2013-02-26

Moral Emotions in Nonhuman Animals

Daniel A. Hampikian
University of Miami, danielhampikian@yahoo.com

UNIVERSITY OF MIAMI

MORAL EMOTIONS IN NONHUMAN ANIMALS

By

Daniel A. Hampikian

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

May 2013

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

MORAL EMOTIONS IN NONHUMAN ANIMALS

Daniel Hampikian

Approved:

_____ Mark Rowlands, Ph.D. Professor of Philosophy	_____ M. Brian Blake, Ph.D. Dean of the Graduate School
_____ Colin McGinn, Ph.D. Professor of Philosophy	_____ Michael Slote, Ph.D. Professor of Philosophy
_____ Hans-Johann Glock, Ph.D. Professor of Philosophy University of Zurich	

HAMPIKIAN, DANIEL A. (Ph.D., Philosophy)
Moral Emotions in Nonhuman Animals (May 2013)

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Professor Mark Rowlands.
No. of pages in text. (249)

I defend the view that some nonhuman animals can be morally motivated by empathic emotions. First, I argue that we are justified in ascribing to some animals phenomenal consciousness, the conceptual capacities to represent values and intentional objects, and the relevant behavioral and physiological similarities to human emotional states. Subsequently, I argue for a model of basic emotions that I call the Awareness of Physiological Vehicles (APV) account. According to the APV account, an animal's emotion is best thought of as a focal awareness of an intentional object and a peripheral awareness of sensations of physiological states as indicators of value. Next, I address various skeptical worries from Cognitivist and Kantian views of moral motivation that without the capacities for critical self-scrutiny of actions and motivations animals cannot be morally motivated. Finally, I give two compatible and plausible explanations for why we are justified in ascribing moral motivation to animals in the absence of being able to justifiably morally praise or blame them for exhibiting emotionally motivated behaviors. I conclude by considering some moral and experimental implications of this view of animals as complex emotional and moral beings.

For Jake

Table of Contents

Introduction

In this dissertation I defend the thesis that nonhuman animals can be morally motivated by empathic emotional states.[1] This thesis is intended to advance a debate that has been recently occurring between animal researchers and philosophers who have claimed that morality is an evolved trait that both humans and some animals possess (albeit not to the same degree of sophistication), and opposing ethicists who have argued that such social emotions are not sufficient for morality.[2] These opposing historical and contemporary ethicists; including Kant, Korsgaard, Aristotle, and Dixon, hold that moral motivation requires the capacity for reflective and critical scrutiny of one's motives and one's possible responses to the morally relevant features of a scenario, making it highly unlikely that any nonhuman animals (hereafter referred to as "animals") can be morally motivated.[3] Compounding this debate are arguments from various theorists in animal cognition and philosophy of mind implying or explicitly claiming that it is theoretically suspect or hopelessly problematic to attribute animals *any* emotions that are similar to and comparable to human emotions.[4]

---

[1] Empathic emotional states are, roughly speaking, states in which an organism has an emotion that is more appropriate to the situation of another organism. This somewhat provisional definition is in keeping with developmental psychologist Martin Hoffman's systematic investigation into various kinds of empathy, all of which he thinks are characterized by having a motive base of "an affective state more appropriate to another's situation than one's own." (Hoffman, 2000, p. 4)

[2] The claim that morality is an evolved and shared trait is argued for in Bekoff and Pierce's *Wild Justice* 2009, and De Waal's *Good Natured: The Origins of Right and Wrong in Humans and Other Animals* 1996. The opposing views are found in Dixon's *Animals Emotion and Morality* 2009, and Korsgaard's, "Morality and the Distinctiveness of Human Action" in De Waal, *Primates and Philosophers* 2006.

[3] The explicitly opposed ethical views and their historical basis in Aristotelian or Kantian philosophical thought can be found in Dixon's *Animals Emotion and Morality* 2009, and Korsgaard's, "Morality and the Distinctiveness of Human Action" in De Waal, *Primates and Philosophers* 2006.

[4] For views committed to the denial of animal affective consciousness see Carruthers (1996, 2004), Descartes (1649) Gallup (1970), Rolls (1999), Rosenthal (1986, 1993), Ryle (1949) Skinner (1974), and Watson (1925). For views committed to the denial of the existence or translatability of animal beliefs and concepts, and by implication emotions that involve beliefs or concepts, see Davidson (1975, 1982), Chater and Heyes (1994), and Stitch (1978). Reasons to think emotions require complex normative judgments that are almost certainly inaccessible to animals can be found in Nussbaum (2001), Sartre (1971), and Solomon (1973).

Accordingly, I begin Chapter One by addressing several skeptical reasons that have led theorists and researchers writing on animal emotions to doubt that animals have the necessary capacities for such emotions. I address the worries that have led such theorists to doubt whether animals possess some of the necessary elements of moral emotions by qualifying the way in which we can justifiably ascribe emotions to both humans and animals, and by clarifying the nature of the evidence that supports such ascriptions. These necessary elements include phenomenal consciousness, the conceptual capacities to represent the implied values and intentional objects of emotional states, and the relevant behavioral and physiological similarities to human emotional states.[5] I then legitimize and clarify ascriptions of emotions to nonhuman animals by systematically laying out the presuppositions that go into ascribing animals emotional states (that are similar to and comparable to human emotional states), and then arguing that these presuppositions are defensible despite various skeptical concerns.

After the qualifications necessary to justify each of these presuppositions are explicitly stated, I defend in part two of Chapter One a model for animal and human emotions that is informed by empirical evidence and that takes on the relevant qualifications. According to this model, which I call the Awareness of Physiological Vehicles (APV) model, an animal is in a basic emotional state just in case:

1. The organism has a peripheral awareness of locally contributing sensations of physiological states not merely as states of the body, but as *vehicles of value,* or in other words, as contributory representations

[5] An additional worry is that there is no clear cut way to demarcate animals with emotional states from those with no (or fewer possible) emotional states. Still, there are a variety of nonhuman animals (including all mammals) that I hope to show clearly have emotions. The issue I am concerned with in this dissertation is not where to draw a line between emotionless and emoting animals, but rather *which* emotions that animals with the relevant capacities might justifiably be held to have, what *form* those emotions take, and the nature of our *evidence* for ascribing these emotions.

indicating the *psychological value* that the intentional object has to the organism, and

2. The organism's global phenomenological state includes a *non-mediated* focal awareness of an intentional object *having the value that is indicated by their peripheral awareness* of sensations of physiological states.

Following this, I elaborate how the APV account best explains the neurological, evolutionary, developmental, and behavioral evidence for the common structure of at least the basic emotions of fear, sadness, happiness, anger, *suffering* (by which I mean pain that functions as a vehicle of value in the total emotional state of the organism), and *delight* (by which I mean pleasure that functions as a vehicle of value in the total emotional state of the organism). Then I compare this account to the standard reductionist options for explaining emotions in terms of awareness of bodily changes or judgments (or a combination of the two). Subsequently, I distinguish the APV account from both a simple adverbial account endorsed by Mark Rowlands and an influential and initially promising somatic account defended by Jesse Prinz that considers emotions to be embodied appraisals. The standard reductionist options and the simple adverbial account are found to be deficient in various ways, and although I hold that the embodied appraisal account is correct to postulate that (some) organism-environment relations that bear on well-being are represented in basic emotional states, I conclude that it is ultimately unsatisfactory as a general explanation of the structure of emotions in humans and animals. The APV account, however, is able to account for the intentional structure of emotions in a way that avoids the difficulties that render Prinz's embodied appraisal account implausible.

In part three of Chapter One, I argue that emotions are best thought of as flexibly developing in sophistication as an organism's capacities for empathy, memory, association, expectation, and categorization develop in tandem with its emotional capacities in its social and environmental interactions. Then I review the neurological, behavioral, and evolutionary evidence for this developmental conception of emotions, and conclude Chapter One by arguing on the basis of this conception that animals ought to be ascribed what I (following Lazarus) stipulatively refer to as "compassion" and "benevolence" (in a sense to be specified). They can be ascribed "compassion" in so far as there is converging evidence that animals are motivated by empathic suffering to act to alleviate the suffering of another, where the nature of this suffering is understood as experiencing pain or aversive emotion more appropriate to another in a way that internally represents the negative psychological value of the other's suffering. Similarly, they can be ascribed "benevolence" in so far as they are empathically motivated to act to promote the well-being of another. This, I argue, is powerful prima facie evidence for the legitimacy of ascribing to animals at least two kinds of morally motivating mental states (and the corresponding dispositions).

After developing the plausible APV model for ascribing and comparing human and animal emotions of the same kind, I criticize a contemporary cognitivist approach to emotions in Chapter Two that I refer to as "strong cognitivism." This approach reduces emotions to (or explains emotions purely in terms of) judgments, evaluations, beliefs, or appraisals.[6] Strong cognitivism makes it implausible, or at least highly problematic, to ascribe moral emotions to animals because according to strong cognitivism moral

[6] The proponents of this kind of view I shall be criticizing include Solomon, Nussbaum, and Dixon, but any view that explains emotions purely in terms of normative judgments or explicit linguistic evaluations will also be subject to the following criticisms.

emotions consist of several complex moral judgments or evaluations. However, strong cognitivism is implausible as an account of human (moral and amoral) emotions for several reasons, the most well-known of which is that it is extensionally inadequate because it cannot account for a wide variety of emotional states that occur without an evaluation or that occur despite the agent making a contradicting evaluation.

The main proponent of strong cognitivism that I focus my criticisms upon is B. A. Dixon, but I will also criticize the strong cognitivist views that have been advanced by Solomon and Nussbaum. I will defend the claim that these specific forms of strong cognitivism are untenable because strong cognitivism as a general kind of explanation cannot fully explain the nature of human (or animal) emotions. This kind of explanation fails, I argue, for the following four reasons.

First, this kind explanation simply does not have the resources to explain a wide variety of mental states that are pretheoretically considered to be paradigmatic emotions such as strong and unreflective fear, love, and empathic concern. Second, strong cognitivism cannot explain how emotions are value determining in a way that can occur prior to explicit evaluations and in a way that can override or influence rational judgments. Third, it conflates the negative condition that an emotion cannot be had while an agent sincerely apprehends a judgment that contradicts a necessary implication of that emotion, with a positive property that is essential to all emotions.[7] Fourth, this kind of account cannot fully explain the phenomenology and physiology that accompanies emotional states.

[7] For example, while I cannot be angry with John for wronging me if I sincerely apprehend the judgment 'that John did not in fact wrong me,' I can be angry with John without explicitly thinking that he wronged me, perhaps because I am unaware of this implication or otherwise preoccupied.

Finally, I argue that strong cognitivism imposes too strict a criterion for moral motivation in the human case. Actions motivated by an awareness that another is suffering and a desire to alleviate that suffering are, when a human has them, clearly an emotional state that motivates actions we would pretheoretically characterize as moral even in the absence of explicit moral judgment.[8] Strong cognitivism denies that this is possible, which is a reason to reject it as an explanation of moral emotions as well as more basic emotional states.

With the critical reasons to reject the possibility of animals being ascribed potentially morally laden emotions addressed and the APV account defended and distinguished from some leading contemporary somatic and cognitivist views, I turn in Chapter Three to the issue of whether empathic emotions exhibiting or reflecting caring concern for another are sufficient for moral motivation. I begin by addressing a challenge from Korsgaard, who defends a Kantian view of morality according to which moral action necessarily requires the capacity for rational self-reflection and normative self-government of one's purposes or motives for acting. As Korsgaard and others have noted, this puts moral behavior beyond the capacity of almost all nonhuman animals. Accordingly, I challenge the adequacy of the general metaethical picture that requires rational reflection and normative self-government for moral behavior. I do so by presenting counterexamples to the asserted necessity of normative self-government for moral motivation and by arguing that self-critical capacities can even be morally inappropriate in certain social situations involving caring concern. Then I show that the justification for moral praise and blame and the justification for ascribing moral

[8] Here I do not mean to imply that all desires necessarily involve emotions (mere preferences may lack emotional aspects), but rather that an emotional state of compassion necessarily involves a desire to alleviate suffering.

motivation are conceptually distinct issues. I argue that while normative self-government and an understanding of moral norms might be necessary for ascribing praise or blame, these capacities are not necessary for moral motivation. Finally, I conclude Chapter Three by arguing that when (some) social emotions motivate altruistic actions, we ought to suspect that apparently moral *behavior* in humans and some animals is in fact *morally motivated*. However, even if there is a common pattern of emotionally motivated cooperative, altruistic, kind, and helping *behaviors* that are found in all mammals, this is compatible with some mammals being motivated by mere reflexive feelings or by egoistic goals rather than genuine concern with the well-being of others.

Accordingly, in Chapter Four I explain why some emotions, especially ones that we are justified in attributing to animals, are correctly *classified* as *moral*. To accomplish this I argue for and supplement two separate but compatible explanations for why we are justified in ascribing moral motivation to animals exhibiting emotionally motivated altruistic behavior. In part one I begin by explaining Prinz's view of the conceptual relation between (apparently moral) emotional states and morality, which is somewhat dependent on his flawed conception of emotions as embodied appraisals and has unfortunate relativistic consequences. While moral relativism is embraced by theorists like Prinz (in the case of human morality) and Bekoff and Peirce (in the case of nonhuman morality), I will argue that the implications of relativism are reasons to reject such views and endorse Rowlands' moral proposition tracking approach and Slote's empathy based Sentimentalist approach, both of which offer compatible and non-relativistic explanations of why some animal emotions are properly classified as moral.

The first approach, Rowlands moral proposition tracking approach, holds that we are justified in ascribing a moral emotion to an animal just in case 1) the emotional state E is a content-involving emotion, 2) there exists a proposition, P, which expresses a moral claim, and 3) the truth of this proposition is required if E is not to be *misguided*. (Rowlands 2012, pg. 79) That is, if an animal is in an empathic emotional state such as compassion there is a moral claim, such as "this animal's suffering at the hands of another is immoral," that must be true in order for the evaluative content of the empathic emotion to not be *misguided*. While the notion of being misguided requires significant theoretical unpacking, the basic idea is that an emotion is misguided just in case it entails an erroneous evaluative proposition. For example, if Smith feels indignation towards a police officer who writes him a ticket for speeding, this emotion entails the erroneous evaluative proposition (that need not be entertained by Smith) that the police officer has wronged him. In this sense, Smith's indignation bears a tracking relation (that I will explain in greater detail in Chapter Four) to the truth of a moral proposition that Smith need not entertain in order to have the (moral) emotion. Similarly, the content of the moral proposition "this animals suffering at the hands of another is wrong" is a true moral claim independently of whether or not the animal motivated by empathic compassion is explicitly entertaining the moral (and evaluative) proposition that, if true, ensures this emotion is not misguided. After explaining Rowlands' tracking relation account, I argue that it is usefully supplemented by the APV account of empathic emotions. This supplementation allows me to explain not only why the moral emotions of animals track the truth of moral propositions, but also why the emotion is structured such that the value of the well-being of another animal is represented internally.

Importantly, this supplemented account is neutral with respect to what ultimately makes a moral proposition true or false, and hence its plausibility is independent of any particular theory of morality being correct. However, the case for animal morality is improved to the extent that we are able to specify how certain emotions function to morally motivate both animals and humans in a way that avoids making morality relative or contingent.

This leads me to argue for an explanation of moral justification and moral meaning recently defended by Michael Slote that implies that certain emotional states actually contain, because of their empathic nature, a necessary component of (objective and a priori) human morality that is sufficient for moral motivation in animals.[9] In this way I will argue that it is indeed plausible to regard animal actions as motivated by moral or immoral concerns on the basis of whether or not those actions display the presence or lack of empathic concern for the other. This will serve three purposes. First, we will have one possible and plausible explanation for what makes moral propositions objectively true or false so that we can with greater legitimacy and specificity apply Rowlands' moral proposition tracking apparatus to ascribe moral emotions to animals.[10] Second, this explanation allows me to state more precisely the relation between human moral agents capable of making objective moral judgments *and* being morally motivated by emotions, and morally motivated animals that are only capable of the latter. Finally, we will have a psychologically realistic explanation of the mechanism of moral sensitivity that specifies the phenomenological structure of an animal's mental state when

---

[9] Defended in his *The Ethics of Care and Empathy* (2007) and *Moral Sentimentalism*. (2010) Some animals may well have rudimentary capacities for moral approval and disapproval, as I suggest in the conclusion, but this is a separate issue from whether or not they can be morally motivated.

[10] As this explanation is one *possible* way of explaining the objective and a priori character of moral emotions in animals, the dialectical import of this explanation is illustrative and as such does not entail that the case for animals entertaining moral emotions stands or falls with the plausibility of moral sentimentalism.

they act out of caring concern for the well-being of another, and that clarifies the relation of this mental state to moral motivation.

After making the case for moral sentimentalism and explaining the role of empathy in an objective and a priori conception of human morality, I argue that the best explanation for how certain emotions function to morally motivate both animals and humans can be provided by the proposal that empathic modes of presentation of basic emotions contain another's welfare as part of their value-laden content. As we shall see, according to Slote's moral sentimentalism we are justified in morally approving of such emotions just in case we are (or would be) empathically warmed by the empathic motivation of agents. I supplement Slote's moral sentimentalism with the APV account of basic emotions in order to argue that this empathic warmth can express the value of an animal's empathic behavior to us through its contribution to our phenomenological state of moral approval as a vehicle of value. This supplementation helps moral sentimentalism overcome some potential problems with accounting for the accuracy conditions and appropriateness of empathy.

Animals only have morality if the empathic emotions they act on are more than just static reflexes that are genetically hardwired into the animal's nervous system. This is why I systematically present evidence in Chapter One for concluding that empathic emotions are developmentally modified over time to detect, with increasing sophistication, the well-being of others. Empathic emotions so construed are not preprogrammed reflexes, but genetically primed developing capacities to detect values important to the organism's psychological well-being, including the suffering or delight of others. When all goes well, these abilities gain increasing sophistication as animals

learn through normal mother-offspring and conspecific interaction, trial and error experience, implicit and explicit teaching, and socialization how to better interpret the phenomenological contributions of bodily states when those states function as background vehicles of value. After animals learn through the development of empathy to detect, discriminate, and act to alleviate the suffering of others and promote the well-being of others, then we are justified in claiming that these emotional values constitute moral motivation for their actions. To anticipate the combination of the APV account with Rowlands' tracking proposal and Slote's arguments for the necessity of empathy to moral motivation, I will argue that an animal is morally motivated just in case:

1. The organism has a peripheral awareness of locally contributing sensations of physiological states not merely as states of the body, but as *vehicles of value,* or in other words, as contributory representations *indicating* the *psychological value* that the intentional object has to the organism,

2. The organism's global phenomenological state includes a *non-mediated* focal awareness of an intentional object *having the value that is indicated by their peripheral awareness* of sensations of physiological states, and

3. Empathy, or having emotions, thoughts, or feelings more *appropriate* to the situation of the other organism, is the (motivating) component of the psychological value that the intentional object has to the organism through the phenomenological experience of a physiological disturbance or involuntary behavior.[11]

In the Conclusion I separate the critical arguments from the positive account of emotion and moral motivation put forward in this dissertation. While I defend a specific account of moral motivation and a specific account of basic emotions, the possibility and

[11] The notion of appropriateness, like the idea of an emotion being misguided, also requires significant theoretical unpacking, and so must remain for the time being rather vague until the full explanation I provide in part two of Chapter Four.

plausibility of ascribing animals moral emotions does not stand or fall on the correctness of these particular theories. Rather, these accounts together suggest a highly plausible way to explain how animals might be morally motivated. Yet even if these explanations should turn out incorrect, the critical arguments presented in this dissertation are sufficient to dismiss any well founded and unbiased reason to refrain from ascribing animals moral emotions. That is, the critical arguments in this dissertation establish that whatever the relation between empathic emotions and morality ultimately turns out to be, it is not necessary for moral motivation that an organism critically reflect on their purposes and goals or formulate linguistic or theory laden concepts about morality and the welfare and rights of others. When social mammals exhibit behavior that is emotionally motivated and empathic, we have no reason not to ascribe them moral motivation. However, if the positive theories of emotion and moral motivation that follow are correct, we have every reason to regard them as moral creatures.

Chapter 1: The Structure and Ascription Conditions of Human and Animal Empathically Modulated Basic Emotions

Summary

The aim of this chapter is to theoretically legitimize and clarify ascriptions of emotions to nonhuman animals. I do so by 1) systematically laying out the presuppositions that go into ascribing animals emotional states (that are similar to and comparable to human emotional states), 2) arguing that these presuppositions are defensible despite various skeptical concerns, 3) explicitly stating the qualifications necessary to justify each of these presuppositions, and 4) developing a model for animal and human emotions that is informed by empirical evidence and that takes on the relevant qualifications.

