought with the non-linguistic propositional content that "M is wrong". The question here is not why it is wrong to steal – that is a matter for Rowlands' Conditions (3) and (4). It is sufficient for Condition (2) that the wolf knows the moral fact and deems her motivations morally right or wrong in relation to these facts.

#### **2.4. Rowlands' third and fourth conditions for moral agency**

I will analyze Conditions (3) and (4) together because the same objections apply to both. While these conditions are distinct, they both depend on the ability to know and understand the principles upon which a moral claim is based. Condition (3) requires an ethical analysis, while Condition (4) requires a metaethical analysis.

**Condition (3): A moral agent has the ability to understand why a motive or action is a good or bad one. This requires being able to scrutinize a motivation in terms of a preferred moral theory. An agent must be able to understand the basis of moral facts.**

**Condition (4): A moral agent must understand what makes something right or wrong. He must be able to examine the principle upon which the moral facts are based. He can examine their mutual consistency, their scope and applicability, and can reject or modify them accordingly.**

The idea in Rowlands' Condition (3) for moral agency is that a moral agent must not only know that a motivation is good/bad or right/wrong, but must be able to identify the principles upon which the moral facts are based. For example, a moral agent might understand that stealing is wrong in terms of universalizability – only do what you could will as universal law for all rational agents. Or one might understand the moral fact in utilitarian terms – do whatever promotes the greater good, and surely the greater good involves a sense of security about property. The sentimentalist might understand the moral fact in terms of empathizing with the suffering of a robbing victim and thereby disapprove of the motivation to steal. The contractarian may see stealing as violating an agreement (tacit or otherwise) – you don't take mine and I won't take yours so that we can live in peace. And so on. In other words, a moral agent must be able to conduct an ethical analysis of her judgment that some motivation or action is right or wrong.

The idea in Condition (4) is that a moral agent must not only know the principles upon which a moral fact is grounded, but must be able to examine and evaluate these principles. For example, a Kantian that considers stealing wrong may also deem it wrong for the powerful to exploit the powerless. So how should he judge the actions of an exploited starving sweat-shop worker who steals from her rich exploitive multinational corporate employer? Condition (4) requires that a moral agent is able to analyze a principle such as universalizability, and how it might apply here, and decide what is wrong or right by either rejecting the principle and deeming the worker's action justified, or retaining the principle and deeming the actions of the worker and employer universally and unjustifiably wrong. To do this, he must be able to justify why right/wrong action depends on principles such as universalizability, or justify why he should reject a Kantian approach and adopt another one. A moral agent must have sufficient

understanding of what constitutes right and wrong. In other words, a moral agent must be able to conduct an effective metaethical analysis of his ethical standards.

I have two objections to these conditions: (a) Rowlands' criticism of metacognitive accounts of moral normativity apply equally to his metacognitive account of moral responsibility, and (b) while Conditions (3) and (4) could result in a higher degree of moral autonomy, I will argue that moral responsibility does not require them, and that moral autonomy and moral agency should be conceptually distinct. I will further argue that the result of Rowlands' conceptual conflation erroneously denies the status of moral agent to many 'normal' adult humans (those whose cognitive functions are not significantly impaired).

*(a) Rowlands' criticism of metacognitive accounts of moral normativity applies equally to his metacognitive account of moral responsibility.*

Rowlands criticizes metacognitive accounts of normativity on the grounds that they appeal to a property at the meta levels which the first-order level lacks. He argues that such accounts fall into regress. He criticizes the claim that scrutiny allows control over motivations, thus conferring normativity. I will argue that his Conditions (3) and (4) are subject to the same criticism when applied to understanding and responsibility. If understanding the basis of moral facts and principles is required for moral responsibility, and this falls into regress, then many normal adult humans should be absolved of moral responsibility on Rowlands' model.

A general metacognitive account of moral autonomy is as follows:

- (1) To be morally autonomous, it is necessary that a subject has the ability to effectively, critically, and morally scrutinize her motivations and actions.
- (2) The ability for critical moral scrutiny requires the ability to reflect on reasons as one's own reasons for endorsing or rejecting a motivation.
- (3) Reflecting on reasons as one's own reasons requires metacognition.
- (4) Animals don't have metacognition.
- (5) Therefore animals don't reflect on their reasons.
- (6) Therefore animals don't engage in moral scrutiny.



(7) Therefore animals are not morally autonomous.

Rowlands accepts this account of moral autonomy and applies it to his account of moral agency. On his model, moral agency is conflated with moral autonomy. Moral agency/autonomy requires moral responsibility, and moral responsibility requires metacognition (herein I will refer to his account as a metacognitive account of responsibility). He denies that moral autonomy is necessary for moral action. If moral autonomy is necessary for moral action, then Rowlands' concept of moral subjecthood is in trouble. He argues that some animals can be moral subjects that can act for moral reasons, and that moral reasons are normative. However, he claims that some interpretations of metacognitive accounts of moral autonomy erroneously deny that animals can act for moral reasons because such accounts attribute normativity to self-government, or control (herein I will refer to such accounts as metacognitive accounts of normativity). Rowlands must show that moral subjects can act for normative reasons without having moral autonomy. He denies that metacognitive moral scrutiny can imbue the motivations of an actor with normativity because a lack of control at the first order will be reiterated at every level of metacognition, hence such accounts of normativity fall into regress. Hence, the moral scrutiny that is necessary for moral autonomy is not necessary for normativity. Hence, the lack of metacognitive abilities does not preclude an animal from acting for morally normative reasons.

Rowlands' argument that metacognitive accounts of normativity fall into regress is:

- (1) According to metacognitive accounts of normativity, moral scrutiny confers normativity on the motivations of a subject via the medium of control.
- (2) To control her first order motivational states, a subject must control her higher-order evaluations and assessments of the states, and then control her evaluations and assessments of those evaluations and assessments, and so on, which results in regress.
- (3) If a subject's meta-level scrutiny of her motivations is clouded and shaped by contextual factors over which she has little awareness and control, then this scrutiny cannot imbue the subject with control over her first-order motivations because the lack of awareness and control over such contextual factors is reiterated at every level.

- (4) If the ability to engage in effective moral critical scrutiny cannot provide a subject with control over her motivations (due to regress), then it cannot be control that transforms the motivations into normative states.

I have two main goals here: first, I will challenge premise 1 and argue that many metacognitive accounts of normativity do not attribute normativity to control, but rather to understanding the moral facts and principles upon which they are based. For example, I interpret the Kantian as attributing normativity to the ability to value humans as ends in themselves. This ability confers normativity on motivations. Control follows from first understanding what one should do, and why one should do it. This is the same direction that Rowlands' account of moral responsibility takes. Second, I will tweak the above argument and show that it applies to Rowlands' metacognitive account of moral responsibility. When we replace 'normativity via control' with 'responsibility via understanding' the charge of regress still stands. Hence, if regress precludes moral autonomy from conferring normativity on the moral motivations of actors via control, then it precludes moral autonomy from conferring moral responsibility on the moral motivations of actors via understanding.

In his criticism of metacognitive accounts of normativity, Rowlands targets the notion that higher-order reflection can confer a property on motivations that they are presumed to lack at the first order. In particular, he takes aim at that notion that critical scrutiny can confer normative control on motivations, which is presumed to be lacking at the first-order. Typically, metacognitive accounts of normativity deny that first-order motivations have normative dimension or exert normative force. It is only via higher-order reflection that motivational states become normative. Rowlands argues that such accounts attribute normativity to a level of intentional control that is only achieved via effective critical scrutiny – normativity is thus conferred via control. He argues that such accounts are subject to a regress fallacy, as per

premises 2 to 4 above, and cites Christine Korsgaard's account of normativity as an example. I will argue that Korsgaard does not attribute normativity to control, but attributes it directly to self-reflection, thereby challenging premise 1. I will argue that her account of normativity points to understanding one's motivations in the same way that Rowlands' account of moral responsibility does. Hence, his account is equally subject to the criticism of regress as outlined in premises 2 to 4. Perhaps all four of his metacognitive conditions for moral agency could be criticized on the grounds of regress, but here it is most pertinent to Conditions (3) and (4) because I assume that they cannot be met without metacognition, whereas I have argued that Conditions (1) and (2) can be met without metacognition.

Korsgaard, following Kant, argues that morality requires a form of self-consciousness that she says is unique to humans. Her account of moral normativity is as follows:

- (1) As agents with free will, humans construct practical identities through self-conscious reflection on the grounds of our beliefs and actions.
- (2) Practical identities constitute our reasons for action and living.
- (3) We value from our practical identities.
- (4) Practical identity is only possible if we value ourselves as humans.
- (5) If we value ourselves as humans, and acknowledge that each of us is one among many, then we are rationally required to value other humans.
- (6) Our rationally constructed practical identity justifies the claims morality makes on us.

To act morally, one must be able to perceive and think about the grounds of beliefs and actions as grounds. On her view, a dog may be conscious of an object as fearful, and that is the ground of his action, but he cannot be conscious *that* he is fearful of the object and *that* he behaves a certain way as a result: "We are conscious of the potential grounds of our beliefs and actions as potential grounds" (Korsgaard, 2008, p.30). This is what allows the capacity for normative self-government that for her is the essence of morality and is unique to humans. She argues that because we humans value ourselves, we must value humanity, which makes moral identity a

necessary practical identity. We are rationally required to value ourselves, and if this is so, then we must be rationally required to value others because of the depth of our social nature. She does not mean that we are moral because we are social, but rather that our reasons are public and sharable, so the “justifications of morality can and should appeal to [our social nature]” (Korsgaard, 1996, p.136). The normative force of any consideration for action can be shared with others. In this way, we become obligated to one another by putting ourselves in their shoes. If you value your humanity, then you want others to respect and value it too, and if you want others to respect and value your humanity, then you must value theirs. On Korsgaard’s view, we are rationally required to empathize with one another, and can choose actions that we can will to be universal laws with normative force. We can act for a moral reason if and only if we can reflect on the possible grounds of our motivations.

Korsgaard attributes the normative dimension, force or grip of our motivations to our rationally constructed practical identity, which we construct by self-consciously reflecting on the grounds of our beliefs and actions. From this reflective construction, we get our reasons for action, which we can endorse or reject according to the value we place on ourselves as humans. Self-government (control) follows from valuing (normativity), not the other way around, as Rowlands claims. Hence, normativity is the result of critical scrutiny, and control is the result of normativity. We can control our motivations in virtue of the normative force conferred by the value we place on humanity via self-reflection. Korsgaard’s account of moral normative force is certainly metacognitive, but it is not so clear that metacognition confers control on motivations and via control, normativity. Rather, metacognition confers value and we endorse or reject our motivations, thus controlling them, based on our rationally constructed values. Hence, the ability to self-consciously reflect on the grounds of our beliefs and actions makes us morally responsible

via the ability to understand the basis of our rationally constructed values - the moral facts and principles. Rowlands' account of moral responsibility reaches the very same conclusion. I will argue that the charges of regress that he applies to metacognitive accounts of normativity equally apply to his metacognitive account of moral responsibility.

Let us return to Rowlands' argument that metacognitive accounts of normativity fall into regress. When we replace the property of 'normative control over our motivations' with that of 'understanding our motivations', the metacognitive argument still falls into regress:

- (1) According to Rowlands' metacognitive account of responsibility, moral scrutiny confers responsibility on the motivations of a subject via the medium of understanding.
- (2) To understand her first order motivational states, a subject must understand her higher-order evaluations and assessments of the states, and then understand her evaluations and assessments of those evaluations and assessments, and so on, which results in regress.
- (3) If a subject's meta-level scrutiny of her motivations is clouded and shaped by contextual factors over which she has little awareness and control (which affects her understanding), then this scrutiny cannot imbue the subject with an unclouded understanding of her first-order motivations because the lack of awareness and control over such contextual factors is reiterated at every level.
- (4) If the ability to engage in effective moral critical scrutiny cannot provide a subject with an unclouded understanding of her motivations (due to regress), then it cannot be understanding that confers responsibility on the motivations of a subject if we are to hold most normal adult humans morally responsible.

To understand her first order motivational states, a subject must understand her higher-order evaluations and assessments of the states, and then understand her evaluations and assessments of those evaluations and assessments, and so on, which results in regress. If a subject's meta-level scrutiny of her motivations is clouded and shaped by contextual factors over which she has little awareness and control, then she can little understand that of which she is little aware. She cannot understand that of which she is not aware, and she cannot deliberately control that of which she is not aware. If a subject's meta-level scrutiny of her motivations is clouded and

shaped by contextual factors over which she has little awareness and control, then this scrutiny cannot imbue the subject with a clear understanding of her first-order motivations because the lack of awareness and control over such contextual factors is reiterated at every level. Later in this section, where I discuss my second objection to Conditions (3) and (4), I will show why this regress is a problem if we are to hold most normal adult humans morally responsible.

Rowlands' arguments support these tweaks to his charge of regress regarding control. He explicitly denies that metacognition imbues motivations with normativity, whether it is control or understanding that is in play. Replacing control with understanding still results in the failure of metacognition to transform a motivation into a normative state. Since he assumes that no non-human animals have metacognitive capacities, and some animals can act for moral reasons, then it cannot be metacognitive understanding that makes a motivation normative. His thesis that some animals can act for moral reasons depends on those reasons having a normative status, or they could not be moral reasons. Recall that his account of normativity is based on an objective consequentialist account of morality (discussed in Section 1). Normative status is conferred on motivations by their possible accordance or discordance with objective moral facts. Normative force is not internal to the subject, but is an external feature of situations. Rowlands never attributes moral normativity to understanding, but he does attribute moral responsibility to understanding. However, when we replace 'normativity' with 'responsibility' in his criticism of metacognitive accounts of normativity, the charge of regress still applies. Rowlands claims that scrutiny cannot confer normativity on a motivation via control or understanding because the meta-levels are vulnerable to contextual factors over which a subject has little awareness and control. If he is right, then scrutiny cannot confer responsibility via understanding because of those contextual factors.

Let me be clear that I do not deny that metacognition is necessary to try to understand moral facts and principles. Rather, I oppose the claim that understanding the moral facts and principles is necessary for moral responsibility. Below, I will argue for objection (b) - that metacognition is required for moral autonomy, but not for moral responsibility. The contextual factors that cloud a moral agent's understanding of these facts and principles will be reiterated at every level, so the degree to which one's understanding is clouded affects the degree to which a moral agent is morally autonomous. I consider moral autonomy to be an ideal to which humans can aim, but never fully realize. As such, the regress in Conditions (3) and (4) do not preclude them as conditions for moral autonomy, as it is itself regressive due to our contextual factors. The more one is able to identify and understand such factors, the more autonomous one can become, but they will always prevent perfect autonomy. Ultimately, we are all constrained by being human – a contextual factor that is likely inescapable. Few, if any, can be fully aware of or understand all that underlies our motivations and moral principles.

Rowlands could reply that Conditions (3) and (4) only require that a moral agent can try to achieve a reflective equilibrium between his motivations and his evaluations of those motivations. It is sufficient that he can try to understand moral facts and the principles upon which they are based to be morally responsible for his actions. Rowlands criticizes this idea in relation to normativity rather than responsibility. The idea is that one can decide which motivations to act on and which to reject in virtue of metacognitive abilities. A metacognitive subject's motivations thus belong to the space of reasons rather than the space of causes. But Rowlands insists that a non-metacognitive moral subject acts for moral reasons, and not just on the basis of mere causes. A moral subject must be reliably sensitive to the good/bad making features of situations and have intentional content directed at those features, which guides the

moral action, whereas a moral agent must be able to reflect on the good/bad making features and his intentional content in order to understand them and be responsible. Such an argument places the motivations of a non-metacognitive moral subject squarely in the space of causes, and Rowlands' thesis is threatened if he accepts the result of this reply. His moral subject must be able to act for moral reasons, not just respond to moral causes.

Rowlands claims that the notion of metacognition conferring normative status on motivations is suspect because we are at the mercy of our motivations at every level, so he should accept that the same is true when it comes to metacognition conferring responsibility. He could reply that the regress problem does not preclude that Conditions (3) and (4) are necessary for moral responsibility, because moral autonomy is necessary for moral responsibility. If moral autonomy is regressive in nature, then its conditions can be regressive. I will argue that the results of Rowlands' model results in fewer human moral agents than is plausible, unless one accepts moral skepticism or hard determinism (which he denies).