I begin in part one by arguing that several theoretical challenges that have been advanced by skeptics of either the comparability or the existence of animal emotions fall short of undermining the likely possibility that there are similar and comparable kinds of emotional states of animals and humans. However, these skeptical challenges do allow us to identify likely sources of errors that might be made in comparative claims about animal and human emotions. Accordingly, I address these potential sources of error by qualifying the way in which we can justifiably ascribe emotions to both humans and animals. Then, in part two, I clarify the nature of the evidence that supports such ascriptions. I argue that this evidence suggests a model of the phenomenological structure of basic emotions that I call the Awareness of Physiological Vehicle (APV) account. I suggest the APV model not as an account of all emotions, but as an account of the way that a subset of similar and usefully compared human and animal emotions are structured. If the proposed criterion of emotional ascription and the view of basic emotions I am

suggesting is correct, then there is convincing neurological, evolutionary, and behavioral evidence for the presence of the same kinds of basic emotions in animals and humans. I conclude part two by arguing that there are also good philosophical reasons for endorsing formal similarities in the intentional structure and value-laden content of these basic emotional states that justify the claim that some animals possess (some of) the same kinds of basic emotional states as adult human beings.

In part three, I argue that emotions are best thought of as flexibly developing in sophistication as an organism's capacities for empathy, memory, association, expectation, and categorization develop alongside its emotional capacities in its social and environmental interactions. I review the neurological, behavioral, and evolutionary evidence for this developmental conception of emotions, and conclude by arguing on the basis of this conception that animals ought to be ascribed "compassion" and "benevolence" (in a sense to be specified) as well as (at least) the basic emotional states of fear, sadness, happiness, and anger. They can be ascribed "compassion" in so far as there is converging evidence that animals are motivated by empathic suffering to act to alleviate the suffering of another, where the nature of this suffering is understood as experiencing pain or aversive emotions more appropriate to another in a way that internally represents the negative value of the others' suffering. Similarly, they can be ascribed "benevolence" in so far as they are empathically motivated to act to promote the well-being of another.

Part One: The Skeptical Arguments against Ascribing Basic Emotions to Animals

A continuous philosophical and scientific tradition that extends at least as far back as Plato and Aristotle holds that there are a subset of emotional states, including fear, anger, sadness, and happiness, that can be justifiably ascribed to both humans and animals.[12] However, the justification for such ascriptions has been challenged by both historical and contemporary philosophers and scientists.[13] The general form of the arguments that skeptics of animal emotions advance is to challenge the legitimacy of a presupposition purportedly involved in ascriptions of emotions to animals. For instance, to claim that an animal feels love for their offspring seems problematic in so far as we have reason to doubt that animals have a concept of the self and an ability to discriminate between themselves and their loved ones, both of which seem necessary for love. Additionally, love of one's offspring seems to imply the capacity to have knowledge of

[12] Plato's *Symposium* and *The Republic* and Aristotle's *De Anima* and *Nichomachean Ethics* develop or assume views of animal and human emotions along these lines. Both ancient Greek philosophers held that there were passions such as fear and love that animals and humans possessed; passions that needed to be either extinguished entirely and replaced by reason or moderated and shaped by habit and rational reflection in order to achieve the good life. Darwin's arguments and evidence for the evolutionary continuity between expressions of emotions in humans and animals set a more recent precedent for this tradition; one that is continually expanded and developed by many of his contemporary followers in evolutionary psychology and cognitive ethology. Darwin's *The Expression of the Emotions in Man and Animals* (1872) is the initial corpus of evidence that has been expanded on by contemporary ethologists, psychologists, and neuroscientists. It is also noteworthy that many insights of contemporary neuroscience into emotion have been the result of (often cruel, in my opinion) animal experiments where portions of the brain are removed or electrically stimulated and the presence of behavioral and neurobiological correlates of emotions are monitored (LeDoux (1996) and Panksepp (1982) are two prominent examples). The idea that animals have some basic emotions like fear also has contemporary appeal, and many ethologists are now comfortable describing and explaining animal behavior in emotional terms. For instance, Bekoff (2010 and 2007) offers explicit examples of how robust scientific research proceeds with justified explanations involving animal emotions.

[13] Carruthers (1996, 2004), Descartes (1649) Gallup (1970), Rolls (1999), Rosenthal (1986, 1993), Ryle (1949) Skinner (1974), and Watson (1925) have all denied that animals are affectively conscious beings. Davidson (1975, 1982), Chater and Heyes (1994), and Stich (1978) have argued that it is hopelessly problematic to ascribe beliefs (and by implication concepts) to animals. Finally, Nussbaum (2001), Sartre (1971), and Solomon (1973) have separately argued that emotions require complex normative judgments that are almost certainly inaccessible to animals. The qualification "conscious" is necessary because a variety of neuroscientists, psychologists, biologists, and philosophers are committed to the idea that the term "emotions" refers to unconscious neurological conditions and psychical drives as well as conscious emotions. My second assumption below addresses this point.

another's well-being and desires, and the capacity to judge that the well-being and desires of one's offspring are as or more important than one's own. These presuppositions, if true, cast the ascription of mental states like love to animals into doubt in so far as we have reason to believe that animals lack these cognitive capacities.

Even more basic emotions can be seen to be highly problematic when ascribed to animals. How, for instance, could an animal fear a threatening entity without possessing the *concept* of a threat, a *conception* of what the entity that threatens them *is*, and a *judgment* that the entity is threatening? Recently several philosophers, ethologists, neurologists, and psychologists that are persuaded by these kinds of concerns have separately challenged the idea that animals have (conscious) emotions. This chapter addresses these challenges and explains why we can be justified in ascribing at least the so called "basic emotions" of fear, sadness, anger, and (the occurrent feeling of) happiness (or joy) to creatures that lack the linguistic, evaluative, and introspective abilities of adult humans.[14]

---

[14] Fear and anger are intended to be examples of relatively uncontroversial '"basic emotions" that are often ascribed to both animals and humans, although it is noteworthy that animal fear is more undeniable than animal anger. The phrase "basic emotions," carries some theoretical baggage, and various somewhat divergent lists of basic emotions have been offered by philosophers and psychologists. Prinz gives a good comparative review of these proposals, which range from lists of four to fifteen basic emotions (Prinz 2004 pp. 86-94). The reason for the lack of agreement is that the list that a particular theorist proposes is hostage to both what the theorist takes an emotion to be and to the form of investigation that they privilege. The list will change, for instance, depending on whether the most important method of investigation is held to be discerning and investigating the conceptual requirements, the evolutionary fitness, the neurological correlates, or the behavioral correlates and constancy of basic emotions. This methodological choice in turn reflects what a particular theorist takes an emotion to be. Most lists converge on at least fear, anger, happiness/joy and sadness being among the basic emotions (disgust is also usually included, but in the way that ethologists and psychologists conceptualize disgust makes it seem more like a pure physiological reaction akin to hunger or thirst, and so I have excluded it from consideration here). Later I will argue that we are at least justified in ascribing to animals these four "basic" emotions. Fear is, of course, the paradigm of a basic emotion because of its distinctive neurological circuits, clear evolutionary advantageousness, relatively constant and instinctive behavioral manifestations, and simple conceptual requirements (that something be represented as threatening to the organism). Lust and love are usually excluded from the list of basic emotions, and while the reasons for excluding the former are most likely

I will make two initial assumptions. The first is that an epistemic inability to provide strict biological criteria for which organisms have emotions does not imply that there are no nonhuman species for which ascriptions of basic emotions are justified. One of the main projects of this chapter is laying out the way that the cognitive abilities of organisms, and the underlying neurological structures that make those abilities possible, constrain the types of emotions we can ascribe. An implication of this kind of project is that some organisms that are biologically classified as animals will lack the requisite capacities and some will possess them.[15] There may be no clear cut way to make this distinction in advance of a complete understanding of the relation between neurological systems and kinds of emotional states, and one of my contentions is that current neurobiology is a long way from understanding the link between phenomenal affect consciousness and the unconscious neurological subsystems that make it possible. Accordingly, the first assumption I make is that while there is no clear cut way to demarcate animals with emotional states from those with no (or fewer possible) emotional states, there are a variety of nonhuman animals, including all mammals, that clearly have emotions. The issue I am concerned with is not where to draw a line between emotionless and emoting animals, but rather *which* emotions that animals with the relevant capacities might justifiably be held to have, what *form* those emotions take, and the nature of our *evidence* for ascribing these emotions.

The second assumption I make is that while some emotions might be unconsciously realized, all emotions are at least potentially conscious phenomena, and an

---

based on a conception of lust as a pure physical desire, the reason for excluding the latter are precisely the complex cognitive capacities that love seems to require.

[15] For instance, insects, worms, and sponges, all of which fall under the Kingdom Animalia, are not promising candidates for beings that experience similar and comparable emotional states to adult humans.

emoting animal must have the capacity for conscious emotions. Since this is not the way that some neuroscientists and psychologists use the term "emotion," this is a bit of a stipulation on my part, but one that is defensible for the following reasons.

To the extent that emotions are philosophically interesting, they have a phenomenologically specified value conferring structure. If I am desperately in love, for instance, there must be a conscious conferral of the value of my beloved to me. The ascription conditions and nature of conscious emotions (as opposed to unconscious processes that underlie emotional states) are also prioritized in this chapter for methodological reasons, since only conscious emotions are relevant to the ethical claims I will later be making about justified ascriptions of moral emotions to nonhuman animals. Undoubtedly there are a variety of interesting questions we can ask about either the unconscious neurological processes underlying these states or the possibility of unconscious components of these states. However, what ultimately makes the emotion of love (or any other emotion) philosophically interesting is firstly, that it is a state which includes one of our most treasured (and sometimes our most deplored, as in the case of unrequited love or abusive love) phenomenal feelings, and secondly, *that it represents the beloved to us as having a value that is of great importance*. This point generalizes to all conscious emotional states, which are states that are value conferring and themselves valuable (or deplorable, depending on the emotion and the circumstance) to us. Accordingly, I will simply assume that the human and animal emotions that are of philosophical interest, and hence the emotions that I will be referring to and arguing about, are conscious and value conferring. The philosophically interesting question I am considering is whether we are justified in ascribing *these* kinds of emotions to animals.

With these two assumptions in mind, let us consider several theoretical and empirical challenges that have been advanced by skeptics of (the existence or the comparability of) animal affective consciousness. I will argue that while these challenges fall short of undermining the possibility or likelihood of similar and comparable animal and human emotions, they nevertheless should be taken seriously in so far as they allow us to identify sources of error that might be made in a comparative claim about animal and human emotions. First, however, it will be helpful to get a little more precise as to what an emotion is in the human case. This quick tour of the emotion literature will not (yet) provide us with necessary and sufficient conditions, but it will allow us to move forward with an informed and relatively accepted characterization of emotions.

Theories of emotion are almost as numerous as emotion theorists. As a result a colorful plethora of assimilations of emotions to judgments, moods, perceptual states, thoughts, physiological states, feelings, or some conjoined or causally related subset of these, have been proposed in the literature. However there are good reasons for resisting any of these reductions; reasons that each reductionist theory uses as ammunition against its opposition.[16]

For our purposes it will suffice to take stock of some of the reasons to resist a reduction so that we may consider the possibility of similar or comparable animal emotions with an informed (if somewhat negative) characterization of what an emotion is in the human case.

---

[16] This state of affairs should lead us to suspect, at a quick glance, that emotions are either not a natural kind (Griffiths (1997) defends this possibility) or, as I suspect, that emotions are really just an irreducible and primitive sui generis kind of mental state that we can merely characterize by explaining the conditions of the possibility of types of emotions and the relation of these types of emotions to other kinds of mental states.

To that end, notice that emotions, like thoughts and perceptions, represent a state of affairs, but unlike perceptions they do so in a value-laden way. Fear, for instance, represents the state of affairs that the agent is fearful about with a particular phenomenological character that is different from a thought or perception about that state of affairs – it represents a state of affairs as threatening. For example, I can perceive the staggering view of the city of Miami below from my 33rd floor apartment, judge that I am perfectly safe behind the guardrail, and yet still fear falling because the height that I am perceiving is phenomenologically represented to me as threatening. Furthermore, suppose that the guardrail was not there, and because of this I judge that the ledge and the high altitude are incredibly dangerous. It is still an open question whether or not I am also afraid of the height. As a trained and competent high-rise window cleaner I could take the rational judgment of danger in stride and walk out on the balcony without a hint of the phenomenological presence of fear. Phobias are examples of the opposite divergence of rational belief, judgments, and emotions. An agoraphobic can correctly judge that open spaces are not dangerous, rationally desire that they not fear open spaces, and yet still *fear* them. If these considerations are correct, emotions are not beliefs or rational judgments because they are neither necessary nor sufficient for the presence of an emotional state.[17]

[17] A quick reply from those who believe emotions are identical to value judgments would be that cases where a value judgment is made but no emotion is present are in fact cases in which the judgment is not *sincerely* believed. However, to say that people suffering from phobias or even more mild emotions that are counter to their rational judgments do not sincerely believe their judgment does not straightforwardly imply that value judgments *are* emotions, since it is consistent with emotional states simply being the sincerity conditions of some value judgments. This does not imply identity, it downright implies a difference between the judgment and its sincerity conditions. Nor will a further second order judgment be able to function as the sincerity condition of a lower order judgment, since this second order judgment would itself require sincerity conditions; generating an infinite regress.

This is not to say that emotions are just instinctual modular feelings completely isolated from cognition. Because of the intentional structure that emotions share with evaluation, an emotion is also usually distinguished from physiological feelings (tickles, itches, twinges) and global moods (depression, angst), neither of which need take a particular intentional object and both of which underdetermine the emotional state you are in. As Lyons points out, anger, fear, and jealousy have overlapping and underdetermining negative physiological feelings (of a raised heart rate, clenched stomach, etc.).[18] What individuates these feelings as feelings of anger or feelings of fear seems to be the evaluation that these feelings lead you to make of the situation that you are in and the way that you conceive of the target of the emotional state.

In the human case, emotions such as love or hate are emotions about a certain object that is represented as hateful or loveable.[19] This means that to have an emotion about an object, you must also have a representation about that object. In other words, the agent must possess a concept whose content is what the emotion is about, *and* a conscious, information-bearing representation whose content expresses the value-laden way that the object of the emotion is characterized. To sum up: an emotion is minimally characterized as a state that represents its intentional object in a value-conferring way

[18] See Lyons (1980) 115-130. Being depressed might involve any number of negative emotions or none at all, and instead merely amount to a general feeling of lethargic negativity about everything. Moods may be about everything that one's attention settles on or nothing in particular, and are therefore usually taken to be different from emotions which are about specific states of affairs. Nothing much turns on this distinction for our purposes, since moods may just be emotions with vague or general intentional objects. Later I will argue that moods are indeed intimately related to emotions because they are simply defocused emotions with vague, general, or yet-to-be-determined intentional objects.

[19] Although this result will make it harder to justify attributions of emotions to animals, it is necessary to characterize emotions in this way if we are to be justified in making comparative claims about human and animal emotions. I will later argue that animals have the representational and conceptual capacities that this characterization of emotions requires.

through a phenomenologically negative or positive experience that is not reducible to judgment, perception, mood, physiological feeling, rational desire, or belief.

It is important to note, however, that the conclusion that the assimilation of emotions to these other kinds of mental states is misguided does not also imply that some of these mental states are not necessary conditions of emotions. We should also note that part of this minimal characterization of emotions as presenting the intentional object in a phenomenologically *negative* or *positive* experience has been left intentionally vague. It is difficult to state precisely what makes a phenomenological experience of an emotion positive to the agent, since some emotions might feel uncomfortable but yet be useful, and some emotions might feel delightful but be detrimental. An emotion like mild performance anxiety, for instance, might feel unpleasant, but be considered a positive emotion to an experienced speaker or performer in so far as it motivates and drives them throughout the performance. Similarly, one can strongly love a person even when it is known that this feeling is detrimental to one's well-being. In this case the emotion involves a positive feeling but because of its known implications it might be thought to be negative to the agent. Phenomenological positivity and negativity also vary with context in a perplexing way, so that a typically positive emotion might impact an agent in a certain context in a negative way. An emotion like love, for instance, can be negative in some cases (abusive or unrequited love), and positive in others (mutual romantic love or friendly companionship). The same holds for emotions like anger, since to some unfortunate souls anger is considered a positive state despite its typically destructive influence on social relations, while to most of us anger is a negative state to be avoided if at all possible. For these reasons I have chosen to use the terms "negative" and

"positive" to merely mean that the emotional state has some kind of presentational impact on the agent, an impact that is value laden in a way that supersedes pleasure and pain and that is variable even within a given kind of emotional state. At this point, this is as much clarity that is possible at this level of analysis of these primitive phenomenological properties of emotional experience.[20]

If animals and humans are to have comparable and similar emotional states, then the kinds of emotions that they share must be such that they meet the requirements imposed by the above minimal characterization of emotions. It follows that in order for animals to be correctly ascribed emotions it must first be shown that animals are phenomenally conscious and that they possess emotional concepts.

The term "emotional concepts" as I intend it refers to the two kinds of conscious, information-bearing representations that are necessary for an agent to have an emotion about an intentional object:

1. The representation of the intentional object itself and
2. The value-laden information bearing representation that is required in order for the intentional object to be presented in a value-laden way to the agent.

If it is correct to ascribe emotional states to animals, then some account of the form of consciousness and of the emotional concepts employed by animals that is independent of human linguistic and critical introspective capacities must be given.[21]

---

[20] Later I will argue that when a pleasure contributes to representing the positive emotional value of an intentional object, we ought to understand the emotional state of the organism to be one of "delight," and when pain contributes to representing the negative value of an intentional object to an organism we ought understand the emotional state to be one of "suffering."

[21] There are good reasons to think that some animals are aware of their own reactions and there is overwhelming evidence that (at least) orangutans and chimpanzees (the genetically closest species to humans) have at least a minimal concept of the self that allows them to identify their reflections as

With this rough characterization of an emotional state as one that is at least potentially phenomenologically conscious, value laden, and intentionally structured in such a way that the conceptual repertoire of the agent constrains the type of intentional object the emotion can be about, we can now turn to the first of the skeptical challenges facing ascriptions of emotions to animals.

The most direct way to refute the idea that animals and humans have comparable and similar emotional states is to simply deny that animals are conscious. For different reasons, the now scientifically and philosophically disreputable traditions sparked by Cartesian mechanism and behaviorism considered animals to be unconscious biological machines that are incapable of experiencing sensations, thoughts, or emotional states. More recently, Higher Order Thought (HOT) theorist Carruthers and Higher Order Linguistic Thought (HOLT) neuroscientist Rolls have separately defended the view that animals cannot have phenomenal consciousness because they lack the capacity for higher order thought.[22] If phenomenal consciousness requires cognitive capacities that many or all nonhuman animals lack, then the unconscious physiological responses that constitute most animal emotions are distinct in both ontology and ascription conditions from

---

representations of their own bodies (see Gallup, G. G. Jr. 1977 Self-recognition in primates: A comparative approach to the bidirectional properties of consciousness. *AM Pyschol.* 32:329-38 for the original experiment, and see also De Waal *Good Natured* pp 67-68 and Griffin *Animal Minds* pp. 249-250). The point I wish to make here does not require empirical investigation into the possibility that certain species of animals have or lack introspective capacities. Instead, I will present a rather straightforward argument by example (to be supplemented shortly with further arguments) to show that self-reflection does not seem to be a necessary condition of being in an emotional state. Consider that I can be angry without any self-reflective awareness of my anger or even myself when I am consumed by rage after being shouted at from a moving car. In this case I am only aware of the person shouting, and this intentional object is presented to me in a phenomenologically negative and value-laden way. This plausible and common example of emotional states precluding or superseding any self-reflective awareness of the state itself implies that an animal may also be in a state of anger (by being aware of the intentional object in the relevant way) without being self-reflectively aware of the mental state of anger itself. In general powerful emotions often preclude or supersede attempts at self-reflection in humans, and so there is no reason to think that self-reflection is a necessary requirement of being in an emotional state. I will provide further arguments supporting this point shortly.

[22] See Carruthers (1996, 2004) and Rolls (1999) p. 262.

conscious human emotions. Assuming that it is a necessary component of being in an emotional state that there is something it is like to be in that state, then this would also entail that many animals simply do not have emotions, even though they have the neurological and behavioral correlates of emotional states.