*(b) While Conditions (3) and (4) could result in a higher degree of moral autonomy, moral responsibility does not require them. Moral autonomy and moral agency should be conceptually distinct. The result of conceptual conflation erroneously denies the status of moral agent to many 'normal' adult humans (those whose cognitive functions are not significantly impaired).*

On my view, moral agency is distinct from moral autonomy. My *prima facie* concept of moral agency is in keeping with Rowlands' concept of moral subjecthood, except that I take moral responsibility to be a requirement. It is sufficient for moral agency that one can act for moral reasons and is responsible for those actions. A moral agent knows how to act according to a tacit code of conduct, or moral practice, and can govern his behavior by it (Gert, 2012). The guiding code is moral in that it involves avoiding and preventing harm to others and sharing resources, which include food, territory, and reproductive resources. This might be considered a



descriptive account of morality, but that does not preclude moral responsibility. I argued in Section 1 that moral responsibility is necessary to morality, whether it is a descriptive or normative account. The latter moves us into the area of moral autonomy because the focus is on how the content of a moral code should be determined, but the skills necessary to make such assessments (e.g. Conditions (3) and (4)) are not necessary for moral responsibility. It is sufficient for moral responsibility that such a code exists; that its content is morally normative in the sense of exerting normative force on the individuals, society or group members involved; and that such individuals have sufficient cognitive capacities to know what kinds of actions their moral code prohibits, requires, discourages, encourages, and allows, and must have sufficient volitional ability to use the moral code as a guide for their behavior. Those lacking these characteristics are not subject to moral judgment (Gert, 2012). Much more will be said on this in Section 3, where I will provide a more detailed account and defense of moral agency.

My *prima facie* concept of moral autonomy is in keeping with Rowlands' account of moral agency except that I deny that moral autonomy is required for moral responsibility. A morally autonomous agent acts on motives, reasons, or values that are her own (Stoljar, 2014). To be autonomous is to be one's own person, to be directed by considerations, desires, conditions, and characteristics that are not simply imposed externally upon one, but are part of what can somehow be considered one's authentic self (Christman, 2011). In other words, an autonomous moral agent must be able to try to scrutinize and set himself apart from the contingencies of his existence - of place, culture, education, intelligence, personality, social norms (including moral norms), social relations, etc. We humans are necessarily fallible and vulnerable to our contingencies, so attaining perfect moral autonomy is an ideal that one can aim for, but likely never reach.

Rowlands denies that actualizing ideal moral autonomy is necessary for responsibility, but claims there is no responsibility without a significant “effective” capacity for autonomy. I do not oppose his concept of moral autonomy, but I oppose the claim that it is required for moral agency. Since Rowlands conflates moral agency and moral autonomy, every moral agent must be able to evaluate his motivations with an understanding that can try to separate itself from its contextual factors. I have argued that this separation is untenable, in that contextual factors are reiterated at every level of metacognition. However, that does not preclude some moral agents from making the attempt and, if reasonably successful, increasing their moral autonomy. I conceive of moral autonomy as the endeavor to understand morality and its content, to construct and critique moral theories and ethical standards, and it is most likely unique to humans. It is an ideal toward which some humans can aim. The cognitive and metacognitive abilities that allow us to do this are what allow our moral practices to change and develop in ways that the moral practices of other animals cannot. Conflating agency and autonomy, as Rowlands does, results in denying moral agency to many of those who are normally considered to be moral agents – “normal” adult humans.

Identifying and understanding moral principles requires cognitive and metacognitive capacities that may be unique to humans, but not all human moral agents are able to do so to a significant degree. For example, in the Kantian case, we must understand the moral facts as based on the principle of valuing rational agents as ends in themselves. But this leads to the question of why we should value ends in themselves. Many normal adult humans might not be able to understand whether or not Korsgaard’s account is question begging in that she presupposes a valuing practical identity as the place from which we value. I am willing to go so far as to claim that most humans could not do so. Korsgaard could respond by claiming this

principle as a primitive – we are valuable to ourselves in the first order, but we cannot act for moral reasons unless we know that we value ourselves – we must be able to reflect on the belief that we are valuable – we must know that we are valuable in order to value. But she argues that the valuing creatures must be constructed, and are constructed by valuing, and this falls into regress. According to Rowlands, moral responsibility requires that a moral agent whose moral principles are of the Kantian ilk, must be able to effectively scrutinize and understand such challenges to, and implications of, their principles. One would need to be able to ask and answer the question of why we should value ourselves and others. This is a tall order for many normal adult humans. I must admit, that even with formal training in philosophy, the best answer I can offer is “most of us just do”. I might then propose an explanation that an instinctive value of oneself in the first order might have evolutionary benefits and when one is a social animal, that value extends to others on whom you rely for survival. But the expert Kantian would deny that this constitutes valuing, so I have failed to understand the principle grounding the Kantian principle of valuing ends in themselves. The implication is that, since I sometimes use Kantian principles in my own moral reasoning and I do not understand the basis of these principles, then I must not be responsible for the motivations that I accordingly endorse or reject.

Perhaps I am being uncharitable and am taking Conditions (3) and (4) too far. In fact, Rowlands claims that a moral agent’s understanding can be far less than perfect. Moral responsibility may depend on having the cognitive ability to reason in the way illustrated above, but it does not depend on being very good at it. Rowlands says that moral agency is a matter of degree, but he also claims that all non-human animals (including those who are moral subjects) are clustered at the non-agent end of the spectrum and most adult humans are clustered at the agent end. He further claims that if it is a matter of choosing one category or another, he places

all non-human animals, very young humans, and severely cognitively impaired humans in the non-agent category with no moral responsibility. Since he conflates moral agency and moral autonomy, only morally autonomous agents are morally responsible. But just how autonomous must one be to be responsible? Rowlands says that it is only necessary that a moral agent's critical scrutiny of moral facts and principles be effective enough so that she is "not hopeless at it". Recall that a moral subject's sensitivity to the moral facts must be reliable, and this means getting it right somewhere between once and always. If Rowlands' criteria for the efficacy of scrutiny are as broad, then for a moral agent's scrutiny to be effective, she must correctly understand the basis of moral facts at least once. Rowlands' account of moral normativity depends on the claim that there are objective moral facts. A moral agent cannot be hopeless at understanding the basis of these facts, and because they are facts, it follows that they have true explanations. A moral subject is only morally responsible if and only if she effectively understands the basis of moral facts – she must get it right at least once, or more likely sometimes, in order to not be hopeless at it. I will argue that this might not be the case for many normal adult humans, especially in satisfying Condition (4) – rightly understanding the principle of principles.

For example, Rowlands' account of moral agency likely precludes normal pre-historic adult humans. Explicit moral principles are a more recent phenomenon for humankind, but it is implausible that we just as recently became moral agents (Shapiro, 2006). Basic moral precepts that forbid certain harms and promote helping others pre-date the rise of the first city-states (~4,500 BC), and are commonly found in hunter-gatherer societies. In all likelihood, these precepts have been present in such societies for at least 50,000 years (Campbell, 2013). Most of our existence was "pre-principled" in the sense that our ancestors were likely not able to analyze

the basis of the facts of their moral codes in any effectively impartial or unbiased way. Their understanding of these facts was probably grounded in emotion and superstition, despite their metacognitive capacities.<sup>3</sup> It is highly implausible that our cave-dwelling ancestors with modern human brains were morally autonomous, but it is even more implausible that they were not responsible for their moral obligations to each other, although the expression of which might look very different than in more recent agrarian or industrial societies. Pre-historic humans presumably had the requisite metacognitive abilities to satisfy Conditions (3) and (4), but the conceptual framework for these conditions probably did not exist. In the western tradition, the view that moral responsibility requires moral autonomy is a modern development: “Putting moral weight on an individual’s ability to govern herself, independent of her place in a metaphysical order or her role in social structures and political institutions is very much the product of the Enlightenment humanism” (Christman, 2011). Pre-historic humans may have been unable to think in terms of moral autonomy, but this would not have precluded them from moral responsibility within their communities.

Let’s look at some of the features related to cognition that are commonly considered to qualify a human as a morally responsible in current western society - old enough to know better, reasonably sane and reasonably intelligent. We do not hold toddlers to the same standard as teenagers, but we hold teenagers to the same standard as adults (albeit with more room for forgiveness). We do not hold an adult who has an insufficient grip on reality, or one who has the cognitive abilities of a young child, to the same standard as a “normal” adult. Rowlands would

---

<sup>3</sup> The received view in evolutionary psychology is that human brain anatomy has not changed in about 50,000 years. The oldest human fossil from which an entire genome has been extracted belongs to a man who lived about 45,000 years ago in Western Siberia. It is therefore plausible that the human brain structures required for metacognition were present in this man, and that these brain structures functioned the same way that they do in present-day humans.

say this is because such individuals do not have the ability to understand the nuances and basis of moral facts. I will argue in Section 3 that all that is required to hold one morally responsible is that they have the ability to know what kinds of actions their moral code prohibits, requires, discourages, encourages, and allows, and must have sufficient volitional ability to use the moral code as a guide for their behavior. They must be capable of knowing the content of their moral norms and when and how to apply them in morally loaded situations. In other words, induction into a moral practice is all that is required to make someone a responsible moral agent in that practice. If I am right, then if some animals can belong to a moral practice - which Rowlands' arguments support - then some animals can be moral agents.

The implication of Rowlands' account of moral agency is that humans who are normally considered moral agents would lose that status. Normal adult humans vary widely in their understanding of moral facts and their basis (assuming that there are moral facts). For example, normal human adults who have been raised in Canada since at least the 1980's have been taught that racism is unequivocally wrong. The norm is to embrace or at least respect the multicultural make-up of Canada, and to avoid discrimination based on race or culture. As this norm becomes more deeply entrenched in society, many do not question it and may be at a loss to explain why racism is wrong, and may not be able to explain what makes it wrong. Nevertheless, normal adult (and pre-adult) humans are held morally responsible for racist motives/actions because they should know better. On the other hand, elderly folk who were socialized before such anti-racist values became the norm are often excused for mildly racist comments or attitudes – the disapprobation is proportionate to the harmfulness of the attitude. Often their racist comments carry the best intentions: “Those Chinese kids are so good at math.” Nevertheless, the capacity to question and try to understand what underlies all racism and its harmful effects regardless of the

positive/negative character of the stereotype, does not determine who is morally responsible for racist motivations. Both the 'normal' young Canadian and the 'normal' elderly Canadian are morally responsible for their racist motivations and actions because both should know better. The elderly Canadian has the disadvantage of being inducted into a moral practice where racism was the norm – she didn't always know better - so that is why she might have more leeway. The capacity to question racist or anti-racist norms and understand the principle of principles upon which they are based has no bearing on whether or not they are morally responsible for racist attitudes. However, it is this capacity that allows the moral autonomy that can change beliefs about such norms.

Human moral judgments of praise/blame have a contextual character regarding the cultural, historical, personal and situational contexts. For example, we count Aristotle as being a responsible moral agent with highly developed metacognitive skills, despite his apparent failure to question the practice of slavery, or his failure to deem it wrong. His actions and motivations are appropriate targets of praise/blame and we can hold him accountable. However, we do not judge his acceptance of slavery with the same degree of disapprobation that we would judge a post-abolishment proponent of slavery. He had moral concerns regarding the welfare of slaves, but not their freedom. His ideal moral autonomy was hampered by his encultured acceptance of the practice of slavery and the lack of the Enlightenment concept of personal autonomy. My view should not be interpreted as moral relativism because I am not here making any claims about how moral facts should be determined or about their ontological status. I am claiming that moral responsibility is not contingent on moral autonomy in our everyday ascriptions of moral responsibility.

If understanding moral facts and principles separates moral agents from moral subjects, then Rowlands is left with the “uncomfortable” counterintuitive result that a substantial number of normal adult humans are not morally responsible, and are therefore not moral agents. One might understand moral facts according to principles of religion, rational thought, or moral conscience. Accordingly, meeting Condition (3) could involve the following: Why is stealing wrong? Because God says so; No rational agent could universalize it; it makes the victim suffer. Moving onto Condition (4) requires that one can ask and answer: What justifies God’s rules? What justifies universalizing principles? What justifies caring about another’s suffering? Most adult humans would end up making circular arguments appealing to their first order perceptions of right and wrong and could not be counted as having an understanding that would qualify them as moral agents on Rowlands’ model. The contingencies that led them to perceive certain motivations as right and wrong in the first order will operate at every level of scrutiny. Hanging moral responsibility on moral autonomy is incongruent with the phenomenal character of moral responsibility and, due to regress, inappropriately absolves too many normal adult humans. If we are going to live in a world peopled with morally responsible agents, then it is too much to insist that moral responsibility requires the ethical and meta-ethical analyses that Conditions (3) and (4) require.

In the next section, I will construct a sketch of the constitution of moral agency, which will oppose my view of what is required to act for moral reasons to Rowlands’ externalist view. I will argue that acting for moral reasons is a function of the individual’s form of life (including her cognitive capacities) and her social environment. Induction into a moral practice confers responsibility and non-human animals can belong to such practices. I will argue that moral



agency is indeed group relative, and Rowlands' worries about determining group boundaries or committing to inappropriate ascriptions of responsibility can be dispelled.

### **Section 3: Moral agency in practice**

Can animals act for moral reasons? Rowlands argues that some animals can on the basis of possessing morally-laden emotions – that is, emotions that have intentional content, which includes factual and evaluative content. These mental states are morally-laden in that they are directed at salient moral features of situations. Further, we can evaluate whether or not such emotional responses are correct by determining whether or not they track the truth of a moral claim that the intentional agent in question need not entertain. Such animals can be moral subjects on Rowlands' view, but not necessarily moral agents. Moral subjects are not morally responsible whereas moral agents are, and moral responsibility comes down to whether or not one is morally autonomous with sufficient understanding of moral facts and the principles upon which they are grounded.

Rowlands does not sufficiently show that animals can act for moral reasons. I argued in Section 1, that acting for moral reasons entails moral responsibility, and that acting for moral reasons requires that the agent can grasp those reasons. I will further develop these two arguments in this section. Acting for a moral reason requires that an intentional agent can make moral judgments or evaluations that direct her intentional action. Further, acting for a moral reason requires that such action is assessable according to action-guiding moral norms. Taken together, anyone who acts for a moral reason can be held morally responsible for such actions. There is a lot to un-pack here, and that will be my undertaking. My claim is that some animals can be moral agents, so let us begin with a sketch of what that might look like.

### 3.1 A non-metacognitive account of moral agency

Moral agency requires that an agent can act for a moral reason, and can be considered responsible for these actions. Hence, a moral agent is morally responsible. I consider the following five conditions sufficient for moral agency:

- (1) There is intentional action on the agent's part. Moral action is intentional action for a moral reason.
- (2) The agent is embedded in a social group that has [implicit] social norms regarding harms to others and sharing resources, which can include food, territory, and reproductive resources. Such norms are moral in that they regard characteristically moral matters.
- (3) Moral action relates to the moral norms of the group in which the agent is embedded, but is contextually bound to his form of life (e.g. kind of species, developmental stage, etc.) in interspecies or otherwise heterogeneous groups.
- (4) The agent can evaluate his motivations in accordance with his normative practice and intentionally act in accordance or discordance with the norms. To do this, the agent must have sufficient cognitive capacities to know what kinds of actions the moral code prohibits, requires, discourages, encourages, and allows, and must have sufficient volitional ability to use the moral code as a guide for his behavior.
- (5) Failure on the part of an agent to follow the norms is typically met with disapproval of other members (or member) of her group which can include sanctions, such as physical punishment, shunning, ostracism, exclusion from food sharing, etc.