Luckily, several conceptual considerations make HOT and (somewhat more so) HOLT very implausible as theories of phenomenal consciousness. In the literature HOT views have been convincingly criticized by Rowlands, whose arguments I will merely explain and supplement.[23] HOT and HOLT are both theories of consciousness that hold, very roughly, that you are conscious just in case you have a thought (or a disposition to have a thought) about a mental state. To be conscious in the sense we are interested in is simply for there to be something that it is like to be in that mental state. In keeping with tradition I will refer to this kind of consciousness with the term "phenomenal consciousness." Paradigm cases of phenomenal consciousness are perceptions and emotions, both of which necessarily have a distinctive phenomenology, or something that it is like to be in that state and without which you would not be in that state.

Conscious experiences are also usually characterized by necessarily having intentionality, or aboutness, in the sense that conscious experiences are always about an object or property. That is, an experience is always an experience that is *of* or *about* something.[24] So you might see a red rose, and think of its delightful scent, and both your consciously entertained propositional thought and your perception are *about* the rose.

---

[23] The original criticisms can be found in Rowlands (2009) pgs 182-187 and Rowlands (2001) "Consciousness and Higher-order Thoughts," *Mind and Language* 16, 3: 290-310.

[24] Nausea, depression, tickles, and the like present interesting challenges to the claim that all conscious experience is about something, yet I think this is merely a consequence of a semantic implication of the

To understand Rowlands' argument against HOT theories, it is important to note that consciousness can be ascribed to creatures or mental states. Mental states are not conscious of things, but creatures are conscious of things through conscious mental states. Consciousness can also be transitive or intransitive, where transitive consciousness is the property of being conscious of something, and intransitive consciousness is the property of being in a state of consciousness. Intransitive consciousness can be ascribed to creatures (for instance, when they are awake instead of in a dreamless sleep) and states, which are intransitively conscious when we are consciously entertaining them. In general, higher order theories of consciousness hold that a state is intransitively conscious if and only if a creature is transitively aware of it.

The higher order awareness of a creature can be characterized in various ways, and depending how it is characterized you arrive at a different kind of higher order theory of consciousness. HOT holds this higher order awareness is having a thought about the intransitive state, HOLT (higher order linguistic thought) holds that the first order

---

term "aboutness" localized to explanations of intentionality that use the term "aboutness." The phrasing in such explanations might be thought to imply that depression, nausea and the like must be about a particular and determinate object of consciousness. Yet the point I am making here is that a conscious experience of nausea in the sense of being nauseated is either about the feeling of nausea (or the physiological reactions of the body), the discomforting state of one's entire body, or a vague, yet to be determined, or general state of affairs. Similarly, a conscious experience of depression is about the state of affairs that depresses one, and this may be a fairly complex or general state of affairs, which you are aware of in the depressive mode of experience. Alternatively you might be aware of the inward feelings and thoughts that cause the depression, or of the state of depression itself. Nausea, depression, and the like may also be thought to sometimes not take any particular intentional object and instead be about the situation one is in or about life in general. Still, they are states contributing to the character of conscious experience and individuated by their value-laden contribution, and the total experience of that particular state is still about something in the sense that it is directed towards a state of affairs. In the case of depression, it is also worth mentioning that it might be the case that the intentional object that the emotion is about may be everything one experiences, feels, and thinks. Depression, like other moods, may have no particular intentional object because the representational scope of moods is yet-to-determined or so general that it is particularly wide ranging. We need not determine the correct way to characterize moods, however, since to the extent that a mood enters into a conscious experience that conscious experience will be about something, even though the mood itself might be a feeling that is not about anything in particular. In any case, the arguments I will advance against skeptics of animal consciousness do not depend on the universal quantification of the intentionality characteristic of experiences.

thought is in a mental language or somehow linguistic in nature, HOP (higher order perception) holds that this awareness of an intransitive state is perceptual in nature, and self-reflexive theories of consciousness (SR) hold that this awareness is a self-reflexive perception or thought. The higher order theories I am concerned with are those that rule out emotional consciousness in virtue of a lack of a capacity to have a thought about one's own mental states. To the extent that HOP and SR theories do not require this, they need not concern us here. To the extent that they do, however, the arguments here equally apply to these theories since they target this particular aspect of higher order theories.

There are really two versions of higher order awareness theories of consciousness: the theory may be actualist, in which case the higher order thought that makes a state conscious is occurrent. Alternatively, the theory may be dispositional, in which case the disposition to have higher order thoughts is what makes a state conscious. The general form that higher order theories take is to explain intransitive state consciousness in terms of transitive creature consciousness, where transitive state consciousness occurs if and only if a creature is transitively conscious of that state. For HOT and HOLT theories, transitive state consciousness occurs when, and only when, the creature has a thought (or a disposition to think) about a mental state. The thought may be occurrent, in which case the HOT theory is of the actualist variety, or the creature may only be disposed to have that thought, in which case the theory is of the dispositionalist variety.

With these distinctions in mind, let us look at the dilemma Rowlands develops against the general structure of explanation invoked by HOT theories. The first horn of the dilemma is that if the higher order thought is also intransitively conscious, then the

fact that the first mental state is intransitively conscious has not been explained. Rather the explanation has simply been deferred to whatever makes the second order thought conscious. We have no explanation of how the second order state becomes intransitively conscious, unless it too is targeted by a further conscious state, and that state is also targeted by another conscious state, and now we are well on our way to an infinite regress. The theory cannot explain how the first state (or any of the states in the regress) becomes conscious at all.

The other horn of the dilemma is that if HOT or HOLT theories accept that the higher order state is not itself a conscious state, then they run into the following problem. If this higher order state is not itself intransitively conscious, then it is unconscious. Let us assume that thinking about something unconsciously makes sense for the sake of argument. Since the higher order state that targets the lower order state is itself *unconscious,* you can have no idea *that* you are thinking that particular higher order state. Nor does it help to introduce a further higher order unconscious state, since any terminal state in the regress that is unconsciously about some lower state cannot be one that makes you aware of that lower order state. If you cannot know *that* you are thinking this higher order unconscious state, you certainly cannot know *what* you are thinking about in this unconscious state. In particular, you cannot know or be aware that you are thinking about the lower order conscious state.

Consider that you are in pain, for instance. If you are in pain in virtue of entertaining an unconscious higher order state that you do not know you are in, it seems impossible that this state of unconsciously thinking you are in pain makes you aware you are in pain. For how can a higher order state that you are not even aware of being in – the

state of unconsciously thinking about having an experience of being in pain – make you aware of the thing that it is about – in this case the state of experiencing pain? How can any unconscious state make you conscious of *something* that it is about while it is unconscious? If you are not aware of being in a certain higher order state, you also cannot be aware of (in virtue of that state) the lower order state that the higher order state is about. An unconscious occurrent thought cannot make you aware of what that thought is about precisely because it is not conscious!

This dilemma strongly indicates that having an occurrent higher order thought is not the kind of property that could explain why a state is phenomenally conscious. Rowlands follows this dilemma for the actualist variety of HOT theories by criticizing dispositional varieties in the following way. Consider again the usual paradigm of a conscious state: the feeling of an occurrent pain. This pain does not seem able to be constituted or explained by dispositions to have the thought that I am in pain. This is because, in general dispositions are not the sort of thing that make something have a certain property or explain in an informative way why it has a certain property. For instance, while the disposition to dissolve in water entails that something that has that disposition is water soluble, this is because being water soluble is a dispositional description of how an item's chemical properties interact in the condition of being in water. Dispositional explanations are descriptions of an entities' tendency to change or interact in various ways under certain conditions.

This can be seen more clearly with the disposition of brittleness, which does not make something break, nor does it explain *why* something breaks. Rather, it describes a tendency of something to break under certain conditions (something that breaks when it

is subject to a relatively weak force is brittle, something that breaks only when an enormous force is applied to it is not brittle).[25] Similarly, the disposition to have a thought that you are consciously experiencing something may be a true description, but it would be extremely odd to say that the disposition is what *makes* you consciously experiencing something in the first place. Rather the disposition describes a tendency of a creature to entertain higher order conscious thoughts about *already conscious* lower order mental states. Here I do not mean to be claiming that there are no occurrent properties that require dispositional properties in order to exist, but rather that the dispositional property of having a higher order (regressive or unconscious) thought under certain conditions cannot be what enables the occurrent property of being (intransitively state) conscious, nor can it *explain* why the occurrent property of entertaining an intransitively conscious state exists. A disposition to have higher order thoughts cannot even be a necessary condition for intransitive state consciousness because it must be a tendency of an organism that *already has those conscious lower order mental states* to think about them with a higher order thought.[26]

While the possibility that animals have phenomenally conscious emotional states is ruled out by certain higher order theories, these higher order theories turn out to be

[25] I disagree, then, with Slote's argument in "The Rationality of Aesthetic Value Judgments" (*JP* 1971) that some things break because they are brittle *and* subject to a relatively weak force and others because enormous force is applied to them. Brittleness is a disposition that is relative to force, context, and the composition of the item in question, but nevertheless it is a disposition of something to break under certain conditions rather than the property that makes something break for all cases of brittleness. A pane of glass might be very thin, and a weak force applied to it, and thereby rightfully be described as normally brittle, while a very thick and very strong pane of glass might be being pushed on by the enormous force of hurricane winds and hence be brittle in the sense that in those conditions it is subject to break if an additional relatively weak force is applied to it. It's resistance to the winds under those conditions just shows that normally we would not describe the glass as brittle, although if it were in a different context its properties interaction with a relatively weak force would change.

[26] It is also worth noting that Rowlands' dilemma is actually worse for dispositional HOT accounts, since in the cases that the disposition becomes occurrent the initial dilemma applies, which makes it even harder to explain why a disposition of this problematic state makes a state intransitively conscious.

independently implausible as accounts of consciousness to the extent that they require a higher order awareness to make the lower order state conscious. We can generalize the point to any higher order theory, including more restrictive versions that require that the higher order state or lower order state somehow contains a mental language like Rolls' HOLT theory. To the extent that such a theory requires that consciousness arises just in case an organism has a thought about a further mental state, it is ruled out as a plausible explanation by Rowlands' dilemma.

However, two important qualifications are brought to light by the HOT and HOLT skeptical implications for animal consciousness. The first is that while we can safely assume that animals are conscious, we should be wary of making comparative claims about the way in which they are conscious. Thinking in a natural language and entertaining higher order thoughts are significant shaping factors in the mode of representation that our conscious states often take, particularly with respect to more complex emotional states. The first qualification is that we must abstract as far as possible from these shaping factors if we are to accurately characterize animal affective consciousness.

The second qualification has to do with Nagel's skepticism about the possibility of discerning the subjective quality of experiences of other forms of life.[27] The reason Nagel is skeptical of this possibility is that the subjective quality of conscious experience essentially involves just one viewpoint: the conscious subject. We are necessarily constrained to our own particular form of consciousness and the imaginative and inferential tools at our disposal. Because of this, we can never know what it is like to be

[27] Nagel (1974).

a conscious being whose subjective mode of representation of phenomenal conscious is different from ours. As Nagel puts it with an apt example, we cannot know what it is like to be a bat by extrapolating from our own case, since this would merely be us imagining, from *our* perspective, what it would be like if *we* were to behave, think, perceive or emote like a bat, rather than gaining experiential knowledge of what it is like from the bat's perspective. And of course, a functional or causal description cannot tell us what the subjective quality of a bat's experience is because such descriptions are compatible with there being very different kinds of experiences or no experiences at all. If we cannot assume even perceptual experiences have the same qualitative and subjective phenomenological character across different forms of life, neither can we assume that emotional experiences are qualitatively similar.

Yet there are some more general properties we can know about emotions regardless of the form of life in which they are realized. If we are careful to limit our inquiry to discerning and describing the structural and functional properties of these mental states, rather than their subjective and qualitative feel, we can discover deep underlying similarities that allow us to make justified ascriptions and comparisons between the emotional states of different species. The solution, then, is to focus not on the subjective and qualitative feel of emotions for other forms of life, nor on the way that our capacities for reflective thought and language shape our own experiences. Rather we should focus on the structure and content of emotional states in a way that does not depend on indescribable qualitative feelings or higher order concepts. In this way we can describe emotional states abstractly and in an empirically tractable way; by explaining

what they are about, how they are structured, what kinds of values they imply, and the physiological and behavioral changes they invoke.

The fact that emotions are phenomenologically characterized as being *about* certain objects and properties is the first abstract and formal feature of all emotional states that does not refer to or imply any specific linguistic, introspective, or subjective and qualitative properties of an emotional state. The intentionality of emotional states is intended here as a necessary but not a sufficient condition for having an emotion, since aboutness might be possible without phenomenal feel. In order to explain the way that the emotions of animals target or are about harmful and beneficial objects and properties, we need to explain how emotions could have such content in the absence of linguistic or introspective capacities.[28] Accordingly, let us consider whether animals can be said to have any concepts at all, and then determine if it is possible to describe the nonlinguistic structure of animals' emotional concepts in a clear and empirically tractable way.[29]

Recall that at least two concepts seem to be essentially involved in the structure of any particular emotion. The emotion's intentional object is represented to the agent in the emotional state in a certain way, which we can refer to as "x's being F", where the subject concept, "x" refers to any particular entity or state of affairs that a being has an emotion about, and the predicate concept, "F," refers to the way that the entity or state of

[28] The assumption I earlier defended that the interesting emotions that humans and animals might share are conscious emotions is relevant here. I won't be taking on the Freudian's and arguing that all emotions are necessarily conscious here, but rather arguing about the emotions that are conscious and value conferring, since these are the philosophically and ethically relevant emotions.

[29] While it might be claimed that the content of animal emotions in non-conceptual, I am in agreement with Jeff Speaks on this issue, who has persuasively argued that there are no convincing (or relevant) arguments for or against the possibility of non-conceptual content largely because of a lack of clarity and agreement on the correct analysis of the terms "conceptual" and "non-conceptual." It is theoretically simpler to assume that animal thoughts and emotions are conceptual (but not linguistic or theory laden), and in the interests of avoiding unnecessary controversy I will make this assumption in what follows. See Speaks' "Is There a Problem about Nonconceptual Content?" *The Philosophical Review,* Vol. 114, No. 3 (July 2005).

affairs is represented to the agent. Ascribing emotions to animals then becomes problematic for two reasons. First, in so far as conceptual ascriptions of value laden predicate concepts are dependent on complex linguistic abilities and introspective judgments, these ascriptions do not seem to be applicable in the same way to both animals and humans. For example, when a capuchin monkey refuses to take a grape after witnessing a conspecific receiving a (more desirable) piece of candy, it is highly unlikely that they do so in virtue of anthropocentric reflections involving theory-laden concepts such as an unjust distribution of resources (although it is likely, in my opinion, that they feel a rudimentary sense of indignation that does not involve reflections on abstract principles of justice). Second, the capacity to have an emotion about x's being F, where x characterizes the entity *as the substance that it is,* depends on the agent having an accurate concept of x. If it is impossible for animals to have a concept of x, or if there is no meaningful sense in which animals and humans share a concept of x (because possessing a concept depends on having a natural language in which that concept is represented), then animals cannot have an emotion about x *as an x*. For example, if an animal is incapable of conceiving of their mother *as their mother*, then it follows that this animal cannot have an emotion about their mother *qua mother*. Even though emotions are not reducible to thoughts or concepts, a defense of animal concept possession is necessary if we are to justifiably ascribe emotions to animals.

It is common and natural to think that animals entertain nonlinguistic representational states that are necessary for them to have perceptual states about (and distinguish between) things like predators, mating partners, and food. Yet the characterization of the specific mode of presentation of these representational states is

problematic. The controversy begins when we ask whether these internal information-bearing states are similar to or comparable to the concepts that constitute conscious human thoughts. There are essentially two skeptical challenges: the first is to defend the possibility that animals have concepts and/or thoughts at all.[30] The second challenge is to determine whether we can understand, study, and model animal concepts and thoughts using qualified *human language* given the vast difference between human and animal cognition. Like the controversy over animal consciousness, this controversy is exacerbated by there being no settled understanding of just what a concept is, or the conditions under which a being possesses a concept, even in the human case.

To anchor this discussion, let us focus on the possession conditions of two paradigmatic human concepts: the predicate *red*, and the predicate *triangle*. The concept "red" refers to all and only red things, and our possession of the concept red is a matter of having knowledge of a certain phenomenal property (redness), classifying and distinguishing red from non-red things, and understanding that redness is different from all of the other properties of red things.[31] By contrast, possessing the geometric concept "triangle" requires something other than the ability to perceptually detect and discriminate the apparent property in the environment. In this case, possessing the

[30] Michael Slote has pointed out to me (in personal conversation) that, depending on one's requirements for concept possession, it might be possible that an animal has thoughts without concepts. I take him to mean that if we think of concept possession as something approximating knowledge of the necessary and sufficient conditions for that concept, then we would want to say that an animal has thoughts but not concepts. It would follow that in many cases humans also entertain thoughts, but not concepts. I think that there are problems with overly stringent requirements for concept possession even in the human case, and will provide an argument for this below. If concepts are just the constituents of thoughts and do not require overly stringent categorization capacities, then it is unlikely that animals should have thoughts but not concepts.

[31] It follows that if a particular animal (or species) is colorblind (or completely blind), then it cannot have the perceptual concept of redness. An animal's perceptual capacities constrain the types of concepts we are justified in ascribing to them, as we should expect. The point is that for any given animal, if they have the relevant perceptual capacities to identify and distinguish an appearance kind, such as redness, then this is what the possession conditions pre-theoretically seem to be.

geometric concept "triangle" does entail having knowledge of the geometric necessary and sufficient conditions (of being a three sided closed figure whose angles add up to 180 degrees and whose sides are composed of straight lines). This possession condition must be sharply distinguished from the possession condition that applies to having the perceptual concept "triangular-appearance," which can be had independently of possessing the geometric concept triangle. That is, just as a small child can identify and discriminate between triangular appearing objects and square appearing objects without knowing how a triangle's geometric properties necessarily differ from a squares, so to it is possible that an animal can possess appearance-concepts in the absence of abstract and uniquely identifying characterizations of the kind of object that classifies the appearance theoretically. Moreover, both a child and an animal can learn to perceptually distinguish triangle appearing objects from triangles to an increasingly sophisticated degree without knowledge of the geometric concept.

Suppose for a moment that possessing a concept is, in the strictest sense of the term, knowing the necessary and sufficient conditions for that concept. Holding fast to this definition of concept possession would mean that most adult humans possess only a very small subset of non-perceptual concepts (mainly logical, mathematical, and geometric ones). This might be thought unsatisfactory as it clashes with the intuition that you possess a concept if you can pick out paradigm cases that fall under that concept or if you understand some basic criteria that, while falling short of necessary and sufficient conditions, make your detection and discrimination of objects or properties that fall under that concept accurate for all cases relatively close to paradigm cases. If this is correct, I would possess the concept of giraffe whether or not I know the necessary and sufficient

conditions for giraffe-hood merely because I understand that paradigmatically a giraffe is a four legged creature with a tall thin neck, strange appendages on the top of their head, and a certain color and pattern of fur. Furthermore, I can extrapolate from this paradigm to relatively similar cases by using these paradigm-based criteria, and identify three legged giraffes with slightly shorter necks and discoloration of their coat as unusual giraffes rather than disabled horses or zebras.[32]

It is helpful to separate the necessary and sufficient conditions for falling under a concept from our imperfect epistemic access to that concept, so that possessing a concept includes cases where we have imperfect epistemic access to the ontological properties of the state of affairs that the concept refers to. Possessing a concept is simply to have a stable mental representation that constitutes beliefs, and is used by the organism to distinguish things that fall under the concept from those that do not on the basis of an implicit or explicit understanding of a paradigmatic or essential property. Knowledge of

[32] I disagree here with Chater and Heyes, who criticize what they call the "prototype" approach to concept possession in psychology for being based on highly questionable assumptions concerning the relation between perceptual clusters and natural kinds, and being difficult to distinguish from the highly problematic "exemplar" approach to concept possession in nonhuman animals. I think that there is a promising nonlinguistic approach to understanding and researching concept possession in such animals (and human pre-linguistic children) if instead of focusing purely on a central tendency of mere perceptual clusters we also consider the learning and discriminatory behavior of animals as it corresponds to observations of paradigmatic behaviors and interactions of those perceptually identified objects in the organism's environment. The paradigmatic behavioral interactions of cats with trees when cats are being chased by a dog, for instance, and the relevant perceptual clusters that allow a dog to observe such behaviors and interactions, both contribute to the dog's cognitive system and enable it to form the (nonlinguistic) concept of a cat and the concept of a tree, which combine to form the belief that the cat is in the tree. Paradigms and perceptual clusters map on to natural kinds in an imperfect way, but then again we weren't looking for necessary and sufficient conditions of natural kinds, but rather appearance-concepts that allow animals to identify, reidentify, and discriminate amongst, various objects in their environment. For prototype based approaches see Medin and Schaffer, (1978); Nosofsky,(1984); Posner and Keele, (1968) Lea and Harrison, (1978); Pearce, (1989) Glass, Holyoak and Santa, (1979); and Smith and Medin, (1981), for functional based approaches that focus on what animals learn to discriminate, see Keller and Schoenfeld, (1950), and Lea, (1984), and for Chater and Heyes criticisms see their Animal Concepts: Content and Discontent in *Mind and Language* (1994), particularly pp. 218-222.

precisely what this mental representation fully entails is not an all or nothing affair, but can be had in varying degrees of sophistication.