If an animal can meet these conditions, then she can act for a moral reason. If she can act for a moral reason, then she can be morally responsible for those actions. Moral responsibility is a function of the agent, her world, and the social context in which she is embedded. This is not to say that anything goes for an agent depending on the situation in which she finds herself, but rather the agent's form of life (species, stage of development, cognitive capacities) determines the kinds of actions for which she can be responsible within her community, which may be heterogeneous in terms of the species, age groups, and cognitive capacities of its member. In section 2.4, I said that my concept of moral agency is in keeping with Rowlands' concept of moral subjecthood, except that I take moral responsibility to be a requirement. By that, I mean that my account involves emotions with propositional content, and allows an objective

consequentialist account of normativity, external reasons and moral facts. In other words, my account allows that there are facts in the world that make a moral agent's actions right or wrong, and the normativity guiding such actions is not dependent on anything in the agent's motivational set. And it allows that information about such things need not be intellectualized, but can be carried and conveyed via affect. Again, all of this needs unpacking. So let us start with what it means to act intentionally.

### **3.1.1 Action for a reason - Intentional action**

Rowlands claims that some animals can act for moral reasons if they satisfy the requirements for being a moral subject. It is not clear that moral subjects need make any evaluations or judgments about what should or should not be done. For Rowlands, it is sufficient for moral action that their behavior could merely be explained in terms of such evaluations, which is the purpose of his moral proposition tracking scheme (discussed in sections 1.1 and 2.3). He claims that morally-laden emotions need not be reducible to evaluations, even though they necessarily involve evaluative content that the subject need not be capable of entertaining. On my view, if a moral subject is not making any evaluations, and is incapable of actually having the implicit propositional content in question, then this subject is merely responding to a cause and not acting for a reason. Merely having an empathetic emotion that regards the welfare of others is not sufficient for acting for a moral reason. Rowlands slips between the terms "acting for" and "acting on the basis of" moral reasons. If a moral subject is better understood as acting on the basis of a moral reason, then it is incoherent to label such action as moral. Although an action and an actor are distinct, it is a mistake to ascribe reasons for an action to an actor who is incapable of acting for those reasons. So what does it mean to act for a reason? What counts as intentional action?

Sarah Buss (1999) argues that intentional action is action for a reason. Reflexive or purely instinctive action does not count. Intentional action requires that an agent is cognitively flexible, and has options in how she could act.<sup>4</sup> It is not enough that one is motivated by her desires and beliefs about how those desires might be satisfied. To act for a reason, a creature must direct her intentional action at a goal or end she sets for herself, and in that sense is self-directed. Evaluative beliefs must play a direct, independent motivating role in that the agent must make a judgment that she has sufficient reason to perform an action. This need not be self-conscious deliberation. The determination that she has sufficient reason to perform an action can be implicit in her judgment that it is obligatory, desirable, appropriate, etc. and such judgments can be primitive. It is sufficient that an agent reaches a conclusion about the relative desirability of options, and in that sense can be entirely first-order – no metacognition required. Peter Carruthers gave us an account of what that might look like – affect can carry information about the desirability of each option (see Section 2.2). She need not know what an intention is, or norms or values. She need only think something like “I will do this”, “this is good”, “worth doing”, or simply “to be done” and the reason can be entirely world-directed – she need not know or think that it is her reason.<sup>5</sup> She need only have the reason as her reason, not represent it as her reason. Rowlands’ proposition tracking method could be used here to track an intentional agent’s judgment that she has such a reason for her action. But when we add Buss’ criteria for

---

<sup>4</sup> Cognitive flexibility is the ability to restructure knowledge in multiple ways depending on the situational demands, such as the novelty, difficulty or complexity of the situation (Spiro, 1995). I emphasize the qualifier ‘purely’ for instinctive action that is reflexive because much complex, cognitively flexible intentional action can be grounded in instinct – including getting your dinner, and how you raise your children.

<sup>5</sup> World-directed reasons consist in having a first-person perspective on the world, that is, a particular point of view of being a self. In contrast, self-directed reasons consists in the ability to think about oneself as such, that is, to explicitly represent one’s perspective as such – an awareness of one as a self (Musholt, 2013). I interpret Buss’ use of the term self-directed to mean that the *goal* is self-directed in that that an agent does something for her reason, rather than merely doing what one has sufficient reason to do. She sets her own goals. Since Buss explicitly states that the reasoning need not be self-conscious, the perspective *qua* deliberator need not be represented as such.

intentional action, then the agent's behavior is not only explainable in terms of evaluation, she is actually making an evaluation. To set their own goals, agents must hold evaluative beliefs about the goals their desires dispose them to pursue.

Let us return to the example of the drooling wolf who has a desire to steal food from the alpha. She presumably has beliefs about how this desire could be realized – perhaps she could sneak over and then quickly grab the food and run, or she could attack the alpha and fight until she wins possession of the food, but is any of this worth doing? This is what it comes down to – the goals she acts on are those she determines are worth doing, are desirable, to-be-done. This is where her practical norms come in. She may feel no desire not to refrain from stealing, and may even be up for the fight, but knows that the right thing to do is to refrain. She need not know why it is the right thing, just know that it is so, in the first-order (the that-clause being embedded – she just needs to have the propositional content, not the form). Buss emphasizes two points: first, among the facts that are relevant to what an agent has reason to do, there are facts whose reason-giving force does not depend on the agent's motives; and, second, the fact that these reasons have reason-giving force plays an essential role in motivating behavior. Norms are such reasons – norms of food-sharing, friendship or kinship – and they are internal to the deliberative process. There is a perspective from which decisions are made – and forming an intention just is making a decision on Buss' view. The perspective is that of agent *qua* intention former rather than agent *qua* one who considers beliefs/desires in forming her intentions.

Following Sartre, Buss argues that an agent cannot reflect on one of her beliefs without detaching herself from it. The essential evaluative belief that an action is worth doing is based on all the facts that the agent considers, but it cannot itself be one of these facts as it articulates the agent's point of view at the time of the decision – as such, it cannot be considered from this

view. At the time that an agent acts intentionally, the evaluative belief that the action is to-be-done cannot be under her scrutiny. Now, suppose that all the reasons one takes into consideration from this point of view consists of facts about the world (e.g. alpha wolf is eating food) and first-order mental states (e.g. desire food) that are assessed in virtue of their affective phenomenal quality, rather than represented as desires one has. Over and above the reasons the agent considers, her practical norms guide her intentional action (e.g. norms about food-sharing). As the agent forms her intention (e.g. to sit and wait her turn), she is deeming that action as the right one. She is endorsing her own action at the time that she initiates it. Therefore, making a judgment about what she should do and deeming an action as right or wrong need not require metacognition. But acting for a reason does require that the intentional agent make a normative judgment (implicit in her action). To act for a reason, an agent need not be aware that they are her reasons for action, but her action must reflect her evaluative belief that her goal is right, desirable, good, worth pursuing. If this is right, then one who is capable of really acting for a moral reason can reasonably be held responsible for his action. To further develop this idea, more needs to be said about the “ought” that guides action for moral reasons.

### **3.1.2 Acting for a reason – normative desires**

Rowlands attributes the normativity of moral action to normative facts that one need not be capable of entertaining. On his view, one can act for a moral reason that one can never grasp. The normativity of the moral fact is entirely external to the agents motivational set. To evaluate the motivations of a moral subject as correct or incorrect, it is enough that they are assessable in light of normative facts, even if such facts are inaccessible to the agent’s reason. I argued in Section 1 that the implication of this view is that such a subject cannot be considered as acting for that inaccessible moral reason. It cannot be counted among his reasons for action, and can

only be counted as a cause of his behavior. To act for a moral reason, it must be possible that the normative force of that reason plays a motivating role for the actor.

Robert Myers (2012) argues that it is insufficient for moral action that motivating states are capable of according with normative truth (or facts), and as such are assessable in light of normative truths. On his view, an agent acts for a reason only if it is the agent's aim to get normative matters right. The concern to get matters right must be integrated into the agent's motivational system. Further, if an agent is not capable of being guided by normative facts, then these facts cannot be reasons for his actions. Myers says that conflating the evaluative with the normative might lead one to think that facts about what would be good for an agent could constitute reasons for them to act, even if they really cannot act for these reasons. Rowlands conflates these concepts in his account of what can constitute the normativity of a moral subject's reasons for action. This is why I see it as a mistake to say that reasons that are categorically inaccessible to an actor can be reasons for his action. Myers is right to say that facts about their good could not actually constitute reasons for creatures to act if these creatures are not actually capable of being properly guided by, and so acting for, reasons. This leads us back to the problem of how some non-human animals might grasp the normativity of moral reasons, which is required for them to act for moral reasons.