What then counts as evidence for an animal possessing a concept? Not all discriminatory behavior is a candidate. As Allen and Hauser have pointed out, ants discriminate live from dead ants by merely detecting and reacting to acidic byproducts.[33] When these byproducts are sprayed on a live ant, other ants will treat this ant as if they are dead and carry them outside of the nest. Partly because they cannot learn that they are making such an error we intuitively hold that they do not have the concept of death or the concept of a dead ant. What examples like this indicate is that some degree of stimulus or perception independence and flexibility with respect to the behavior that the concept mediates is required for justified ascriptions of concepts to nonlinguistic beings. Possessing a concept is not possessing the necessary and sufficient conditions for what falls under that concept, but neither is it mere perceptual discrimination. Rather concept possession must be to have an identifying and classifying ability involving a steady internal representation that lies in between these two extremes. With this rough characterization of concept possession as a discriminatory and classifying ability using an internal representation that is relatively perception independent and that flexibly mediates discriminatory behavior, let us turn to the reasons to be skeptical of animal concept possession.

In the literature there are two challenges to ascribing animals beliefs, and, by implication, the concepts that constitute these beliefs. The first is the problem of holism, from Davidson and Stich, which is the idea that a dense network of beliefs and concepts

[33] Allen and Hauser (1991) pp. 227-232.

that are inter-defined is necessary in order to have even one meaningful belief.[34] In order to have a belief about a triangle, for instance, I need to have a lot of other beliefs about lines, angles, geometric figures, and so on. Similarly, for a dog to have a belief about a cat being up a tree, that dog must have a holistic network of other beliefs about what kinds of things trees are, what kinds of things cats are, the capabilities of cats to go up trees, and so on. Yet, the argument goes, because we do not share a language-based network of holistically determined belief contents with animals, it is hopelessly problematic to assign particular beliefs to animals.

The second challenge to ascribing animals concepts or beliefs is a further worry from Davidson that it is also false to attribute *any* beliefs at all to animals.[35] Very briefly, the argument he presents for this is that in order to have a belief, you must be able to understand when that belief misrepresents. In particular you must be surprised when your belief turns out to be false. But in order to be surprised when your belief turns out false, you must understand that beliefs can accurately represent or misrepresent. So in order to have a belief at all, according to Davidson, you must be able to have a belief about a belief, namely that it can be true or false. For Davidson, the capacity to be surprised is necessary to having a belief, and in order to be surprised you must have a belief about a belief. It follows, on such a view, that you must have a belief about a belief in order to have any beliefs at all, and in order to have a belief about a belief you must have a natural language to represent the first order belief that your second order

---

[34] Davidson (1975, 1982) and Stich (1978). Rowlands has convincingly argued that animals can be ascribed beliefs when they are individuated by their *de re* referential properties rather than the internal mode of presentation of those beliefs. Here I take on the challenge of explaining and clarifying the evidence for structural properties that the internal mode of presentation of emotional concepts and beliefs must take, and so my arguments are intended to build on, but not depend on, Rowlands' arguments concerning the legitimacy of *de re* belief ascription in *Animal Rights* (1998-2009) pp. 196-202.

[35] Davidson (1975, 1982).

belief is about. The same argument would equally work for concepts: in order to have concepts at all you must be able to understand when they misrepresent, which requires having a belief about a concept.

A number of theorists have pointed out that this argument fails because it over-intellectualizes surprise. That is, it depends on animals not experiencing surprise when their belief misrepresents or when they misidentify something as falling under a concept.[36] Yet surprise behavior is a relatively common phenomenon in animals (you can verify this simply by fake throwing a ball for a dog to fetch and watching his or her puzzled reaction when it does not land). Understanding when your belief or concept misrepresents just amounts to an emotional[37] reaction of puzzlement or surprise when your expectations are violated.

Now since I'm arguing that beliefs and concepts are necessary for emotions, I cannot base my answer to Davidson entirely on the presence of (the potential *emotion* of) surprise, so instead I want to take a different route and suggest that certain communicative and error corrective behavior that happens when a concept or belief

[36] DeGrazia (1996) p. 149, Rowlands (1998/2009) pp. 209-211

[37] I am not sure whether it is correct to consider surprise an emotion, but here are some reasons for why I call it an emotional reaction here. Surprise can be, like other emotions, either "negative" or "positive" in the sense that I can be pleasantly surprised in some contexts and fearfully surprised in others. Surprise also seems to take an intentional object, which is the thing or situation that you are surprised about. Finally, there is a necessary phenomenological characterization of surprise: unlike the state of *coolly* reassessing your beliefs as you observe that a certain prediction or expectation is not met, a reaction of surprise entails that there is something specifically *hot*, (but perhaps irreducible and psychology primitive, and hence resistant to analysis) that it is like to be in that state. It would be interesting to see if surprise reactions have not only the same paradigmatic behavioral manifestations but also the same neurological pathway activations across a variety of species. I suspect that just as Damasio, Panksepp, and LeDoux's work on fear have shown that there are shared and preserved neurological and biochemical activation conditions across species (as well as behaviors), empirical research would confirm this speculation about surprise. This prediction is also empirically testable along the same lines of fear reactions – once a stimulus evokes a fearful reaction in anticipation of a pain or threat, the behavior, physiology, and neurology of an animal scan be assessed when the stimulus is not followed by the expected pain or threat. See LeDoux (1996) and Panksepp (2005) and Damasio (1994).

misrepresents is evidence of concept possession.[38] That is, if an animal were to alter their behavior in a way that displays the presence of a minimal semantic web of representations that are independent of rigidly determining perceptual cues, and if they are observed to systematically detect and correct for *errors* of conceptualization and classification, this would sufficiently address the challenge concerning the possibility of recognizing misrepresentation. Importantly, such abilities entail the possibility of recognizing when beliefs misrepresent without having a natural language or a belief about a belief, and so constitute counterexamples to Davidson's overly stringent requirements for belief ascription. When the expectations an animal has about their environment are not met, and they react by altering their behavior in a way that shows they recognize and correct errors of expectation and classification, then we have evidence for the more intuitive thesis that (animal and human) beliefs ontogenetically begin with simple environmental expectations and classifications rather than suddenly springing up in only human organisms(or language trained animals) when they acquire the ability to have second order thoughts via representing their thoughts in a natural language. There are two well documented empirical examples illustrating precisely these abilities in play.

The two examples of animal behavior evidencing conceptual capacities that lie in between the extremes of perceptual identification leading to rigidly determined behavior and full concept possession are the predator specific alarm calls of vervet monkeys[39] and prairie dogs[40], respectively. Both species have been observed to learn, use, and respond differently to at least four types of predator specific alarm calls. Vervets, for instance,

[38] Here I follow the work of Allen (1999) and Newen and Bartels (2007).
[39] For alarm calls in vervet monkeys, see Cheney and Seyfarth (1990).
[40] For alarm calls in prairie dogs, see Slobodchikoff, C. N. (2002).

make different sounds depending on whether the potential predator that is identified is an eagle, leopard, python, or baboon. The conspecifics that hear this cry vary their behavior accordingly, so if the vervet monkey utters an eagle attack alarm call other monkeys that hear the cry will hide under cover, but if they utter a python alarm call they will instead climb a tree. Similarly prairie dogs make different alarm calls in response to different types of predators (dog, coyote, human, or hawk) and vary their barking frequency proportionality to the relative speed of approach of that predator. The behavioral responses of prairie dogs also varies systematically with the type of alarm call and frequency of the call. These communicative calls and corresponding behaviors are usually taken to show that these species have a minimal semantic net with which they can represent and understand four kinds of predators (and in the case of prairie dogs, predator speeds). This stable representation of predators (the alarm call) functions across a variety of perceptual circumstances and allows them to identify and respond to types of predators from many different conditions of observation.

Importantly, positive and negative reinforcement is used in the corrective learning process of young vervet monkeys, where false or inaccurate calls are either punished or not positively reinforced by the other monkeys reiterating the call. This is quite possibly, Cheney and Seyfarth suggest, a mechanism for teaching infants and juveniles to utter alarm cries with greater accuracy and precision.[41] Empathic imitation, by which I mean the mimicry of actions of others (even when those actions are only appropriate to the others' situation), where that mimicry is motivated by an empathic emotion that the mimicking agent feels, is widespread in the mammalian kingdom and so also more than

[41] See Cheney and Seyfarth's "Vocal Development in Vervet Monkeys" *Animal Behavior* 34, 6 p. 1648 (1986).

likely plays an instructional role through the mimicking of adults who serve as the model for behavior and vocalization in the development of more accurate and precise alarm calls.[42] Greater weight is given to the alarm calls of more senior monkeys who have more accurate discriminatory abilities and are observed to be more closely attended to than younger monkeys when they utter an alarm call. The presence of a minimal semantic network, discrimination across a variety of perceptual circumstances, relative independence of a stable representation, error correction and detection, and learning on the basis of error detection, are all evidence that vervet monkeys possess at least four distinct predator concepts, which are abstracted from specific perceptual cues, and not based in a human language, but nevertheless which can be studied empirically.

Since both vervet monkeys and prairie dogs are communicating an understood representation about a feared attack from a predator, these alarm calls are of particular interest as they seem to implicate not just predator specific subject concepts but also the predicate concept of danger. Vervets and prairie dogs make a variety of noises that do not indicate danger to their conspecifics, and they respond to the alarm calls with physiological changes and behavior that would be symptomatic of fear in human beings. This suggests that they differentiate between calls that represent threatening scenarios and those that do not. Since the behavioral symptoms of fear correspond with these neurological symptoms, we have more evidence that somehow the predicate concept of x being a threat is represented internally. If it is present, the predicate concept of "threatening," does not rigidly determine behavior in the way that detection of acidic byproducts does in ants. Instead, the predicate concept when present in an emotional

[42] This claim is based on the evidence and arguments that the well-known ethologist Frans de Waal presents in Chapter 3 of his *The Age of Empathy* (2009). See in particular pp. 55-56 and 59-61.

state flexibly motivates a variety of fear responses including freezing, fleeing, and defending oneself depending on circumstances. It is also subject to developmental error correction, first because successful empathic mimicry is presumably pleasurable or less discomforting then the absence of such mimicry, and second since young vervets will be punished or not positively reinforced if they make a threat call when there is no threat just as they will be if they make the wrong predator call. While this concept is internally represented to the vervets in a way that we cannot precisely determine, we can still characterize it generally and locate it in a web of other concepts that represent other emotionally relevant features of the environment. It is also possible to describe the parameters of this concept and what it tracks through the results of careful and systematic ethological observation of error detection and correction behavior.

Recall that there were essentially two worries attaching to ascriptions of beliefs or concepts to animals, the first of which was that because animals could not understand when a belief or concept misrepresented they could not have concepts at all. Error detection and corrective behavior suffices to dismiss that worry. However, animals have a minimal semantic web in which such things as danger, type of predator, food, and offspring, are inter-defined that might still be radically different from a human beings linguistic/semantic web of concepts. Since the web of holistically determined meanings of an animal's conceptual repertoire is, seemingly, utterly foreign to the web of holistically determined linguistic and theory based concepts and beliefs for a human, the project of explaining animal concepts using qualified human language seems hopelessly problematic. That, in any case, is the objection from content holism.

Yet there is significant overlap between the meanings of certain human and animal concepts that, to borrow a term from Lazarus, track "core relational themes," or organism-environment relations that bear on well-being.[43] Organisms that have concepts (and many that do not) must be able to differentiate between food, predators, prey, young offspring, mating partners, and other environmental threats and benefits that bear on their well-being in order to survive and reproduce. Being able to track and flexibly respond to core relational themes like threats and benefits are traits that are evolutionary advantageous. While adult humans track threats and benefits to survival and reproduction in more sophisticated and culturally dependent ways than animals, this conceptual discriminatory capacity with respect to basic biological needs (hunger, thirst, etc.) is surely one humans share with many other species. The kinds of meaningful concepts that we can and should ascribe to animals begins with those that identify and discriminate between basic organism-environment relations that bear on well-being and reproductive success. The complexity and specificity of the concepts we can justifiably ascribe to nonhuman animals increases proportionately to the evidence we have for increasingly sophisticated discrimination, classification, and error detection and correction. However, because of the objection from content holism, even the overlap of core-representational themes will not ensure the legitimacy of ascriptions of beliefs to animals because there is no guarantee that the human concepts we use to describe the behavior of animals have the same holistically determined meanings as the concepts that an animal actually uses to navigate their environment.

[43] Lazarus, R. S. (1991) and see also Prinz (2004) pp. 14-16.

Recall that Davidson (1975, 1982) and Stitch (1978) were skeptical of the possibility of ascribing animals beliefs (and by implication, concepts that constitute those beliefs) that are comparable to human beliefs for precisely this reason. Both theorists have independently developed arguments purporting to establish that, because of the vast difference in the meaning of human linguistic concepts and animal concepts, it is hopelessly problematic to use human language to ascribe beliefs (and the concepts that constitute those beliefs) to animals. Essentially these arguments were based on the notion of content holism, or the idea that when we have a belief, such as the belief that "the squirrel ran up the tree," the concepts that constitute this belief receive their meaning from a web of interconnected beliefs and concepts; in this case beliefs about squirrels being furry, fast, climbing mammals; trees being sturdy and tall, having branches and leaves, and so on. These concepts (about branches, leaves, fur, climbing, mammals, and so on) also receive their meaning in part from their interconnected relations to other linguistic concepts, none of which seem to be concepts or beliefs that could be possessed by an animal. When we say, then, of an animal such as a dog, that he believes "the squirrel is in the tree," we are (according to the arguments of Davidson and Stich) illegitimately ascribing concepts and beliefs to the dog that a nonlinguistic animal cannot possibly possess.

To make matters worse, Chatter and Heyes have additionally offered skeptical arguments against the ascription of nonlinguistic *concepts* to animals. Chatter and Heyes' argue that there is no empirically tractable sense of concept to be found in human and animal psychology and philosophy that applies to both humans and nonhuman animals. This is because, they argue, we have no well-defined conception of what it

means for a nonlinguistic creature to have a concept (all proposals are either dependent on having natural language, ill-defined, or empirically intractable), and hence we cannot assess whether or not such creatures have concepts. (Chatter and Heyes, 210)

However, Mark Rowlands has offered a solution to both the problem of content holism and the problem of being unable to precisely specify and empirically determine the nature of animals' nonlinguistic concepts, which is based on the idea of relativizing content ascription to contexts. According to Rowlands, we need not specify the nature of animal concepts or determine precisely how those concepts are embedded in dense network of nonlinguistic beliefs in order to use human language to describe the behavior of nonhuman animals. This is because when we ascribe a belief such as "the squirrel is in the tree" to an animal (such as a dog barking at the foot of a tree), we can conceive of this ascription as *tracking* the content of the animal's belief, where tracking is understood as the following relation between propositions:

> (*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**. (Rowlands, *Can Animals Be Moral?* pg. 58)

Rowlands suggestions is that when we ascribe *p* to an animal, we are essentially providing an explanation of the behavior of the animal that is subject to empirical testing along the lines proposed by Allen (behavioral evidence of error identification and correction). When we ascribe to animal a propositional belief, Rowlands argues, we are not literally saying of the animal that they believe the linguistic proposition *p.* Rather, we should understand the ascription of *p as a way to explain the behavior of an animal.* The

actual belief *p** entertained by the animal nevertheless bears a conceptual relation towards *p*, because when we explain the behavior of an animal by saying that they believe p, we are actually ascribing to them *some* belief *p** whose truth is guaranteed by the truth of proposition *p*. In the case of the dog barking at the foot of the tree, when we ascribe to the dog the belief *p* that "the squirrel is in the tree", this ascription should be understood as an explanation of the behavior of the dog that entails that the dog has a different belief, *p*,* whose truth is guaranteed by the truth of the proposition *p*.

Recall that the problems of content holism and the lack of an account of the nature of nonlinguistic concepts both occur because human concepts receive their meaning (or are "anchored," as Rowlands puts it) from the context of a dense network of interrelated *linguistic* concepts that we cannot legitimately ascribe to animals. The tracking relation allows us to anchor content-meanings to contexts, so that we can describe animal behavior using human concepts and beliefs that *track* the animal's concepts and beliefs, no matter what representational states and what dense network of interrelated representational contents that give those states meaning for the animal are actually entertained by the animal. Rowlands formalizes this tracking relation and the anchoring of beliefs contents to contexts as follows:

> Let us begin with the idea of a nonanchored, context-free proposition, *p* . This proposition can be anchored to a context. I shall use the expression [H: *p*] to denote *p* anchored to the human context, and [C: *p*] to denote the same proposition anchored to the canine context. (ibid 59)
>
> Then it follows that:
>
> Anchored proposition [H: *p*] will track the anchored proposition [C: *p**] if and only if

> (1) If [H: *p*] is true then [C: *p**] is true, and
>
> (2) this truth-preservation obtains because of a reliable asymmetric dependence between the concept expressed by the subject term of [H: *p*] and the concept expressed by the subject terms of [C: *p**]. (ibid 60)

Using the example of a dog, Hugo, who chases a squirrel up a tree and then barks at the foot of that tree, the explanation of the dog's behavior using the belief that "the squirrel is in the tree," works as follows:

> (1) If [H: the squirrel is in the tree] is true then [C: the chaseable thing is up there] is also true, and this truth preservation obtains because of reliable asymmetric dependence between the concept expressed by "squirrel" (as anchored to a human) and the concept expressed by "chaseable thing" (as anchored to a dog, such as Hugo). (ibid 60)

The solution to the problem of holism and the problem of a lack of knowledge about the precise nature of an animal's concepts is to ascribe beliefs that are anchored to human context and that reliably makes true the animal's belief anchored to the animal's conceptual repertoire. There is no precise or particular belief that we ascribe to the animal when we anchor content ascriptions in this way, since "there will be a set of context-bound propositions {[C: *p**], [C: *p***] . . . *[C: p* n *]}" whose truth is reliably guaranteed by the truth of [H: p]. (ibid 62) In the case of the dog Hugo, he might for instance actually belief either that "the chaseable thing is in the tree," or "the fuzzy thing is in the tree," or "the eatable thing is in the tree." Yet whatever member of this possible set of beliefs that Hugo actually entertains is guaranteed by a reliable asymmetric relation to the truth of the anchored proposition that "the squirrel is in the tree."

The notion of a reliable asymmetric relation is important here, as it specifies the logical relation between concepts that must obtain if the ascription is to be justified. Mark illustrates this logical relation with an example where the world's most famous

pianist is afraid of dogs and happens to encounter the dog Hugo. If we imagine that, instead of barking at a squirrel, Hugo barks at the foot of a tree where the world's most famous pianist has climbed up because of his fear of dogs, then while the truth of the anchored proposition [H: *p*] "the world's most famous pianist is in the tree" guarantees the truth of the proposition [C: *p**], "the chaseable thing is in the tree," it does not do so via a reliable asymmetric relation. This is because:

> It is not true that, for all x, if x is the world's most famous living concert pianist then x is chaseable. His less timorous predecessor, let us suppose, was not. But it is true that for a dog such as Hugo (but maybe not for a Pomeranian: C is, in this way, different for Hugo than it is for a Pomeranian), for all x, if x is a squirrel then x is chaseable. This connection is a reliable one. (ibid 63)

The reliable asymmetric relation is, Rowlands suggests, best captured by universal quantification over the concepts expressed by the subject terms of propositions [H:*p*] and [C:*p**], respectively; where for any subject concept X partially constituting anchored proposition [H:*p*] and for any subject concept Y that is concept X's de-anchored counterpart and that partially constitutes anchored proposition [C:*p**], if something is an X then it is a Y. This can be illustrated through the pianist example by considering that it is not true that, for all X, if X is the world's most famous pianist, then X is a Y, or a chaseable thing (for Hugo). (ibid 63) The more courageous predecessor of this particularly timid pianist is not chaseable; whereas for all squirrels it is true they are chaseable things (for Hugo). (ibid 62-3) This theoretical apparatus allows us to legitimately use anchored human beliefs to explain the behavior of animals as occurring because they entertain some belief [C: *p**] that is anchored to the context of their holistically determined web of beliefs and whose truth is guaranteed by the truth of [H: *p*] via a reliable asymmetric relation. This translation schema and the continuity of core

representational themes across a variety of species (at least all mammals) suffice to dismiss the challenge from content holism and the lack of precision that is possible about the exact nature and content of an animal's nonlinguistic concepts. The skeptical worries from content holism and nonlinguistic concept specificity thus fail to establish that it is hopelessly problematic to understand, empirically study, and model animal concepts.

They do inspire two more important qualifications that will constrain comparative claims about emotional concepts. The first qualification is that when we characterize an animal's emotional state we must carefully abstract away from human concepts and beliefs that are language or theory based and that normally characterize that emotional state. For instance, when a vervet monkey utters an alarm cry at the sight of an approaching python, and other vervets respond by uttering the same cry and searching the ground for predators, we can take this as partial evidence that the emotion of fear is felt towards a ground predator by the vervets. To distinguish the conceptual content of the emotion felt by the vervets from a human's holistically individuated belief about a predator, we might qualify these concepts as at least four distinct concepts of "a certain kind of dangerous animal," and abstract away from any human beliefs that are language or theory based regarding the particular danger, being an animal in general, and being that particular kind of animal. We can accomplish this abstraction via Rowlands' de-anchoring strategy and the reliable asymmetric relation that obtains between the relevant propositions. In this way we can explain an animal's discriminatory behavior by using the (reliable, asymmetric) tracking relation that obtains between linguistic propositions that are plausible explanations of the animal's behavior and whose truth guarantees the

truth of the belief that the animal actually entertains (anchored to the context of the animal's web of beliefs).

Yet we should be cautious here: simply because a human cannot perceptually detect the difference between a triangle and a triangular appearance does not mean that a human lacks either concept, only that they cannot detect the property of the apparent triangle that makes it in fact not a triangle. An adult human would not be able to detect that a small apparent triangle is in fact not a triangle because there is microscopic space where two lines should connect, but this does not mean they lack the concept of a triangle. While we cannot rule out that an animal has a certain complex concept, the more language dependent or theory dependent the concept is, the less likely it is that an animal will possess it.

The second qualification is that determining what kinds of concepts should be ascribed to animals requires careful ethological observation and research that pays particular attention to what kinds of things a certain animal discriminates between, how this ability to discriminate operates, and how they acquire this ability. By careful observation it has been discovered that vervet monkeys and prairie dogs actually make more fine grained distinctions then the general concept 'approaching dangerous thing.' Emotional concepts necessary for experiencing fear of a predator are thus well confirmed by ethological observation, but we could not determine this simply by speculating about the minimal requirements for perceptual input and behavioral output reactions.

To summarize the qualifications inspired by these defeated skeptical challenges to animal emotion concepts and emotional consciousness:

1. Phenomenal consciousness, while present in animals, is likely radically different for beings with radically different cognitive capacities and so we cannot make justified claims about the identity or similarity of the *subjective quality* of emotional states across a variety of species.

2. However, we can make claims about the identity of kinds of emotional states as long as those emotional states are not essentially and solely constituted by raw phenomenal feelings but are at least partially constituted by their structural and functional properties.

3. The structure of emotional states is such that they involve concepts that determine what the emotion is about and how that state of affairs is represented *as valuable*, and so the range of emotional concepts and value concepts that are available to a given organism constrain the kinds of emotional states a being can have.

4. Emotional concepts (predicate concepts like *threat* and subject concepts like *ground predator*), while apparently present in animals, must be carefully qualified when described by abstracting away from human language or theory based concepts through the de-anchoring tracking apparatus proposed by Rowlands and by ethological observation and experimentation determining the kinds of identifying, re-identifying, discriminating, classifying, error detecting, and error correcting behaviors that the animal engages in.

With these initial theoretical challenges to the possibility of animals having emotions met and the relevant qualifications in place, let us now consider the nature of evidence for emotion ascription in general and determine what the nature of this evidence suggests about the structure of basic emotions that animals and humans are likely to share.

Part Two: The Evidence for Animal Emotions and Their Phenomenological Structure

It is important to separate, at this point, arguments and evidence for the ascription of emotional states to animals from the form that such emotional states plausibly take in

animals. While it is true that what counts as evidence for ascribing something depends on the nature of the state to be ascribed, and part of the nature of the state described will be specified by the conditions under which it is justifiably ascribed, there is an important distinction between the two. The ascription conditions for an emotional state will be equivalent to our best available evidence for the presence of that emotional state. These can be partially known independently of the ontological conditions for what an emotional state is just as there can be evidence that something is or contains gold before it is known exactly what gold is.

There are essentially four forms of evidence one can have for the presence of an emotional state. While no form of evidence is infallible, together they constitute the best available ascription conditions. The first form is simply introspecting and, by attending to your experiences and the way you are feeling about yourself and the world, determining if you are in an emotional state. Of course, you can be mistaken as to whether you are in any particular emotional state, either because of self-delusion, confusion as to what the emotional state is, the possibility of unconscious or intractable emotional states, or a failure of attention (to mention just a few sources of first person error).

If, however, we want to determine whether someone *else* is in a certain emotional state, their *sincere* self-report is usually taken as constituting evidence for the presence of that emotional state. Again, this form of evidence is fallible. Much turns on the concept of sincerity, for anyone can mistakenly or with the aim to deceive agree or disagree that they are in a particular emotional state. What determines whether or not they are sincere in their assent or dissent, we might ask? There are four possibilities:

1. they are truthfully reporting on their phenomenological state to the best of their abilities, in which case all the sources of error associated with introspection are possible,
2. their self-report as to whether they are in a particular emotional state is confirmed and supported by their behavior, which is characteristic of being in that emotional state,
3. their self-report is confirmed and supported by certain physiological changes they undergo, which are characteristic of being in that emotional state,
4. or their self-report is appropriate to the environmental and social situation they are in, which paradigmatically invokes the emotion in question or to which the emotion in question is the paradigmatic or appropriate response.

So reports of being in an emotional state depend on sincerity, and sincerity in turn depends on these four further types of evidence. It is worth drawing attention to the fact that a first person report is a form of behavior, albeit one that gives us much more precision then other forms. Pacing rapidly back in forth can be done because of jealousy, anger, frustration, self-loathing, or impatience, and a verbal report is far more helpful to the project of specifying which state the agent is in than observation of this nonverbal behavior. Of course, as the Ekman studies (and many others) have shown, there are also subtler and involuntary nonverbal behaviors, particularly with respect to facial expression in humans, which are more determinate and often taken to be more reliable than self-reports.[44] Minimally, small and uncontrollable behavioral and physiological changes that express emotional states are needed as corroborating factors when determining whether self-reports are sincere and whether introspection is accurate.

[44] Ekman conducted a number of multicultural studies of minute changes in human facial expressions of emotions that led him to conclude that there were some universal human expressions of basic emotions. (Ekman 1971 and 1977)

However, behavior and physiological changes are also incomplete, fallible, and indeterminate evidence for the presence of emotional states. The reason that behavior is fallible and incomplete has already been stated with respect to verbal behavior. Any behavior that is done purposefully, or with the intent of the agent, can be done not as an expression of an emotional state but for some other purpose, such as intent to deceive. Yet let us take a closer look at physiological changes and unconscious reflexive behaviors, which can, for our purposes, both be classified as physiological disturbances.

Because we can study physiological disturbances right down to the neurological systems that get activated, they present a powerful form of evidence for emotional ascription when strongly causally correlated with a type of emotional state. When triangulated with behavioral evidence and paradigmatic "appropriate" responses to situational and social settings, physiological changes are a convincing form of evidence for the presence of an emotional state. These distinctions should help to clarify what was really at state in the famous Schachter and Singer experiments, which are usually cited in defense of cognitive views of emotion that consider emotions to be or contain evaluations.[45]

In the experiments, subjects were injected with adrenaline to cause physiological arousal and put in various happiness or anger invoking circumstances (insulting questionnaires and an offensive actor or simply a room with a joking around actor), and then interviewed as to what emotional state they were in while their nonverbal behavior was observed. The Schachter and Singer studies suggest that the nonverbal and emotionally expressive behavior varies according to whether the subject is put in negative

[45] Schachter, S. and Singer, J. (1962).

or positive social circumstances following the adrenaline injection. This indicates that (some kinds of) physiological arousal underdetermines but primes subjects for an emotional state. Schachter and Singer then posited that it was the subject's interpretation of their physiological state, an interpretation that was influenced by context and their evaluation of that context, that determined their emotional state.

However, as has been noted by De Sousa, subjects cannot be said to be in precisely identical physiological states (there are, trivially, fine grained neurological differences if their behavior and emotional evaluation is different).[46] Additionally, as has been noted by Prinz, some subjects reported being happy to the interviewer even when their nonverbal behavior was indicative of anger (these results were interpreted as insincere self reports to please the interviewer).[47] Since each form of evidence (self-report, behavior, physiological changes, and environmental context) is incomplete and underdetermining with respect to the emotional state the subject is in, we should expect evidential conflicts of this sort. The fact that adrenaline evoked arousal does not completely determine or function as sufficient evidence for a particular emotional state is an implication of the fact that these four forms of evidence are just that; evidence for an emotional state *but not the actual presence of it*. When these four forms of evidence agree they constitute the *best available* (but still incomplete and fallible) evidence.

The ascription conditions for an emotional state are then as follows. To the extent that we have confirming and fine grained neurological evidence, behavioral evidence, contextual evidence, and self-reports, we are justified in ascribing an emotion. There are

[46] DeSousa (1987) pp. 54-55.
[47] Prinz (2004) pp. 70-72.

of course further breakdowns of the types of patterns of neurological activity that are causally correlated with specific kinds of emotions, as well as increasingly fine grained breakdowns of the minute and reflexive facial expressions, eye movements, and behavioral patterns associated with these emotions. The details of these breakdowns are an important empirical matter, and as such not to be determined by philosophical theory.

What is important for our purposes is the dependence relationship between a self-report and the sincerity conditions for that self-report being a matter of behavior, context, and physiology. It is plausible that these nonverbal ascription conditions mirror the following three essential phenomenological components of emotions. The first of these components is that an emotion has a value-laden phenomenology that is correlated with and often expressed by involuntary behavior and physiological changes. The second component that the contextual ascription condition suggests is that through these experienced physiological changes and involuntary behavior the agent is focused on a particular situation that is experienced as bearing some value to the organism's psychological wellbeing. Finally, the emotion is motivating with respect to actions that lead one to avoid the harm or seek the benefit of the object or state of affairs that is the emotion's intentional object. To illustrate what I have in mind, consider the example of the emotional state of occurrent happiness (or joy)[48] and how these three forms of evidence each mirror a phenomenological component of the emotion and suggest the

[48] By both the term "occurrent happiness" I mean the immediate state of (possibly unreflective) joyfulness about some state of affairs (the intentional object), and do not mean to suggest that eudaimonistic happiness is an emotional state, nor do I have in mind any kind of "flourishing," but simply the state of joy that animals experience when they are playing or content, like the quite obvious happiness of dogs when playing with each other. That occurrent state is what I have in mind by the term "happiness." I will later explain the relation between happiness (an occurrent emotion) and joy (a mood).

following structural analysis of how the conceptual and phenomenological requirements interact with the involuntary physiological and behavioral changes or states.

The verbal report of occurrent happiness depends for its sincerity on behavioral evidence (including both involuntary and voluntary behaviors), physiological evidence (neurological evidence, changes in heart rate and muscles, perspiration, etc.), and contextual evidence (is the context one in which the emotion of happiness is appropriate). An emotion of occurrent happiness motivates the organism to be near, acquire, or engage with the intentional object and engage in typical happy behaviors for that species (tail-wagging for dogs, laughing for humans, etc.), some of which contribute to the phenomenological value of the intentional object in a way to be specified. Additionally, an emotion of happiness requires a pleasant phenomenological experience (warmth, attraction, etc.) that is the result of the physiological changes that constitute typical evidence for an organism being happy (the presence of endorphins, opiods, activation of the relevant brain areas, etc.). Finally, the emotion of happiness requires that the environmental context that the organism is in be represented as beneficial to the organism's psychological well-being. In this way the three primary forms of evidence each strongly suggest a phenomenological component of the emotional state of occurrent happiness, and the same can be said for the other basic emotional states.

Fear, for example, can occur in the absence of an internal judgment or externally uttered verbal report that something is threatening, but not without the phenomenological structure of presenting something as threatening. Fear can also occur in distinctly nonthreatening contexts, but not without the agent phenomenologically experiencing that scenario as if it is threatening, and representing the threat and its

threatening character internally with a steady internal representation that allows for flexible behavioral response. Fear can occur without any particular physiological changes (although some physiological change is causally necessary), but not without some pleasant or unpleasant feeling of physiological change, such the clenching of the gut, speeding up or slowing down of the heart, skin prickling, sweating, etc., all of which are evidential or sincerity conditions of certain emotions. Fear can also occur in the absence of behaviors characteristic of fear, such as flight or freezing, but not without motivating one to somehow escape from or remove the threat.

An emotion like fear might be had in the absence of one or all of the evidential ascription conditions (behavioral, physiological, and contextual). However, each ascription condition tells us something about the phenomenological character of emotions because the best explanation of the fact that behavior, physiology, and environmental context are ascription conditions is that they are expressions of different dimensions of the phenomenological character of emotions. In particular they tell us that the most likely structure of an emotion will involve the components that I claim are mirroring these ascription conditions. In other words, the reason that behavioral, physiological, and contextual sources are evidence for the presence of an emotion is that the structure of the emotions phenomenology is such that an emotion represents the intentional object at least partly through the experiential contribution of behavioral and physiological changes and at least partly through the perceived context.

Emotions essentially feel like something different from ordinary thoughts and sensations; in fact, they are the very paradigm of our most dramatic, extreme, and exotic phenomenological states. We use narrative and value laden features, contextual factors,

physiological descriptions, and descriptions of behavior to describe a certain kind of emotion. However, these features, factors, and descriptions are often expressed in poetic and artistic ways rather than as analytic value judgments or identifications of threats and benefits. This is because while we can describe the parameters and necessary conditions of emotions abstractly, the phenomenological feeling through which these abstract conditions are expressed is something that cannot be completely described from the third person scientific standpoint.

These considerations suggest a way to abstractly characterize emotions that can be ascribed to both humans and animals while avoiding the pitfalls of attempting an accurate and kind determining description of the subjective phenomenology of particular emotional states for particular organisms. These ascription conditions for the sincerity of self-reports suggest that the structure of emotions is such that an animal or human has a basic emotional state just in case:

1. The organism has a peripheral awareness of locally contributing sensations of physiological states not merely as states of the body, but as *vehicles of value,* or in other words, as contributory representations *indicating* the *psychological value* that the intentional object has to the organism, and

2. The organism's global phenomenological state includes a *non-mediated* focal awareness of an intentional object *having the value that is indicated by their peripheral awareness* of sensations of physiological states.

I call this account of basic emotions the Awareness of Physiological Vehicles (APV) account. The concepts of vehicles of value, peripheral and focal awareness, and indication that are employed in this formalization require significant further explication.

First, in order to understand what I mean by 'vehicles of psychological value,' consider how a physiological state of a painfully empty stomach can indicate something beyond an internal representation of its own state to the organism. If an organism is in presence of delicious food, the physiological sensation of having a painfully empty stomach contributes to the phenomenological state of desiring that delicious food. Similarly, a felt physiological state of trembling, tensing, heightened heart rate, and so on can indicate the threatening nature of the emotional stimulus eliciting this response. This is, very roughly, what I mean by physiological states functioning as vehicles of value. Experiencing such physiological states as vehicles of value is to be aware of these states as indicators of psychological value.

The notion of indication employed in the formalization should be understood as teleological. The physiological states that the organism peripherally senses have the evolved *function* of indicating or representing the organism-environment relations of (at least) threats, frustrations, losses, and benefits. The notion of teleological representation that I favor (although this theory of basic emotions is not committed to any particular account of teleological representation being correct) is that defended by Millikan.

Following Millikan, I will understand teleological theories of representation as those that explain the representation relation of mental states in terms of the relational proper function of the mechanisms that produce those mental states.[49] The proper function of a mechanism, trait, or process is what it is supposed to do or what it should

[49] Here I am indebted to Rowlands' defense of Millikan's account of teleological representation in his *Animal Rights: Moral Theory and Practice* (pgs. 213-217). The original view can be found in Millikan's *Language, Thought, and Other Biological Categories*. Here I will use Rowlands' arguments for a teleological account of the referential component of mental representation in animals to explain how sensations of physiological states can be sensations of those states as states that indicate, or refer to, normative properties of intentional objects of emotions. All references to Rowlands in this section are to *Animals Rights: Moral Theory and Practice.*

do. A heart, for instance, has the proper function of pumping blood, sperm has the proper function of fertilizing the ovum, and a hammer has the proper function of hammering nails. Hearts do not always pump blood, however, just as sperm does not always fertilize the ovum and hammers are not always used to hammer nails. Hence the proper function is what an item should do, which is its evolved or intended *normal* function. The normal function of an entity is thus not necessarily what it actually does, but rather what it has been designed to do or what it is supposed to do. (Rowlands 213)

Mechanisms with proper functions can acquire this normative quality through the purposes of a designer or the blind process of natural selection. In the latter case, certain mechanism, traits, or processes enhance the reproductive capacities of the organism by enabling it to cope with its environment. Proper functions are thus relational in the sense that they are relative to the environmental feature they detect or respond to and ultimately relative to gene reproduction. Chameleons' skin, for instance, has the relational function of making it the color of its immediate environment, which evolved through selection pressures and the advantages afforded by the camouflaging mechanism that helps it to avoid predators. (Rowlands 214)

It is in this sense of having an evolved normal function that teleological theories claim that certain neuronal mechanisms have the proper function of producing representations of features of the environment relevant to the organism's survival and reproductive success. Hence Rowlands proposes that the referential component of the representation relation between mental states and the environmental features they are about is: "If cognitive state S is produced by mechanism M, and if the proper function of M is to produce S in environmental circumstances E, then, according to the teleological

theory, S represents, or is about E." (ibid 215) In this way the internal states of an organism represent environmental objects and properties "in virtue of the fact that the mechanisms which produce such states have evolved to produce them in circumstances where a given environmental item is present." (ibid 216)

This account of the representation does not entail either that mental states themselves have relational proper functions or that all mental content is exhaustively accounted for by relational proper functions. Rather, the teleological explanation of evolved mechanisms with the proper function of producing certain representations is intended to account only for the way that a mental state can refer to or indicate items in the environment. (ibid 215-216)

Now the proper functions relevant to this theory of basic emotions are those of the physiological changes that enter into an organism's peripheral awareness. It is the evolved proper function of these physiological states is to indicate to the organism the relevant normative properties of intentional objects of emotions. This includes (at least) the normative properties of being a threat, benefit, loss, or frustration for the organism. When an organism is in a state of fear, their peripheral awareness of the physiological changes of trembling, tensing, raised heart rate, and so on, is a peripheral awareness of these physiological states indicating the threatening value of the intentional object. The value indicated by physiological changes will vary depending on the emotional state they organism is in. I will clearly define these values for the basic emotions below. For now I will focus on the emotion of fear for simplicity. The reason that an awareness of these physiological states can be an awareness of them as indicating normative properties of the intentional object of the emotion is that these states have the relational proper

function of indicating such properties. I have at this point explained how physiological states function to indicate normative properties to an organism. However, the mode of awareness of these normative properties (being a peripheral awareness) has yet to be explained. It is to this mode of awareness I now turn.

The difference between peripheral and focal awareness has been explained by Kriegel as "primarily a distinction regarding the distribution of attention: focal awareness of something is highly attentive awareness of it, peripheral awareness of it is less attentive awareness." (Kriegel 360) He adds that since this distinction depends on attention and attention is a matter of degree, peripheral and focal awareness are poles on a single spectrum. (ibid 360) As I shall be using these terms, the focal awareness of an intentional object of an emotional state is focal relative to the peripheral awareness of sensations of bodily states. In other words, when one is in an emotional state, one attends more to the intentional object and less to the sensations of bodily states.