Can some non-human animals aim to get normative matters right? Myers provides an holistic account of desire whereby desires typically (but not exclusively) involve a disposition to act in ways that accord with the contents of one's normative beliefs. In other words, desires can be normative in that they can be shaped by the beliefs that an agent has about his reasons for trying to bring about some goal. Myers argues that an agent's desires can be sensitive to his normative judgments because their desires are generally aiming to get normative matters right.



On my view, aiming to get normative matters right need not require metacognition. Granted, Myers' holistic account is worded in such a way that it could be characterized as a metacognitive activity. However, building on my interpretation of Buss' account of intentional agency, I argue that one need not believe that one has reasons as reasons in order to have reasons. All that is required is a first-order belief that one's goal is worth pursuing – that it is the right thing to do. Buss argues that norms are internal to the deliberative process, so one need not be aware that one is trying to get normative matters right in order to be guided by them – it is sufficient that one aims to get them right in practice. If we think of normative truths as being what is pro-social or harm-reducing, then we can conceive of our hungry wolf as being guided by norms of cooperation and kinship that are requisite for successful social living (more will be said about the possibility that some non-human animals can have such norms when I discuss moral practices)<sup>6</sup>. Since such reasons must be accessible to her, it is tempting to think she would have to represent them as such. But as I argued above, these reasons can be implicit in her judgment of how she should behave. The evaluative belief that her goal is worth pursuing is the reason she has for pursuing it, so she need only be disposed to act in ways that will bring this goal about. The point to take away is that normative facts that are external to her motivational set can shape her desires because she has a systemic tendency to desire to get matters right (prudential or moral). In other words, she can aim to get normative matters right, and satisfying this aim requires that she has access to the normative facts of the matter.

---

<sup>6</sup> I have made a couple of controversial claims here. My claim that we can think of normative truths as being pro-social is not a claim about what constitutes normative truths. It is simply that it is reasonable to presume that normative truths would support pro-social rather than anti-social behaviours. Secondly, the suggestion that some non-human animals might be capable of cooperation is contentious (Tomasello, 2009), but there is no room here to defend this idea. Suffice it to say that there is plenty of evidence that suggests the possibility (de Waal, 1996, 2006, 2009), and if cooperation can be achieved without metacognition, then we need not discount it as possible for some non-human animals.

How might some non-human animals come to have normative beliefs, as distinct from evaluative beliefs? Myers distinguishes them in terms of their relevance to desires. Normative beliefs are those that lead to normative desires. If an agent has normative beliefs, then he will exhibit a systemic tendency to form desires that accord to them – hence, taken as a system rather than individual states, desires tend to aim to get normative matters right. Myers looks to Claudine Verheggen’s (2006, 2007a, 2007b) account of interpersonal externalism, based on Davidson’s account of propositional content determination, to explain how normative content might be fixed. Broadly, the contents of people’s propositional attitudes are fixed by their interactions with one another and the objects in their environment. Through interactions with others and with objects in their environment, a creature can start to think about those objects under some concepts and not others. This certainly applies to moral reasons. I argued in section 2.4 that pre-Enlightenment western society might not have had the conceptual framework to think about moral autonomy. The concept of a morally autonomous individual, one who has the ability to govern herself independent of her place in a metaphysical order or her role in social structures, developed socially and over time. In present western society, typically, we highly value moral autonomy and aim toward it as an ideal. As such, many agents act for this reason. We caution our children against peer-pressure and encourage them to ‘think for themselves’. Each individual, when confronted with the attitudes of others, can take steps to clarify their own attitudes. In this way, the propositional attitudes of the individual wolf who learns that stealing food leads to unpleasant interactions with others, can be shaped and fixed interpersonally. She can thus form evaluative beliefs about some kinds of outcomes as being worth promoting and other as not worth promoting, or worth resisting. Now there can be more to her propositional content than “avoid attack”. She can have a reason for resisting stealing, and that reason is to get

normative matters right. Taken as a system, her desires have a systemic aim to accord with the facts, including those about reasons for action. The fact of the matter is that she has pro-social normative reasons for promoting the good in her group. This information can be carried by affect, rather than intellectualized – it feels good to conform (Christensen et al, 2004). Such moral reasons are accessible to her faculties of practical reason. She knows what to do in certain social situations. She can judge her goal as to-be-done, according to the norms of her social practice, which are facts external to her motivational set, but not external to her reason. Of course, this begs the question that non-human animals have social norms. So let us now turn to that contentious question.

### **3.2 The moral practice**

Our actions and motivations are judged in society according to society's norms, but we are held accountable only when we are generally considered to be capable of being guided by these norms. I understand social norms to be informal, implicit "rules" that govern attitudes and behaviors in social groups. Failure to follow these norms is typically met with punishment or normative sanctions, which can include exclusion from the group, physical attacks, etc. I consider moral norms to be social norms that govern attitudes and behaviors which involve harm, resource sharing, or mating. The morality of a reason for action fundamentally involves the potential for preventing or causing harm to others. Moral responsibility requires normative social practices that involve cooperative and interdependent relationships with others. There could be no approbation or disapprobation without a normative social context.

My concept of a moral practice is in keeping with that of Rowlands, which is inspired by Wittgenstein's community view of rule-following, wherein the normativity of semantic rules (their correct and incorrect application) is essentially social. Wittgenstein argues that there is no

intrinsic meaning to signs (linguistic or non-linguistic) and they can only be interpreted with a practice. Some animals can have normative communicatory practices, such as vervet monkeys with their alarm calls, and dogs with their play-signals, as discussed in Section 2.2. The important point is that individuals can learn to adjust their signs in accordance with the norms of the practice, and misuse (or norm violations) can meet with disapproval. The little vervet who indiscriminately cries “snake” soon learns that others react unpleasantly. As he learns the normative “code of conduct” for communication, he learns to signal threats more accurately. This Wittgensteinian idea that content is only possible in the context of a practice can be extended to the notion of moral norms. Language only comes to have meaning in a social context, as do moral norms. Christina Bicchieri (2006) has argued that social norms can be considered as a “kind of grammar of social interactions”, in that a system of norms specifies what is acceptable and what is not in a society or group, and it is not the product of human design and planning. The instincts and physical structures required for language or moral behavior are in-born, but for either to have meaningful content, they require the context of normative social practices.

If I am right, then acting for moral reasons requires that the agent can belong to a community, wherein normativity is embedded in social practices and is a function of the relation between the individual, the world and the social environment in which the agent is embedded. The Davidsonian concept of interpersonal externalism discussed above lends support to the idea that normative content is socially fixed. Social norms are the unplanned results of individuals’ interactions (Bicchieri and Muldoon, 2014). There is debate about whether or not non-human animals can have social norms or merely precursive proto-social norms due to the claim that they require shared intentionality to be collectivized, and non-human animals are not proven capable

of shared intentionality (von Rohr et al, 2011). My claim is only that some animal groups can have normative systems, and this is shown by their communication practices and social practices. Whether the norms of these groups can be considered social norms or proto-social norms is not what is at stake here. My goal is to show the form that moral agency takes, and show that it is not different in kind for human and non-human animals. The content of human moral codes varies widely, due in part to our capacity for moral autonomy, but social norms for controlling harm, resource sharing and sexual reproduction are fundamental.

Now we have a context for making ascriptions of moral responsibility – the moral practice. Can it apply to some non-human animals? Can we ascribe to them moral responsibility? I have argued that the possibility for such ascriptions is co-extensive with acting for moral reasons, in that acting for a moral reason must be assessable in terms of moral responsibility. If an agent acts for a moral reason, then it must be possible to hold the agent responsible for that action, and therefore she can be praised or blamed for her action. On my view, none of this is possible without a moral practice, within which responsibility is conferred. Right away, questions about how to think of the boundaries of moral responsibility pop up, as to whether or not moral responsibility can extend beyond the limits of a social group, but I will address those later. First, I must sketch out how one might become a responsible moral agent. An agent can only be morally judged for his action if he has sufficient cognitive capacities to know what kinds of actions his moral practice prohibits, requires, discourages, encourages, and allows, and he must have sufficient volitional ability to use the moral practice as a guide for his behavior. Sufficient volitional ability is the ability to do otherwise – to intentionally act differently than the moral norm prescribes. Such an agent must be cognitively flexible and have more than one option from which to choose, which can be as simple as to-do or not-to-do. When an individual

meets these requirements, he can be considered fully inducted into his moral practice. Others in his group expect him to behave in certain ways. If he violates moral norms, he would be subject to the associated sanctions, such as ostracism or attacks.

Now let's look at some real-life examples of animals acting for moral reasons (on my view). Masserman et al (1964) conducted laboratory experiments on Rhesus monkeys where the subjects were fed only if they pulled a chain which delivered an electric shock to another unrelated macaque who was in plain view through a one-way mirror. If they refused to pull the chain, they starved. One monkey chose to starve for almost two weeks rather than harm his conspecific. Most of the monkeys (87%) chose starvation over harming another. Further, the incidents of refusal to pull the chain was higher in those monkeys who had previously been shocked, suggesting that since they knew what it felt like, they could empathize with their conspecifics. The researchers also controlled for dominance relationships, and found that they played no significant role. They concluded that a majority of rhesus monkeys will consistently suffer hunger rather than secure food at the expense of electroshock to a conspecific; and, this sacrificial pattern is induced primarily by visual communication, remains characteristic for individual animals, and is enhanced by familiarity or previous experience of shock, but is not significantly related to relative age, size, sex, or dominance. It is not unreasonable to conclude that these monkeys were intentionally acting for the moral reason that it is wrong to harm another macaque under these circumstances, were aiming at what is right (pro-social); and were guided by their evaluative beliefs and normative desires. But can we hold them morally responsible for their actions? We cannot, but other macaques in their group might. Norms about harming others and learning empathy are no doubt socially learned elements in groups of highly social mammals, such as macaques. We can say that these macaques have moral obligations for

which they are responsible. As Shapiro quips, “While many philosophers don’t think other animals can have obligations, it could be that macaques think they do!” (Shapiro, 2006, p.365). Perhaps you can see where this is leading in terms of how to reasonably ascribe moral responsibility to some non-humans, which I will articulate later.

If there is moral action in non-human animals, then we should expect to see it most recognizably and abundantly in some of our closest relatives. Indeed, researchers have observed plenty of behaviors that we consider moral when observed in humans (de Waal, 1996). A particularly interesting case, in that it involves bystander intervention that suggests prosocial motivation, is recounted by von Rohr et al (2011). In chimpanzee groups, infants hold a special position of tolerance, as they do in many mammalian social groups. They enjoy extreme tolerance for behaviors that typically elicit negative responses when conducted by adults or older youngsters, such as food-stealing, jumping on others, and even interfering with mating. As such, we should expect to see normative sanctions against harming these special individuals, and that is just what anecdotal evidence suggests. Young chimpanzees that approach a newborn too closely elicit aggression from the mother and learn to observe the infant from a more respectful distance; older chimpanzees self-handicap when playing with little ones seemingly adjusting their behavior to the capacities of the younger playmate; older playmates increase their play-signaling when in proximity to the infants’ mother presumably so that she understands no punishment is required; and so on. Intra-group infanticide is rare – only 5 of 112 infants born between 1964 and 2005 in the Kasekal community of Gombe were killed by someone from within the community. Aggression toward infants has been met with massive vocal protests from several adults, as Frans de Waal (1982) observed when a highly aroused bluffing male grabbed a three-year old infant and swung him against a wall. The vocalizations of the other adults were

the loud and sharp sounds typically emitted by bystanders and are interpreted as protests, indicating “sympathy” for the victim, and seem to occur only in dramatically escalated situations. These vocal protests can reasonably be interpreted as indicating disapprobation regarding the adult male’s attack on the infant.

The above story nicely illustrates conditions 4 and 5 of my sketch of moral agency. It shows how and when non-human animals might express approbation, and indicates that normal adult members of the group are expected to behave according to the norms of the practice (do not harm infants). It also shows how infants are not expected to know how to behave according to the norms of the group, and are thus afforded extreme tolerance. As they grow and learn, expectations for normatively appropriate behavior increase, suggesting that they can only be morally judged for their actions if they have sufficient cognitive capacities to know what kinds of actions their moral practice prohibits, requires, discourages, encourages, and allows.

### **3.2.1: Setting the limits**

My account of moral agency depends on moral responsibility being group relative. This is not to be taken as a commitment to moral relativism, but rather that induction into moral practices require the ability to grasp the normative contents of such practices. Moral agents are only moral agents within their moral practices. For humans, the boundaries of our practices are quite limitless, whereby we expect that some moral norms regarding the most serious harms be applied inter-culturally and inter-species. The concepts of international human rights and animal welfare protection are evidence of this human capacity, which in turn is evidence of our capacity for moral autonomy. However, humans have made mistakes ascribing responsibility inappropriately, such as the fact that humans once treated pigs as if they were morally responsible according to a human moral code. One of the most important requirements for acting



for a moral reason was overlooked – and that is the ability to grasp the moral reason. Pigs cannot be expected to grasp the contents of a human moral code, and therefore should never be held responsible for violating its norms. For an individual to become a moral agent, he must first learn the rules of his moral practice and be capable of perceiving the morally salient features of a situation. Responsibility only comes with being able to know the norms and knowing how and when to apply them in situations. As such, moral responsibility is contingent on a creature's form-of-life.

One could object that there are many cases where there are no tidy group boundaries, where it makes no sense to say an animal can act for a moral reason. For example, who makes up the moral practice of a dog living in human society? For the dog, it is his pack (which may include everybody from humans to parakeets) and he might feel an obligation to protect them from threats (which may include other dogs or humans). A “good” dog learns how and when to inhibit his bite and he learns when it is appropriate to threaten a serious attack. He learns the rules of his moral practice, and that for him is a dog-practice even though it is experienced in a multi-species environment. This is because the intentional content in question is of the dog's mind - it is dog-context bound even in a multi-species social context. If the question is whether the dog can be a moral agent, then it is the dog's intentional content and how he represents his practice that is in question. The context is dog-context, and his accountability is weighed accordingly. Dogs are corrected within such practices, as Rowlands point out in his story about the dog learning to be gentle with the small child. It would be wrong to hold him accountable as a human moral agent, but even humans can hold him accountable in a dog-context. He may be blamed for certain behaviors and praised for others, and when these behaviors involve morally salient features, one could say he is morally praised and blamed. However, it would be ludicrous

to hold him accountable as a full moral agent in a human moral practice – his is a dog moral practice in which humans play a role, just as mine is a human moral practice in which dogs play a role. I cannot be considered a moral agent in the dog's moral practice any more than the dog can be one in mine. We cannot be fully inducted into each other's practices as we have no way of grasping the content of each other's normative reasons. I cannot know how to interpret all the signals a dog can make, and he cannot interpret all the signals I can make. This limits the degree to which our interpersonal interaction can fix the normative contents of our motivational states. Hence, the normative concepts under which we can think about each other's actions are limited by our different forms-of-life. When the differences are great enough, some individuals only qualify as moral patients in one practice, despite qualifying as moral agents in another.

### **3.2.2: Mere convention or moral norm?**

My descriptive account of moral agency could be charged with blurring the distinction between morality and convention. It is not enough that there seem to be behavioral regularities in social groups that seem to involve moral matters. I have two short replies. We could follow von Rohr et al (2011), and distinguish moral norms from mere convention by the type and intensity of the emotional response that a violation elicits. However, this might prove to be quite unreliable. Humans might turn a blind-eye to immoral actions simply because they do not want to get involved, which could lead an observer to conclude that the norm violation was not that serious, therefore it must be a mere convention. Think of the famous case of Kitty Genovese. She was brutally murdered and raped while her cries for help went unanswered by her neighbours, at least twelve of whom heard her, and two of whom knew that she had been stabbed early in the attack which took place over one half-hour

([http://en.wikipedia.org/wiki/Murder\\_of\\_Kitty\\_Genovese](http://en.wikipedia.org/wiki/Murder_of_Kitty_Genovese)).<sup>7</sup> There are many factors that can account for the lack of intense emotional bystander reaction, but it is certainly not the case that the norms violated were merely conventional.