To illustrate how this distinction works with a simple example, consider your current peripheral awareness of the environment at the edges of your vision as you are focally aware of the words on this paper or screen. The awareness you have of the environment at the edges of your vision is one involving less attention then the focal awareness you have of the words your attention is directed towards. Similarly, when in an emotional state you are less attentive to (and hence less aware of) the sensations of physiological states then you are of the intentional object of the emotion. It is also important to note that while you can usually direct your attention to different intentional objects at will, the salience of the intentional objects of emotions will result in your attention being directed in a way that bypasses or overrides your will.

The role of peripheral attention in emotions is vitally important to specifying the mode of awareness that an organism has of an intentional object and the sensations of physiological states involved in an emotional state. To see this, first consider that there are many locally contributing sensations that partially make up the global or complete phenomenological state of, for instance, and approaching predator. When an organism senses an approaching predator, they experience fear and their attention is (usually) focused on that predator. However, the global phenomenological state of the organism involves more than just a representation of an approaching predator. In particular, the global phenomenological state will involve many local contributions of the various sensory modalities and proprioception, including the sound, sight, and smell of the approaching predator.

Importantly, these local contributions also include various sensations of physiological changes. Such physiological changes will typically include trembling, tensing of muscles, raised heartbeat, sweating, and various involuntary behaviors such as crouching, flinching, freezing, or preparing to run. All of these physiological changes make a local contribution that makes a difference to the global phenomenological state of the organism. According to the theory proposed here, feeling fear entails that an organism has a focal awareness of an intentional object, and a peripheral awareness of all of the locally contributing sensations of physiological changes that enter into the global experience. Together, these local sensations indicate the threatening value of the intentional object to the organism. Finally, it is important to note that the present theory does not require a specific physiological profile for each of the basic emotions, but just

that some physiological changes are sensed in a peripheral way as indicating the emotional value of the intentional object.

For humans, it might be possible to direct one's attention away from the predator to specific local contributions to this global phenomenological state of fear. If a human were to attend to any of these physiological changes, they would become focally aware of that change. Being focally aware of one particular physiological change they will then be only peripherally aware of the approaching tiger and even more peripherally aware of the other locally contributing sensations (including other sensations of physiological changes). This might seem to entail that if one were able to completely focus on a physiological change such as one's raised heartbeat and think of it only as a physiological change, then one would no longer be in the state of fear that initially triggered that physiological change. However, as long as one's peripheral awareness still includes other sensations of physiological changes indicating a threat and that awareness of sensations is less attended to than the awareness that one has of an intentional object of an emotional state, then one is still in a state of fear. This is why I specified that the focal awareness of an intentional object of an emotion is focal with respect to the peripheral awareness of physiological sensations that function as vehicles of value. In the case we are considering, a state of fear about an intentional object still occupies one's attention more than the peripheral awareness of other locally contributing sensations of physiological changes but less than the focal awareness of a raise heartbeat. In practice this kind of control over one's attention may be highly problematic because emotions prime us to be attentive to the intentional object and its value.

It may also be possible in principle (although perhaps not for normal adult humans or other mammals) to direct one's attention completely on a physiological state as a physiological state in a way that also excludes all other items of peripheral awareness normally accompanying a state of fear. To do so would be to no longer attend to the intentional object of the emotion or the value indicated by sensations of the normally locally contributing physiological states. This is why a common counterexample to somatic views, which is pointing out that one can be aware of any local physiological change (raised heart rate) without having an emotion (fear), cannot lead us to dismiss this theory of basic emotion. This selective and exclusive attentiveness to a particular local sensation of a physiological change is not the form of awareness that the current theory proposes partially constitutes an emotional state, even though the physiological change may be present in both the emotional state and states of mere introspective awareness of bodily changes.

While in a state of fearing an approaching predator, a human or other mammal will not be focused on any of these local phenomenological contributions as bodily changes. Instead they will be focused on the predator, and they will experience the predator as threatening. Their global experience of the predator as threatening is distinct both from normal sensations of physiological changes and from rational value judgments about the approaching predator. I submit that this difference is that when an organism is in an emotional state they have a peripheral awareness of the phenomenological experience of all the local physiological changes that contribute to their global phenomenological state by functioning as value indicators. A basic emotional state will always involve local contributions of sensations of physiological changes that enter into

peripheral awareness as indicators of value. This is why I call this kind of peripheral awareness an awareness of physiological states as *vehicles of value*.

This account corresponds somewhat to Damasio's "somatic marker hypothesis," according to which the "process of continuous monitoring, that experience of what your body is doing *while* thoughts about specific contents roll by, is the essence of what I call a feeling." (Damasio 1994 pg. 145) Damasio goes on to state that, "When the bad outcome connected with a given response option comes into the mind, however fleetingly, you experience an unpleasant gut feeling." (ibid 173) However, Damasio understands emotions (which he refers to as 'feelings' to distinguish them from unconscious emotional processes) as merely requiring an image of the state of the body juxtaposed, combined, or superimposed over another thought or image for the presence of an emotion. (ibid 146) This analysis of emotion is too broad, as one must be aware of changes in the body as background indicators of value rather than as mere states of the body. For consider that if I come to be aware of my trembling while perceptually experiencing an approaching dangerous predator as a mere bodily symptom of my mild coldness (a biological rather than psychological value) rather than as indicating the psychological value of a threatening situation. In this case I am not in a state of fear even though the situation I am perceptually aware of or thinking about is juxtaposed or combined with my awareness of trembling. The same follows for the other basic emotional states, all of which require the registering of bodily states as indicators of psychological values rather than mere indicators of any body-environment relations (hunger, thirst, hotness, coldness, externally caused pains, etc.).

According to this Awareness of Physiological Vehicles (APV) account, the core relational themes for the basic emotions of sadness, happiness, anger, and fear are values that the organism is aware of in virtue the contribution of their physiological state to the value of the intentional object. The reason that emotions are motivating, phenomenologically distinct from judgments (with same content), can be had without language, and have (as the neuroscientist Panksepp has shown) homologous physiological and neurological substrates across mammalian species, is that the core relational themes are conveyed to the organism through their (felt) physiological states.[50]

This account can also be usefully contrasted with the somatic account that Jesse Prinz has recently defended, where one is aware of X as frightening in virtue of an awareness of physiological changes. This awareness is such that X is presented as frightening through a problematic mediating state of indirect awareness of bodily changes that are calibrated to thoughts and perceptions by reliable causation. (Prinz 99-101 and 144) I will address the problems for Prinz's somatic account in more detail below. My dialectic purpose here is to use the contrast between the APV account and Prinz's somatic account to explain why the awareness-with relation is needed to explain the intentionality of emotions in a way that avoids the usual theoretical problems associated with indirect awareness accounts like the one defended by Prinz.

To begin to see how this mediating indirect awareness of bodily states is problematic, consider how Prinz explains the intentionality of the emotion of being sad about the death of a child (Prinz 62). Because according to Prinz an emotion is an indirect awareness of bodily changes that are calibrated to thoughts and perceptions by

[50] See Panksepp's *Affective Neuroscience: The Foundations of Human and Animal Emotions* for a systematic review of the homologous neural substrates of the basic emotions in mammalian species.

reliable causation, his view seems to imply that the emotion of sadness is not about the child dying at all. This emotion is rather either about my bodily states that I am indirectly aware of, or it is about my judgments and beliefs about the child dying that are calibrated (through a reliable causal relation) to certain modes of awareness of physiological conditions. However, if I am sad about the death of a child, I am not sad about a judgment, a calibration of that judgment to a kind of physiological awareness, or whatever reliably causes my physiological changes. In particular, I am not sad about the reliable causes of any particular physiological state I am in when these causes (alcohol, neurological conditions, conversations about death, etc.) and the intentional object of the emotion (the death of the child) differ; I am simply sad that the child has died. I will have more to say about why Prinz's view fails to capture the intentionality of emotions like sadness below, for now I merely want to contrast the APV account with Prinz's view. According to the APV account, the sadness I feel in this case does involve real or simulated bodily changes, but these bodily contribution (tears, a sinking feeling in one's stomach, etc.) are experienced as background indicators of the value of the intentional object (the death of a child). It is through these bodily changes that we experience the sadness we feel, because when we entertain an emotion we are aware of the value of an emotion's intentional object *with* our actual or perceived bodily changes rather than being aware *of* them as bodily changes (that are calibrated or connected by reliable causation to judgments, perceptions, etc.).

The APV account can also be contrasted with a simpler adverbial account, where one is aware of X as frightening in virtue of the frightening mode of experience that

modifies the experience of X, such that X is presented in a frightening way.[51] The reason that it is preferable to explain the awareness of a felt physiological or behavioral state as a vehicle of value that expresses the value of the intentional object through the subject's background awareness of their physiological state, rather than as an indirect awareness (as Prinz's indirect awareness of embodied appraisal account implies), or as simply an unnoticed mechanism of modulation that changes the character of the (noticed) overall experience (as the adverbial account implies), is that while the value laden contribution to consciousness of felt bodily changes is essential to the nature of emotions (and hence part of the explanandum for a theory of emotion), the physiological state does not enter into conscious experience as a mediating indirect awareness of a physiological state. It seems inaccurate to say that you are not aware of a physiological change at all when you are in an emotional state. There must be an awareness of the contribution of a physiological state without that awareness being an indirect awareness of a bodily state somehow combined with a further mental state. The solution I propose is that there is an awareness *through* a bodily state's contribution to phenomenology that retains the bodily states visceral and felt character as a contributory expression of the value of the intentional object. It is in virtue of being aware *with* a bodily state or involuntary behavioral state that one is aware of that bodily state's contribution to consciousness when one is aware of the positive or negative value of the intentional object. The occurrent bodily changes are not just causally responsible for an emotional state, rather they actually enter into the phenomenology of emotions, which explains why when we have emotions distinctive and relatively localized neurological patterns that are also activated when we are aware of

[51] Rowlands suggests that such an approach can be used as *part of* an explanation of the form of moral sensitivity in (some) animals. See his *Can Animals Be Moral* pgs. 224-227)

bodily states have been identified (Damasio calls these "somatic markers" when they "mark" an image or thought as good or bad for the organism). (Damasio 1994 pg. 173 and 1996 pg. 1415) This account also explains why involuntary physiological changes and behaviors across all mammals have been identified to correspond with at least basic emotional states like fear without implying that organisms are aware of changes in their body as mere changes in their body that happen to accompany particular thoughts or experiences (Le Doux 147, 165-166, and 171-175). Further evidence for this view over the adverbial account is that when the neural mechanisms underlying the background bodily awareness that partially constitutes emotional states (according to APV) are damaged in humans, emotional deficits and irregularities result. (Damasio 1994 pg 62-69, 211 and 1996 pgs 1417-18) Although the adverbial view could explain this as the result of a causal precondition of emotions being absent, what is left unexplained is that when the mechanisms of bodily awareness are absent in patients with neurological conditions like anosognosia (right hemisphere damage in the somatosensory system responsible for both the external senses of touch, temperature, pain, and the internal sense of joint position, visceral state, and pain), a radical reduction in the emotional *experience* reactions results; as evidenced both by *the self-reports of such patients* and their behavior. (Damasio 1994, pgs. 64-5) These causal neurological mechanisms seem to enable the *bodily states* of a normal adult human being to function in the phenomenology of emotion itself, rather than just serving as general neurological causal preconditions for the simple adverbial states account of emotions as not essentially involving a bodily awareness-with relation. This confirms the APV account's suggestion that emotions

essentially involve a background awareness of (actual or neurologically simulated) bodily changes as indicators of psychological value.[52]

Additional support for the APV theory can be garnered from considering psychological disorders where inappropriate emotions, such as a fear of something harmless, to be cases where the visceral contribution of the body inaccurately indicates that something has a certain psychological value. In the case of fear, the visceral contribution of the body inaccurately indicates that a harmless entity or situation is threatening. A theory of emotions ought to be able explain why in these circumstances an inappropriate emotion is felt rather than an appropriate one. For example, when a human feels fear of playful encounters with others rather than feeling happiness around and towards such others against their better judgment, a theory of emotion ought to be able to explain why one mode of experience rather than another is appropriate in these circumstances, and the psychological mechanism that results in an inappropriate emotion. It seems that explaining why the inappropriate mode of experience can persist despite the organism's better judgment is more difficult for the adverbialist account than for the APV account defended here. This is because the adverbialist account cannot appeal to a background awareness of bodily conditions inaccurately indicating the threatening value of playful others via an awareness-with relation, but must rather simply hold that the subject's experience is being modified one way rather than another with no explanation of the psychological mechanisms leading to inappropriate and persisting irrational emotions. By contrast, the APV view, unlike Prinz's account or the adverbialist account,

[52] But see Dunn, et al 2006 for a dissenting opinion concerning the empirical evidence for the somatic marker hypothesis. For my purposes it does not matter whether somatic awareness is neurologically realized in the precise way that Damasio proposes, only that (through some neural mechanism) a background awareness of the physiological changes that constitute well-studied profiles for the various basic emotions contributes to the phenomenology of emotions.

can explain this psychological mechanism while still accurately characterizing the relation of emotions to intentional objects. On the APV account, you are aware of the physiological state of, for instance, trembling, but not just as a physiological state. Rather you are aware of the threatening value of the intentional object *with* that felt state of trembling.

This account of the structure of basic emotions, can be further elaborated by contrasting it with some standard reductionist accounts of emotion. In what follows, I will advance the APV view of basic emotions as the best explanation of the emotions that we share with animals, not with the aim of being the last word in emotion theorizing, but rather as one of many possible ways of explaining how animals have emotions that can express and reflect values and that can be empathically modulated. As we shall see, empathic emotions that express the value of another's well-being are the most likely candidate for explaining how animals can be morally motivated, and so it is important for our purposes to defend the APV explanation of the structure of such emotions against other possible accounts of emotions.

There have been many attempts to explain emotions in terms of physiological states and/or cognitive evaluations, yet no such attempt has ultimately been successful. Since these points have been made before (albeit in different form), I will be brief in my arguments against the standard reductionist options. It is necessary to review their flaws to understand how the APV account improves upon Prinz's attempt to overcome the characteristic difficulties of a somatic account of emotions while retaining the theoretical strengths of reductionist cognitivist views. Let us begin with the James-Lange theory of emotions, according to which an emotion is an awareness of a physiological state. This

account has widely been regarded to be deficient because awareness of physiological changes is neither necessary nor sufficient for being in an emotional state. To see this, consider that the raised heartbeat you feel after too many cups of coffees does not constitute the emotion of fear, nor does the flush you feel when overheated on a summer day constitute anger. Awareness of bodily changes simply is not sufficient for being in an emotional state. Yet neither is this awareness necessary.

When you are in a state of fear about something, your heartbeat may be quickened and you may become aware of this fact, just as when you are in a state of anger you may become aware that your cheeks are flushed, but this awareness is not necessary. You may merely be angry or fearful about the state of affairs that elicits this emotional response, such as feeling intense fear of an approaching attacker, or feeling consuming anger at a grave injustice, with the state of your body never entering into your attention. To be sure, there are many more finely detailed physiological states you may become aware of while in an emotional state, the limiting factor only being that those states be accessible by introspective or perceptual means. A possible response from an advocate of this reductionist option that follows from this consideration would be that being aware of all, or most, of the set of all such states typically accompanying particular emotions constitutes an emotion, rather than being aware of any one particular bodily change.

However, not only will the sets of bodily states you are aware of overlap with each other such that the individuation of emotions will become highly problematic for such a view, we can also imagine that for any particular member of this set of bodily changes that you can be aware of, you can also be aware of that bodily change without having the corresponding emotion (as in the case of feeling your flushing cheeks or

feeling your heart race). It follows that a possible kind being, call them Somatic Aliens, can be aware of every member of the set at once without being in the corresponding emotional state, even if this is impossible for human beings. Such beings would be introspectively aware of every member of a set of bodily changes typically accompanying an emotion, such as the trembling, freezing, raised heartbeat, and so on, that typically accompany fear. However, Somatic Aliens would be aware of those states only as changes in their body rather than as emotional states about the object of fear. Somatic Aliens are emotionless beings with exceptional capacities for introspection and perceptual awareness of their bodily changes, and as such constitute a counterexample to reductionist somatic theories. Our intuitions about such cases as not being instances of emotions but rather instances of bodily awareness indicate that even a complete awareness of physiological changes is not sufficient for an emotion.

A further well-known problem with the James-Lange view is that because emotions are about things other than bodily states, an awareness of a bodily state change does not adequately explain the intentionality of emotions. The intentionality of emotions cannot be explained by such a view without adding something to an awareness of bodily states so that being aware of a bodily state can also be being aware of something in the world (Prinz and Damasio defend views committed to this conceptual move, as do I). For instance, it might be claimed that bodily changes are transitive representations of the world, or that an emotion requires the juxtaposition of a thought with an awareness of a bodily state (Damasio explicitly defends this possibility in *Descartes, Error*, pg 146), yet both of these qualifications make emotions something *in addition to* an awareness of physiological states and changes rather than holding that

emotions are constituted by bodily awareness. Finally, it seems possible that an organism constituted out of a very different kind of physical structure ought to be able to have emotions while not having an awareness of any of the changes of the body that humans, mammals, or even vertebrates typically have when they are in a particular kind of emotional state. That is, physiological changes that are typical of emotions like fear or anger do not seem to be necessary for emotions because creatures that are constituted very differently from us ought not to be considered emotionless on a priori grounds.

A similar thought experiment rules out the view that emotions can be reduced to evaluative judgments. For any particular evaluative judgment that partly constitutes a set of evaluations that you must make in order to be in an emotional state, it is possible that the particular evaluation simply occurs to you as a fact about your relation to the world, and not one you feel any particular emotion about. The danger of the height I am at as I am flying in an airplane is something that I can think about without feeling any fear, as I have become accustomed to air-travel. Someone attracted to the notion that emotions are evaluations might then reply that I would feel fear if I were to make many evaluative judgments at once or in quick succession, such that I become aware of the set of evaluations about the unsafe state of the structure of the airplane, the immediacy and salience of the danger, the relative lethality of the fall, and so on. However, as long as I can make any one of these judgments without feeling the emotion, it seems to follow that a kind of being is possible, call them Cog Aliens, that can be aware of each member of the set of evaluations at once as relational facts without feeling any emotion. If it were not for the actual case of human psychopaths devoid of the normal range of human emotions but able to make sophisticated rational judgments, the conceptual possibility of

Cog Aliens might seem like a mereological fallacy, similar to claiming that since each brick in a building has the property of weighing less than 10 pounds, the building built out of these bricks must also weigh less than 10 pounds. However, there does not seem to be any property of rational judgments (devoid of any actual or simulated physiological changes) that can function relationally with other rational judgments to generate an emotion when these judgments are had simultaneously or in quick succession. Unlike the mereological fallacy involving bricks, the case of judgments seems to be case of mental states that do not have a necessary property or a relational property to other judgments that can function cumulatively or in Gestalt fashion to result in an emotion. This is significantly supported by the case of psychopaths apparently capable of making evaluative judgments but incapable of feeling the normal range of emotions that accompanies these judgments in normal adult humans (although of course whether psychopaths *truly* understand evaluative judgments like "causing unnecessary pain to others is wrong" can be contested). I am not claiming that such a property is conceptually impossible, but only arguing that we cannot even imagine what such a property would be. It follows that we are theoretically justified in concluding that rational judgments are simply not logically sufficient for emotions. This is because it is possible that Cog Aliens could form sets of judgments about the relative value of objects and states of affairs that correspond to the values expressed by emotions, but they could do so as coolly as psychopaths, without feeling one way or another about the values implied by evaluative judgments.

It is also possible to show that forming evaluations is not necessary for emotions. If it is granted that it is possible to feel an instinctual fear of an approaching dangerous

object (a snake lunging for one's face, for instance), without having the explicit thought that this object is dangerous, then it ought to be possible to be an organism that always feels emotions in this way. This may seem somewhat question-begging, but since instinctual fears, joys, rages, and so on are the most common and basic emotions in not just humans but, arguably, all mammals (I will provide behavioral and neurological evidence for this claim shortly), then any account of emotions must explain these basic instinctual emotions that are prior to rational judgments, and this seems impossible for a reductive account. As the degree of cognitive sophistication that is achievable by members of a species decreases, it becomes more plausible to think that the form the emotions of such creatures would take would be precisely of this kind. Reptiles, for instance, might feel fear and aggression always as instincts and never have evaluative thoughts about the intentional objects that these emotions are about. Although some kind of physiological awareness, and some kind cognition, are probably necessary in order to feel an emotion, the thought experiments and problems that I have raised for such theories indicate that one need not perform a conscious evaluation or be aware of a bodily change in order to feel an emotion.