Perhaps, as Andrews (2012b) suggests, the distinction between moral norms and social conventions is arguably a false one. Most conventions contain moral normativity at their center. Shaking hands or touching noses are conventional greetings, but are also gestures of trustworthiness. They leave each one vulnerable to the possibility that the other might harm her. Perhaps this is why it is typically considered offensive to violate such norms of greeting – trust is central to social living. I have tried and failed to think of one social convention that cannot be associated with some aspect of morality, which is not to say that none exist. Most social conventions (if not all) can be explained in terms of harm, sex, or resource sharing.

---

<sup>7</sup> The details of this case are under dispute as the bystander apathy may not have been as extreme as previously reported. See <http://www.newyorker.com/magazine/2014/03/10/a-call-for-help>. The point remains that by-stander apathy is a phenomenon which may lead a naïve observer to mistaken conclusions about moral norms.

## Conclusion

Can animals be moral? That depends on whether or not they can act for a moral reason. Rowlands' moral subject is better understood as responding to moral reasons. Their behavior is caused by factors that are only moral reasons for those agents whose practical reason can possibly access these factors. Rowlands' moral subject might have "moral emotions", but I have argued that this is not sufficient for acting for a moral reason. Empathy is arguably necessary for moral beings. After all, sociopaths characteristically lack empathy and are by definition amoral in that they typically have no regard for right and wrong and often disregard the rights, wishes and feelings of others. That is not to say they cannot fake it due to their metacognitive skills and act for moral reasons. But metacognitive skills are not necessary for acting for moral reasons – affect and a social practice can provide all the information necessary for an empathetic creature to act for a moral reason. I hope I have at least shown that we can think of the moral agency of some non-human animals in such a way that we can make sense of ascriptions of responsibility without putting them in danger of inappropriate moral persecution; and that we can take seriously the possibility of their moral agency when we study their behavior.

## Bibliography

- Andrews, K. (2012). *Do apes read minds?: Toward a new folk psychology*. Cambridge, Mass.: MIT Press.
- (2012 b)“Ape Autonomy? Social norms and moral agency in other species”, *Philosophical Perspectives on Animals: Mind, Ethics, Morals*. Markus Wild, ed. Springer.
- Andrews, Kristin and Lori Gruen. (2014). “Empathy in other apes”. In *Empathy and Morality*. Ed.Heidi Maibom Oxford University Press.
- Bekoff, M., & Pierce, J. (2009). *Wild justice: The moral lives of animals*. Chicago: University of Chicago Press.
- Bicchieri, Cristina and Muldoon, Ryan, (2014) "Social Norms", *The Stanford Encyclopedia of Philosophy* (Spring Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2014/entries/social-norms/>.
- Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. New York: Cambridge University Press.
- Buss, Sarah. (1999) “What practical reasoning must be if we act for our own reasons”. *Australasian Journal of Philosophy* Vol. 77, No. 4, pp. 399-421; December
- Carruthers, Peter. (2008) “Meta-cognition in Animals: A Skeptical Look” *Mind & Language*, Vol. 23 No. 1 February, pp. 58–89.
- Campbell, Chris. (2013) <http://genealogyreligion.net/the-earliest-moral-ethical-laws-were-not-religious>
- Christensen, P. Niels, Hank Rothgerber, Wendy Wood, and David C. Matz (2004), "Social Norms and Identity Relevance: A Motivational Approach to Normative Behavior," *Personality & Social Psychology Bulletin*, 30 (10), 1295-309.
- Christman, John, (2011) "Autonomy in Moral and Political Philosophy", *The Stanford Encyclopedia of Philosophy* (Spring Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2011/entries/autonomy-moral/>.
- Davidson, Donald. (1982). “Rational Animals”. *Dialectica: International Journal of Philosophy of Knowledge*, 36(1982): 317-328.

- Darwin, Charles. (1880). *The descent of man, and selection in relation to sex*. New York: D.Appleton & Co.
- Dennett, D. C. (1998). *Brainchildren: Essays on designing minds*. Cambridge, Mass.: MIT Press.
- (1987). *The intentional stance*. Cambridge, Mass.: MIT Press
- Gert, Bernard, (2012) "The Definition of Morality", *The Stanford Encyclopedia of Philosophy* (Fall Edition), Edward N. Zalta (ed.), URL = [<http://plato.stanford.edu/archives/fall2012/entries/morality-definition/>](http://plato.stanford.edu/archives/fall2012/entries/morality-definition/)
- Flack, Jessica C. and Frans B.M. de Waal. (2000). "Any animal whatever: Darwinian building blocks of morality in monkeys and apes." *Journal of Consciousness Studies*. 7:1-2. pp. 1-29
- Korsgaard, Christine. M. (2010) "Reflections on the Evolution of Morality." *The Amherst Lecture in Philosophy* 5:1–29. [<http://www.amherstlecture.org/korsgaard2010/>](http://www.amherstlecture.org/korsgaard2010/).
- (2008) "The activity of reason", <http://www.people.fas.harvard.edu/~korsgaar/CMK.Activity.of.Reason.pdf>
- (1996). *The sources of normativity*. New York: Cambridge University Press.
- Masserman, Jules H.; Wechkin. S, and Terris, W. (1964) "Altruistic behavior in Rhesus monkeys." *The American Journal of Psychiatry* Vol 121. Dec. 1964. 584-585.
- Musholt, K. (2013). "A Philosophical Perspective on the Relation between Cortical Midline Structures and the Self". *Frontiers in Human Neuroscience*, 7, 536. doi:10.3389/fnhum.2013.00536
- Myers, Robert H. (2012) "Desires and Normative Truths: A Holist's Response to the Sceptics" *Mind*, Vol. 121 . 482 . April, pp.375-406
- Rowlands, Mark. (2012). *Can animals be moral?* New York: Oxford University Press.
- Spiro, R., P. Feltovich, M. Jacobson, and R. Coulson. (1995). "Cognitive flexibility, constructivism, and hypertext: Random assess instruction for advanced knowledge acquisition in ill-structured Domains". In T. Duffy & D. Jonassen (Eds.), *Constructivism and the Technology of Instruction*. Hillsdale, NJ: Erlbaum.
- Stoljar, Natalie. (2014) "Feminist Perspectives on Autonomy", *The Stanford Encyclopedia of Philosophy* (Spring Edition), Edward N. Zalta (ed.), URL = [<http://plato.stanford.edu/archives/spr2014/entries/feminism-autonomy/>](http://plato.stanford.edu/archives/spr2014/entries/feminism-autonomy/)

- Tomasello, M. (2009). *Why we cooperate*. Cambridge, Mass.: MIT Press.
- Verheggen, Claudine. (2006). "How Social Must Language Be?" *Journal for the Theory of Social Behaviour*, 36, pp. 203–19.
- (2007a). "Triangulating with Davidson". *Philosophical Quarterly*, 57, pp. 96–103.
- (2007b). 'The Community View Revisited'. *Metaphilosophy*, 38, pp. 612–31.
- Von Rohr, Claudia Rudolph. Burkart, J. M., & Schaik, C. P. (2011). "Evolutionary precursors of social norms in chimpanzees: A new approach". *Biology and Philosophy*, 26(1), 1-30.  
<http://search.proquest.com.ezproxy.library.yorku.ca/docview/1440893152?accountid=15182>
- Waal, F. B. M. de. (2009). *The age of empathy: Nature's lessons for a kinder society*. New York: Harmony Books.
- (2006). *Primates and philosophers: How morality evolved*. Eds. Macedo, S., Ober, J., & Wright, R. Princeton, N.J.: Princeton University Press.
- (1996). *Good natured: The origins of right and wrong in humans and other animals*. Cambridge, Mass.: Harvard University Press.
- (1982). *Chimpanzee Politics: power and sex among apes*. Unwin Paperbacks, London.
- Wittgenstein, L. (1974). *Philosophical investigations* (3rd ed.). Oxford: Blackwell.