Because of these counterexamples and problems, it will also not help to combine these accounts along causal or constitutive lines. Since evaluations and physiological awareness are both neither necessary nor sufficient for emotions, it follows that holding that emotions are evaluations causing physiological changes or vice versa would merely be adding an unnecessary causal condition to an already flawed kind of explanation. It would still be possible to construct counterexamples as cases where an awareness of a bodily state is not an emotional state and the conjoined evaluation is similarly

emotionless. Constitutive conjunctions of the two kinds of requirements is similarly flawed; as evaluations are neither necessary nor sufficient for feeling emotions, and so unless an awareness of bodily changes is both, adding such an awareness as a causal or constitutive condition will not suffice to explain emotions. The same follows for adding an evaluation to an awareness of a physiological state. So much for the standard reductionist explanations of emotions. What is needed is a specification of the way in which thought and physiological changes combine that explains why some cases of this linkage are cases of emotion and others are not.

Here I must be careful, however, since on my view having a thought (broadly conceived so as to include perceptual identifications) and a bodily change that one is indirectly aware-with as the vehicle of value are necessary for feeling an emotion. If a background awareness of a physiological state causes a thought about the intentional object of the emotion rather than just a thought about that physiological state, such that the physiological state is taken to be indicating the value of the intentional object, then we have, I submit, a genuine emotion. Yet this is not the same as a direct awareness of a physiological state causing an evaluation. I can be aware of my gut clenching, and this can cause me to coolly judge that I am in a dangerous situation, but unless I take my indirect awareness of my gut to be expressing the danger of the situation, then we simply have a cool judgment caused by an awareness of a bodily state.

I will have more to say about the flaws of specific versions of these reductionist views below, but for now it suffices that the APV account, which holds that emotions are constituted by a mental state about an intentional object that is "colored in" by a

background awareness of physiology *as a vehicle of value,* is distinct from and more plausible than these extensionally inadequate reductionist strategies.

According to the APV explanation, you are in an emotional state just in case you are aware of an intentional object and the contribution of the physiological state to phenomenology simultaneously, where one's awareness of the intentional object is an awareness of a figure with a certain value, and that value is given by the value-laden background of that figure. This value-laden background is in turn generated by the physiological state's contribution to phenomenology as a vehicle of value.

The APV account shares many features with the embodied appraisal explanation of emotions that has been proposed by Jesse Prinz in his *Gut Reactions* and *The Emotional Construction of Morals.* In particular it shares with his embodied appraisals account a partial endorsement of an idea initially proposed by Lazarus: that organism-environment relations that bear on well-being are the general structure of the value-laden information that emotions transmits to the organism. I think that Prinz's reliance on the work of Lazarus is well founded and separable from his problematic ontology of emotions. Accordingly, it will be worthwhile to see what kind of organism-environment relations Lazarus thinks might be expressed in emotions and why these relations justify ascribing emotions to animals that are of the same kind as basic emotions in humans. But first, it is necessary to see in more detail why Prinz's embodied appraisal view of emotions is inadequate as an account of even human emotions, and how the APV account can avoid the difficulties that plague this view.

Prinz endorses a version of a theory of emotions that has been defended by James and Lange, where emotions are thought of as perceptions of bodily states. As we have seen, awareness of bodily states, when so roughly construed as a theory of emotion, seem to be neither necessary nor sufficient for experiencing and emotion. However, this view has been rather ingeniously refined by Prinz, who has done so using many of the same contemporary psychological, behavioral, and neurological experiments that have informed and shaped my own view.

As prima facie evidence for his view, Prinz offers a conceptual argument, which he calls the subtraction argument. The subtraction argument asks us to imagine a particular emotion, say fear, and then subtract every possible bodily manifestation of that emotion. When we have subtracted every bodily manifestation of fear, it seems as though we are no longer inclined to call the remaining thought or judgment an emotion. (Prinz 206) However, because bodily manifestations of emotion are essential on both the embodied appraisal and the APV account, this argument cannot be used to establish Prinz view over the APV account I defend. Additionally, this argument shows at most that certain physiological conditions are necessary, not sufficient, for emotions. Finally, this argument is significantly undermined by the possibility of beings with very different physical constitutions from humans having emotions without being aware of anything remotely resembling a mammalian physiological change. The APV view, by contrast, can handle this kind of counterexample as long as such beings are aware of some physiological change as a vehicle of value, however different from the normal mammalian changes.

The neurological evidence for the view Prinz defends comes mainly from the work of Damasio, who has performed neurological studies indicating that when people experience emotions the brain areas that detect bodily changes become active. (Prinz 58) Again, this is not evidence for Prinz's view over the APV account advanced here, which also predicts this result, nor is it evidence for the sufficiency of "embodied appraisals." In any case, the subtraction argument and the work of Damasio are prima facie evidence that some kind of awareness of physiological states are implicated in emotions.

The specific nature of this awareness according to Prinz is as follows. Emotions are composed of two parts: valence markers and embodied appraisals. Embodied appraisals are thoughts and feelings that (1) represent core relational themes, (2) register the body's preparation for action, and (3) prime congruent memories. (Prinz 244) The difference between emotions of negative and positive psychological value is a difference in valence. The negative and positive valence of emotions are determined by inner state Prinz calls "inner positive/negative reinforces" that function like imperatives directing us to sustain positive emotions and eliminate negative emotions. (Prinz 173-4) Yet notice the circularity here: valence determines the positivity/negativity of emotions by functioning as an inner reinforcer. How do inner reinforces reinforce? They impel us to avoid negatively reinforced experiences and seek out positively reinforced experiences. The notion of reinforcement assumes the positivity and negativity that the notion of valence was intended to explain. It follows that valence cannot be explained in terms of inner reinforcement without circularity.

So much for valence markers. Let us turn to the other conjunct of Prinz's account: embodied appraisals. In his own words; "To qualify as an appraisal , a state

must represent an organism-environment relation that bears on well-being." Prinz calls such relations, following Lazarus, a "core relational theme" (Prinz 77), and holds that emotions represent core relational themes without explicitly describing them by tracking changes in the body that reliably co-occur with organism environment relations. Importantly, for Prinz, embodied appraisals are calibrated by judgments (reliably caused by judgments of that kind). This leads to the creation of calibration files, or data structures in long term memory that establish a link between judgments of a particular kind and emotions of a particular kind. Each file contains a set of representations that can each trigger the same or a similar pattern of bodily responses. When this triggering occurs, "the perceptions of the bodily responses caused by representation in a calibration file are emotions." (Prinz 100) Calibration files can contain a variety of representations, from explicit judgments to sensory states (so a new scent on one's wife's clothes or the judgment that one's wife has been unfaithful can be part of a calibration file). Finally, for Prinz calibration files are causes, not constituents, of emotions (hence emotions are embodied appraisals and do not contain calibrating judgments as constituent parts). (Prinz 99-101 and 144)

However, introducing calibration files (clusters of beliefs, memories, and judgments as reliable causes of the awareness of the physiological state) into the embodied appraisal theory results in several problems for Prinz's view. Because neither judgments/beliefs nor modes of physiological awareness are necessary for emotional states, it is quite easy to construct counterexamples where a subject has a "calibration file" that reliably causes a physiological awareness but is not in an emotional state. Imagine, for instance, that via localized electrical stimulation of the RAGE neurological

circuit, every time I clap my hands a neuroscientist activates in my body the physiological symptoms of anger. Being aware of the fact that my body is being stimulated rather than expressing a true emotion, it is no stretch of the imagination to see that I do not actually enter into a state of anger (at clapping? The neuroscientist? Or anger at myself being the agent of clapping?). Still I form a calibration file containing various memories about the result of clapping, my condition, and the expectation that whenever I clap I will enter into this state of physiological awareness of bodily conditions normally associated with anger. Imagine that over time these thoughts become calibrated to the bodily states symptomatic of anger, and eventually come to cause anticipatory bodily reactions that are similarly symptomatic of anger. Still, because I am aware of the classical conditioning that is happening to me, I do not become angry at clapping (or whatever the intentional object of the thoughts in the calibration file is held to be) as I become angry at other things, such as parking tickets, flight delays, or slights from the neuroscientist responsible for this classical conditioning. Rather I am merely aware of the physiological changes being caused by my now calibrated beliefs about clapping. In this instance we have a calibration file (several memories, expectations, and beliefs) that has become a reliable cause of an awareness of the bodily states symptomatic of anger. Still, it is plausible to think I am not in a state of anger but rather in a (calibrated) state of introspective awareness of the bodily symptoms of anger having been classically conditioned to clapping.

A second problem for Prinz's view already touched upon is that what an emotion represents is either underdetermined by the calibration file, as it contains many perceptual and cognitive memories and associations that all function as a reliable cause of the

emotion (many of which may be activated at once), in which case emotions turn out to be about multiple intentional objects. Alternatively, an emotion might not be about the intentional objects of the thoughts contained in the calibration file, but rather solely about bodily states that the agent is aware of when in an emotional state. If the latter is the case, then it is not emotions that are about the joyous times and tearful tragedies of life, but rather the thoughts that cause emotions. Emotions, then, would always be about the disturbances in states of our body.

Alternatively, emotions might, on this view, be held to be at least partially *about* the calibration file and/or the judgments, memories, and beliefs that constitute that file. If this is in fact what Prinz endorses as a way of getting around the intentionality object, it would mean that the emotion is literally about a calibration file and not what that the mental states contained in that file are about, which is the actual intentional object of the emotion. To clarify this dilemma, consider again Prinz's example of sadness at the death of a child (Prinz 62). In this case, the emotion of sadness is not about the child dying at all, it is either about my bodily states or it is about my judgments and beliefs about the child dying that are calibrated to certain modes of awareness of physiological conditions, such that they reliably cause a bodily awareness of physiological states that are typical of sadness.

Neither alternative is very plausible. If I am sad about the death of a child, I am not sad about a judgment or a calibration of that judgment to a kind of physiological awareness. I am also not sad about what causes or reliably causes my physiological changes, at least when the intentional object and the cause of an emotion differ. That is, if alcohol, a neurological condition, or a conversations about death cause me to be sad

about the death of a child, my emotion is not about these causes, it is about the loss of a child. Finally, I am not sad about any particular physiological state I am in; I am simply sad *that the child is dead*. By contrast the APV account handles these kinds of cases well, because the awareness with physiological states is a background awareness of the value of the intentional object by definition, rather than an awareness of the physiological states themselves, their causes, or their calibration files.

Another aspect in which the APV view advanced here is a significant improvement over Prinz's view, is that while Prinz holds that certain differences in anatomy and behavior make it unlikely that there are identical kinds of emotions across mammalian species, his remarks in this regard are significantly undermined by empirical investigations into the behavior and neurology of mammals. Prinz claims that there are differences in the anatomy and expressive behaviors of adult humans and other animals that make it unlikely that they share emotions of the same kind. (Prinz 114-115) For instance, he states that "Infants and animals are emotionally affected by different things, and they express their emotions in different ways." (Prinz 115) If this were true, then we ought not to conclude that infants and adult humans share emotion kinds with animals of other species.

However, this claim is significantly undermined by data from current neuroscience and ethology. The anatomical differences he cites (such as a larger frontal lobe) are generally agreed to not be necessary for emotions, as "removal of the neocortex [and] localized brain-stimulation that evoke specific emotional behaviors, suggest that a series of emotional circuits exist […] for exploration, aggressive defense, fear, and various social initiatives […] in the limbic system [and] can be demonstrated at the

midbrain level." (Panksepp 74) The claim that there are differences in the frontal lobe between adult infants and animals of other species is irrelevant, as basic emotions operate through neurological circuits that are largely independent of these anatomical differences. The neocortex in humans is certainly involved in linguistic expressions of emotion, but of course linguistic abilities are not necessary for the feeling of instinctual fear or any other basic emotion.

Furthermore, although the specific mode of emotion behavior may differ from species to species, when the emotional behavior is characterized in terms of its function, there is enough similarity among at least different mammalian species to conclude that the same kind of emotion is motivating such behavior. As the neuroscientist LeDoux notes, while a fearful infant may crawl away from a threat and a fearful dolphin may swim away from a threat, they both act to escape the danger that their state of fear is about. (LeDoux 122) There are a variety of other ways of escaping or removing threats but these diverse behaviors and expressions of emotions serve the same function (somehow escaping from the threatening thing) and are initiated by the same midbrain located neurological circuits. (Panksepp 75) We can, therefore, dispense with Prinz's claim that animals and humans do not share basic emotions along with his embodied appraisal account of the structure of emotions.

I will have more to say about Prinz in my criticisms of his theory of morality in Chapter Four. For now it suffices that I have differentiated my view from Prinz's embodied appraisal account, and offered some reasons to prefer the APV account. There is something that can be saved from Prinz's view, however, which is the idea that bodily states express organism-environment relations that bear on well-being.

This conception of the values that emotions express Prinz credits to Lazuras, who holds that, "each emotion is defined by a unique and specifiable relational meaning […] expressed in a *core relational theme* […] which summarizes the personal harms and benefits residing in each person-environment relationship." (Lazarus 39) I think Lazarus is on the right track when he proposes that emotions are determined by what he refers to as "appraisals," which involve (rather than explicit linguistic judgment) constantly detecting and evaluating the relevant adaptational conditions of living that require action.[53] (Lazarus 191) The core relational themes of the four basic emotions that we are interested in here – anger, fear, sadness, and happiness – are shared representations of different kinds of values across at least all mammalian species.

I depart from the way that Lazarus and Prinz characterize core representational themes that are represented in emotions for two reasons. First, as they characterize these states, hunger, thirst, local pleasures and pains, coldness, hotness, and sickness all qualify as emotions because these non-emotional states also represent to the organism relations that bear on well-being. The APV account solves this problem of individuating emotions from these more purely physical states of the organism by positing that it is only when a physical state or change is interpreted by the organism as presenting a psychological value are we justified in saying that an organism has an emotion. This is why, as I will argue below, pain and pleasure do not qualify as emotions unless they are background contributions of the negative or positive psychological value even though, in general, they carry information about the organism-environment relation that bears on well-being. I will accordingly understand core relational themes as organism environment relations

[53] Although I will have some significant disagreements with his conception of which organism-environment relations are relevant to emotions.

that bear on psychological wellbeing, rather than the general physical well-being, of animals. Secondly, I disagree with Lazarus and Prinz's characterization of core relational themes when these themes are defined in a way that makes the basic emotions involve cognitive abilities that are uniquely human. I give reasons to think that these emotions should not be so defined below as I explain the four core relational themes for the basic emotions of happiness, sadness, fear, and anger.

Before I argue for the specific core relational themes that I think are represented by physiological and involuntary behavioral changes that function as vehicles of value in at least all mammals, it will be helpful to briefly consider the relation between pleasure, pain, moods and emotions. The standard distinction between moods and emotions is that moods are unlike emotions in that they do not have a specific intentional object, but like emotions in that they share a similar kind of feeling. The emotion of sadness, for instance, is about a specific loss, while depression is simply a constant and general tendency to feel sad. Similarly, when we are in an angry mood, we might find fault with everything or nothing in particular, but feeling the emotion of anger is standardly thought of as requiring a specific state of affairs (or agent) that one is angry about.

I think the most helpful, simple, and clear way to think of the relation between moods and emotions is to hold that the difference in intentional structure between emotions and moods is merely the varying of degrees of specificity and determinacy with which these emotional feelings target the world. When an emotion is completely defocused, as for instance when an organism feels fear of the general threatening scenario of hearing strange noises in total darkness, this instance of a mental state falling under the general category "emotional feeling" of fear is better described as the *mood* of anxiety.

When a general anxiety about this situation becomes fear as the lights are turned on and the source of the noises is identified as an approaching predator, this instance of a mental state falling under the same general category of an "emotional feeling" of fear becomes focused and is better described as feeling the *emotion* of *fear* about an imminent threat. The same can be said of the relation between the general moods of joy, depression, and rage, which are about general, yet-to-be-determined, or vague intentional objects. These moods become the emotional states of happiness, sadness, or anger as the emotional feeling is focused on a particular state of affairs that one feels, happy, sad, or angry about. I will use this classification schema of moods and emotions being different forms of emotional feelings (that vary only by the degree of specificity with which they represent entities and states of affairs in the world) in the following explanation of the core relational themes for anger, fear, sadness, and happiness and the corresponding moods of rage, anxiety, depression, and joy.

As we have seen, emotions cannot be reduced to pleasures and pains even when juxtaposed with thoughts that cause or co-occur with pleasures and pains. This is because pleasures and pains can occur without emotions and the same emotion can sometimes generate pleasure and sometimes generate pain, depending on context. For instance, anger and fear can be exhilarating or distinctly mentally painful and love sometimes causes the most exquisite internal pleasures and the most devastating mental anguish. Moreover, a pain in one's side and the pleasure of scratching an itch are not emotions; these kinds of states simply do not normally convey psychological values as background conditions of a total phenomenological state.

However, pleasures and pains can function in the background of a psychological state as indicating the psychological value of the intentional object, and when they function in this way the organism is in an emotional state. For instance, when one's painful stomach is interpreted as bearing a certain value about something other than one's physical body, such as the threatening value of an approaching predator, then one is in the emotional state of fear precisely because the stomach pain is internally represented as indicating something more than information about one's stomach.

I will stipulate that the negative contribution of pain or the positive contribution of pleasure as an indication of psychological well-being is to be referred to as "suffering" and "delight," respectively.[54] It is worth noting that one's psychological well-being cannot be considered good while an all-consuming pain occupies one's mind any more than one's psychological well-being can be considered to be poor when in the midst of an intense and preoccupying pleasurable state. These kinds of pleasures and pains are different from low level pleasures and pains, and rightfully considered to be manifestations of the emotional states of delight and suffering, respectively. In such

[54] DeGrazia has powerfully argued, that pain in the sense of "an [intrinsically] unpleasant or aversive sensory experience typically associated with actual or potential tissue damage," (107 and see footnote 26) is justifiably attributable to animals firstly because of the similarity of pain behaviors across at least mammalian species. Harrison points out that the human pain detection threshold, which is the smallest stimulus that humans experience as painful 50% of the time, seems to correspond to a similar threshold as evidenced by pain behaviors (escape or avoidance) in at least all vertebrates. Degrazia also points out that all vertebrates demonstrate a capacity for learning, discrimination, and behavioral evidence of pain. These behaviors include avoidance, escape, crying out, getting assistance, and limiting of the movement of the painful body part. There is also overwhelming physiological evidence for pain in at least vertebrates in that anesthesia and analgesia control pain behaviors in all vertebrates (and some invertebrates). 108-109 Following Degrazia, who holds that suffering is "*a highly unpleasant emotional state associated with more-than-minimal pain or distress,*[54]" I also hold that suffering is an emotional state. (116) However, I differ from his terminology regarding pleasure, joy, and happiness. While I understand pleasure as an intrinsically pleasant sensory experience, I use the term delight to indicate an emotional state associated with more-than-minimal pleasure. The nature of the associations are between pleasure and pain and their corresponding emotional states of suffering and delight are explained above, and in this again I differ substantially from what Degrazia has to say about the relation between pleasures, pains, and emotions. See in particular pg. 124-126 for his views on the positive feelings.

cases the pleasures and pains themselves become positive or negative value indicators functioning in of the total psychological state of being consumed by intense pleasure or pain. As such, while pains can be part of a positive emotional state (pangs of love, muscular exertion during joyful play) and pleasures can be part of a typically negative emotional state (exhilarating fear and anger), "suffering" is by definition a pain that indicates a psychologically harmful emotional state, and "delight" is by definition a pleasure that indicates a psychologically beneficial state. There is significant overlap, then, between suffering and its manifestation in states of fear, anger, or sadness, and between delight, and its manifestation in happiness. With the APV account of basic emotions now supplemented by the proposed conception of core relational themes, and the relation between emotions, pains, pleasures, and moods now well defined, let us now consider the specific core relational themes that determine the kinds of values that are represented across at least all mammalian species for at least the four basic emotions.

Consider again the basic emotion of fear. The wealth of empirical research indicating that the cardiovascular response, brain systems, chemical responses, and nervous system activity that occurs in response to threatening situations are surprisingly similar in mammals is best explained by fact that an awareness of bodily states (whether real or simulated by electrical shocks to the neurological circuits causally necessary for fear) as vehicles of value occurs when an animal feels fear.[55] The nerves extending to the gut, heart, blood vessels, and sweat and salvitory glands that give rise to a taut stomach, racing heart, high blood pressure, clammy hands and feet, and dry mouth are activated when a human or animal with a similar nervous system is in a state of fear, and we know

[55] Le Doux (1996) p. 132

that these cardiovascular stress responses are controlled by similar kinds of brain networks and body chemistry in species as different as birds, rats, rabbits, cats, dogs, monkeys baboons, and people (to name just a few of the better studied species).[56] There is also empirical evidence confirming the intuitive idea that the behavioral characteristics of emotional states, when abstractly and functionally defined, are found in many species other than humans. For instance, while fish swim away from danger, birds fly away from danger, and humans run from danger, all three are correctly characterized abstractly and functionally as fleeing a threat. Fleeing, freezing, and defending oneself are three examples of abstractly characterized fear behaviors that are undertaken by animals and humans alike. All of these behaviors have the function of escaping from the threatening circumstances that fear represents.

Bearing this in mind, consider that fear, for Lazarus, expresses the core relational theme of a concrete and sudden danger (although Lazarus prefers to call this fright in order to distinguish it from the more symbolic, existential, and ephemeral threat that is expressed in anxiety). (Lazarus 235) However, there is no reason to restrict our conception of fear to immediate threats, since we usually and commonsensically hold that fear can involve recognizing a danger that is not immediately present. The extent to which far off events can be abstractly conceptualized will of course limit the things that one can be afraid of, but the point is that fear represents to the organism the core relational theme of a danger (that is happening or is anticipated, imagined, or remembered) to the organism's unimpeded survival and flourishing, however spatiotemporally removed the intentional object of fear is for that organism. Following

[56] Le Doux (1996) p. 132

Lazarus let us call the corresponding mood anxiety, which is merely the defocused emotion of fear (fear about something yet to be determined). Anxiety, then, is just the representation of a yet to be determined, general, or vague danger to the organisms survival and flourishing.

Anger, for Lazarus, expresses the organism environment evaluation that an offense has been committed against one or the kin and conspecifics one cares about. Lazarus, who is trying to characterize adult human anger, adds to this core relational theme that the offense must be considered to be a demeaning slight or a threat to one's *identity* (presumably, identity is interpreted *very* broadly by Lazarus so as to include one's ideological and moral values). (Lazarus 222) I would argue, however, that even a physical injury, if unprovoked, can initiate anger in the absence of a perceived threat to one's identity. This, in any case, is certainly true for nonhuman mammals, and there are, as Panksepp has thoroughly argued for, underlying neurological circuits that are shared by all mammals for anger of this sort. (Panksepp, *Affective Neuroscience: The Foundations of Human and Animal Emotions* Chap 10) As Panksepp puts it "*Modern evidence suggests that anger emerges from the neurodynamics of subcortical circuits we share homologously with other mammals. The general localization of these circuits have been identified by localized electrical stimulation of the brain*." (ibid 187) Let us again revise Lazarus' conception, and posit that the core relational theme for anger is, according to the APV account, that a harm has either happened or is anticipated, imagined, or remembered, that prevents the organism from realizing a general goal of the organism. I understand "goal" here very broadly to include making movements that an

organism wants to make or living in a way that is desirable to the organism (especially with respect to living with healthy and safe offspring, kin, and conspecifics).

This way of describing the core relational theme of anger captures that anger is invoked when harm has been done to something that the organism cares about. The harm is generally or typically represented as a surmountable frustration that can be prevented or alleviated by the threat of an aggressive attack or the actual initiation of such an attack. We can refine this core relational theme as the notion that anger represents to the organism that a goal has been frustrated in way that can be alleviated by (at least) initiating an aggressive attack. When the frustration is underdetermined, the organism can be described as enraged, where the mood of rage is merely the defocused emotion of anger and the emotion of anger is understood to be equivalent to feeling a focused rage about a certain state of affairs. When an organism appropriately feels anger there is an obstacle (to the ego, the circle of care, or the general goals of the organism) that is experienced through the bodily changes contribution to phenomenology as frustrating, and the behaviors of attacking (biting, clawing, pouncing), destroying, or aggressively overcoming the surmountable obstacle are functionally similar across species. Again, this is true at the functional level even though the specific way of attacking, destroying, and aggressively overcoming will change depending on the environment, social context, physiology, ontogeny, and phylogeny of the organism.

Sadness expresses the core relational theme of an important loss. While Lazuarus adds the qualifier "irrevocable" as being necessary for sadness, this would restrict our concept of sadness to an overly narrow kind of adult reflective sadness. We can abstract from this qualifier here, except to mention that the reaction of sadness generally

expresses the helpless status of an organism to recover this loss. This emotion also has homologous underlying neural circuits in all mammals (see Panksepp, *Affective Neuroscience: The Foundations of Human and Animal Emotions*, Chap 14). The emotional behaviors of sadness are somewhat unique, as they are typically identified as the lack of normal feeding, drinking, exploring, mating, playing, and other social behaviors. Sad behaviors are those that have the function of expressing an emotional understanding, through the contribution of bodily states to the phenomenological experience of the intentional object, of the value of the entity that is lost or separated from the organism (by death, destruction, or insurmountable barrier or distance). It is noteworthy that the functional description of sad behaviors does not reveal an obvious evolutionary advantage for the organism, since mourning behaviors are often pointless from an egoistic perspective and involve subdued performance (or an altogether lack of performance) of activities that are essential to survival such as eating and drinking. However, the anticipation of the suffering accompanying such behaviors is a powerful motivator to avoid these losses of companions and entities useful for survival. It is also likely that mourning behaviors and subdued normal activities are advantageous to the organism in so far as they allow for the organism to reset and reorganize its values and goals, many of which likely revolved around the lost and mourned entity. When the noticeable lack of normal behaviors, and the downcast posture of sad animals as they withdraw from their environment and their usual activities, are a constant tendency for an organism we have evidence of the corresponding mood of depression, which is minimally characterized as a constant feeling of sadness that has a general, vague, or yet-to-be-determined intentional object.

Finally, the emotion of happiness expresses, according to Lazarus, that we have gained something or are gaining something we desire (Lazarus, 267). Here again I disagree with Lazarus, since during animal or infant play (the archetypical activity associated with feeling joy or happiness) there need be no egoistic thoughts of gaining or having gained a desirable state of affairs. Still, it would be odd to claim that frolicking and playing animals are not happy. Panksepp, who has researched the neurological circuits underlying the joy or happiness expressed by mammals when playing, argues that neurological evidence suggests that there are homologous neurological circuits that function to initiate play and arouse joy during play exist across mammalian species. (Panksepp, *Affective Neuroscience: The Foundations of Human and Animal Emotions* 280-281) Happy behaviors are those that have the function of drawing the organism to the desired goal or to perform the desired activity (and motivating via anticipatory excitement), whether this is by playing (in any of the extremely various modes of play exhibited by animals), copulating, grooming, or being reunited or near kin and companions. The core relational theme for happiness then seems to be rather that the situation one is in with respect to one's environment is good for one's flourishing and *usually* this situation is pleasant and desirable for the organism. I will use the term joy, somewhat stipulatively, to refer to the mood of happiness about something underdetermined, vague, or yet-to-be-determined, and the term happiness to refer to the occurrent emotion (and not eudaimonism) that represents a specific aspect of one's situation as valuable to flourishing and *typically* pleasant and desirable.

The qualifier "typically" is necessary here since it is possible that pleasure and preference-satisfaction can happen without an emotional state of happiness. That is, it

seems possible that one can be in a situation that is pleasant and desirable and yet feel neither occurent happiness nor joy. The core relational theme for happiness also involves a rather vague conception of "flourishing," which is notoriously difficult to account for even in the human case. We can provisionally characterize flourishing for an emotional animal as having a full life that includes:

1. Having the autonomy to explore a diverse environment that allows for accomplishing challenging and rewarding organism specific goals,
2. having a good upbringing with caring family relations,
3. having the ability to play with others and aspects of the environment,
4. forging appropriate relationships with companions and family,
5. being able to engaging in species typical behaviors to satisfy species typical desires,
6. having a comfortable and safe home environment,
7. being as free as possible from suffering (pain, fear, anxiety, anger, rage, sadness, and depression),
8. and having normally functioning bodily capacities and health.

The occurrent (typically pleasant or desirable) happy or joyous state is part of the organism's overall flourishing in that it motivates an organism to perform these activities and draws them to these environmental situations by expressing the value of these activities and situations through the phenomenological contribution of various physiological states of the organism.

To summarize, these are the core relational themes that determine the values represented by the four basic emotions (and their corresponding moods) shared by at least all mammalian species:

1. Fear: A danger (that is happening or is anticipated, imagined, or remembered) to the organism's unimpeded survival and flourishing

2. Anger: A frustrating harm (that is happened or is anticipated, imagined, or remembered) that prevents the organism from realizing a general goal, including making movements that an organism wants to make or living in a way that is desirable to the organism (especially with respect to living with healthy and safe offspring, kin, and conspecifics)

3. Sadness: An important loss (that is happened or is anticipated, imagined, or remembered) that the organism is helpless to recover

4. Happiness: The situation (that the organism is in or that is anticipated, imagined, or remembered) is one that is part of the organism's overall flourishing and *typically* pleasant and desirable for the organism.

The core relational themes are expressed in an emotion as what Lazarus calls "appraisals." Since the way that emotions internally represent values is important to determining whether animals have emotional states that can be morally motiving, it is worthwhile to explain what appraisals are for Lazarus, and why I hold that they can, when appropriately defined, occur through a background awareness of physiological states as vehicles of value.

Lazarus maintains that at the conscious level appraisals can, but do not have to, happen through adult deliberate, conscious, and reflective thought under volitional control. Importantly for the compatibility of appraisals and the APV account, there is also another form that appraisal can take, which is through simple, primitive, automatic, unconscious or preconscious, and rapid emotion processing systems. (Lazarus 128) That is, some appraisals expressed in emotions occur in the automatic and uncontrollable, and, Lazarus adds, "unconscious," modes. However, it is a mistake to consider emotions unconscious because if emotions are by nature value revealing or value transmitting modes of awareness, then even though the appraisal's *generation* might be unconscious

its *expression* must become conscious in order to mean something to the organism and motivate the organism to act in the relevant ways. In any case, these unconscious modes of appraisal are, for Lazarus, similar to what Heidegger refers to as "being-in-the-world," Merleau-Ponty refers to as emotional intelligence, and more modern psychologists have dubbed tacit knowledge (Polanyi 1966) and affordances (Baron and Boudreau 1987) (Lazarus 152-3).[57] It is this form of "appraisals" that is most plausibly attributed to both humans and animals in virtue of a background awareness of physiological states as vehicles of value.

Lazarus suggest that this kind of appraisal contains meaningful values in a way analogous to the way that we can ride a bicycle or type without explicitly representing to ourselves the position of the keyboard or how to balance on a bicycle. We have tacit knowledge of these things that allows us to unreflectively type or ride. Just as the meaning of the keys position in the keyboard is known pre-reflectively (after explicit training) and yet allows for action guidance, and the balance adjustments we make when riding are done without deliberate effort, so too emotions convey meaningful appraisals (without explicit and deliberate reflective effort) to an organism. Importantly, these bodily contributions also make these activities feel fundamentally different from other activities. For instance, walking feels different from riding a bike just as typing feels different from speaking, although the difference does not seem to involve explicit awareness of the fine details of the muscular movements that are occurring without reflective thought. This corresponds to the way that bodily states contribute to the

[57] It is doubtful that there is, as Lazarus suggests, such a simple and straightforward interpretation of Heidegger's conception of "being-in-the-world," which Heidegger explains in its relation to the complex concept of Dasein (a being that must be understood in terms of the possible ways for it to be, including the state of being-in-the-world). See Heidegger's *Being and Time*, pgs. 42, 63-5. As an interpretative foray into Heideggerian philosophy will take us too far afield, I will merely note this complication.

phenomenology as a value-conveying background to the intentional object. Fright feels different from anger, even though the evaluation that each transmits can occur effortlessly and reflexively through our bodily changes and our awareness of these changes as value indicators.[58]

According to the APV account, the core relational themes for the basic emotions of sadness, happiness, anger, and fear are values that the organism is aware of in virtue the contribution of their physiological state to the background of the intentional object. The reason that emotions are motivating, phenomenologically distinct from judgments (with same content), preverbal and meta-verbal (they can be had without language and cannot be exhaustively explained in language without referring to the physiological feelings), and have homologous physiological and neurological substrates across mammalian species, is that the core relational theme is conveyed to the organism through

[58] Much more needs to be said on the representational status of these pre-intentional activities and how they contribute to and interact with conscious emotional states. Rowlands has argued that pre-intentional activities or "deeds," are best conceived of as an "array of on-line, feedback-modulated adjustments that take place below the level of intention but, collectively, promote the satisfaction of [an] antecedent intention." Deeds are, if they meet certain criteria, representational. (*Body Language Representation in Action* (2006) 103) I think that the involuntary physiological and behavioral changes that function as vehicles of representation in an emotion have a similar representational status. These changes have the purpose of tracking (and the ability to fail in doing so) organism-environment relations important to survival and psychological wellbeing. They also bear important information about these relations to the organism through their contribution to phenomenology, and they are subject to modification by individual (ontogenetic) trial and error learning, and therefore not dependent for their character on the immediate environment. Physiological and involuntary behavioral patterns that contribute value to phenomenology are different from "deeds" in so far as they do not *require* (although they are often the product of) an antecedent intention, but because of the previous qualities of these physiological and behavioral vehicles of value, they meet the requirements Rowlands has argued are necessary for being representational in the sense of playing a role in representation that is itself representational. Physiological vehicles of emotional values are able to misrepresent (and so meet the misrepresentation requirement Rowlands argues for on pgs. 218-220), and they are decoupleable from the immediate environment when they are subject to modification by trial and error learning and error identification when emotional vehicles fail to detect the themes they have the initially phylogenetic and then quite quickly ontogenetic purpose of detecting (and so meeting the teleological and decoupleability requirements Rowlands argues for on pgs. 215-218). These vehicles also bear the information about the threat or benefit of the intentional object to the organism (and so meeting the informational constraint) in a way that, if all goes well in the process, mirrors the relevant relational properties the intentional object has to the organism (and so satisfies the combinatorial requirement Rowlands argues for on pgs. 222-223). These pre-purposeful but arguably representational emotional reactions of the body give us value-laden information that help shape our purposes.

their physiological states. The APV account holds that the awareness of these bodily states is not an indirect awareness of these states *as bodily states* (as Prinz maintains), but rather that one is focally aware of an intentional object as have a certain value that is indicated by a peripheral or background awareness of physiological states contributing *as vehicles of value*.

I have defended the APV model not as an account of all emotions, but as an account of the way that a subset of similar and usefully compared human and animal emotions are structured. If the proposed criterion of emotional ascription and the view of basic emotions I am suggesting is correct, then not only is there convincing (though necessarily incomplete) neurological, evolutionary, and behavioral evidence for the presence of similar and comparable kinds of emotions in animals and humans; there are also good philosophical reasons for endorsing formal similarities in the intentional structure and value-laden content of these basic emotional states.

Importantly, the possibility that animals not only have basic emotions but also have more complex emotional states approximating or corresponding to some forms of human compassion and benevolence is also suggested and explained by this model as instances where a basic emotion or feeling is modulated by the psychological mechanism of empathy. For instance, compassion can be thought of as merely "*being moved by another's suffering and wanting to help.*" (Lazarus 288) If we think of compassion in this way, it is natural to regard the mechanism of compassion to be empathic suffering. It is through empathic suffering that the organism represents the negative value of the suffering of another and is thereby motivated to alleviate or resolve that suffering. The nature of this suffering may be an empathically experienced negative emotion, such as

fear or anxiety, or an empathically experienced physical pain that is interpreted by the organism as representing the negative psychological value of the pain that is occurring to the target of empathy.

The plausibility of animals being justifiably ascribed not only basic emotions but also more complex emotions like compassion that are comparable to human emotional states does not, of course, depend on the plausibility of this particular model for animal and human basic emotions. Rather I suggest this model as one way of illustrating how some kinds of complex emotions are justifiably ascribed to both animals and humans. A much stronger argument would be needed to show that the constitutive structure of these basic emotions, as I have modeled them, is essential to all emotional states, including the more cognitively sophisticated human emotions. I suspect this larger project is defensible, but my aim here is more humble.

The overall project of this dissertation is to argue that some emotions that animals can be justifiably ascribed are morally motivating. Two natural candidates for moral emotions are empathically feeling or fearfully anticipating another's suffering and being moved to alleviate or prevent it (which Lazarus refers to as "compassion"), and enjoying or desiring another's pleasure or delight. Taking a cue from Lazarus, let us stipulatively call this latter state "benevolence," where the pleasure or delight of another is empathically represented to the organism as conveying a positive psychological value they are motivated to promote. If compassion and benevolence are not just possible but likely occur in dispositional form in all mammals, they constitute prima facie evidence for moral motivation. However, in order for such states to be morally motivating, they must be more than mere reflexive emotions. Rather, they must be learned and

developmentally flexible responses to the internal representation of the wellbeing of another through empathic emotions. Accordingly, I will now turn to the neurological and behavioral evidence that has been discovered for the presence of learned and developmentally flexible empathically modulated emotional states. In what follows, I will explain and carefully attempt to interpret the philosophical import of this evidence; relying mainly on the recent work of Preston and De Waal, who have investigated and argued for the presence of empathy in all mammals.

Part Three: The Development and Elaboration of Basic Emotions into Learned and Empathically Modulated Emotional States

In this section I examine and interpret the philosophical significance of the neurological, evolutionary, and ethological evidence for the development and elaboration of basic emotions into more complex empathic emotional states in animals. The developmental conception of empathic emotions I will defend is intended to ensure that when we ascribe such emotions to animals we are ascribing the same kind of learned caring response to the well-being of others that we would be if we were to ascribe empathic emotions to humans. This will in turn advance the thesis that animals are morally motivated in so far as this kind of response can be correctly thought of as morally motivating in humans in the absence of an explicit understanding of moral rules or standards.

As our starting point, consider the persuasive arguments of De Sousa in '*The Rationality of Emotions*" for the thesis that we ought to regard emotions in general as

more than just rigid and inflexible subjective responses to stimuli that are preprogrammed by an organism's genes. Instead, he suggests that emotions apprehend something in the world that exists independently of our reactions. In particular, De Sousa argues that emotions constitute the apprehension of axiological properties. (De Sousa xiii) The axiological properties that are apprehended first originate as biological values and then develop as an organism learns certain paradigmatic situations through various social and individual learning processes that appropriately elicit a certain emotional response. In the case of animals rather than the dramatic situation types of humans that De Sousa has in mind, these paradigmatic emotional response eliciting encounters will be situations such as predator threat responses, friendly playing situations, aggressive challenges/responses, caring for offspring situations, and so on. For humans, De Sousa designates these social and environmental situations that a paradigmatic emotional response is appropriate for to be "paradigm scenarios." Paradigm scenarios are learned dramatic situation types that define the reaction, roles, and feelings characteristic of that emotion. In what follows, I will adopt the developmental schema De Sousa suggests, where basic emotions are modified by learning appropriate responses to paradigm scenarios, which in turn give emotions their more complex meanings.

The notion of paradigm scenarios is useful to the APV account in so far as it gives us a useful framework to understand how an initially instinctive emotion like fear can come to be a response to something learned, such as a particular kind of alarm cry. In this section I review the evidence for the increasing sophistication and complexity of animal emotions as those animals develop their cognitive and empathic capacities and learn appropriate reactions to paradigm scenarios by interacting with their environment,

particularly with respect to their social interactions with kin and conspecifics. These interactions are initially primed by genetically inherited neurological mechanisms to detect and respond to environmental features useful to evolutionary fitness, and then develop to reflect increasingly sophisticated and discriminating appropriate response to paradigm scenarios arising out of interactions with conspecifics and the environment.

I submit that an animal's learning of appropriate responses to *social* paradigm scenarios occurs as an organism's basic emotions interact with its capacity to feel empathy, or thoughts and feelings more appropriate to another's situation, as well as its more purely cognitive capacities to remember, associate, anticipate, and categorize various social and environmental features that bear on psychological well-being. By doing so the organism develops emotions that have increasing sophistication with respect to discriminating both the nature of the value that is expressed to the organism and the nature of the intentional object. For instance, when vervets learn through trial and error and social reinforcement (including other vervets encouraging the appropriate responses to fearful paradigm scenarios and discouraging or ignoring inappropriate responses) to respond with greater fear to a senior vervets alarm call than a juvenile's alarm call, and to fear an aerial predator rather than a ground predator depending on the sound made by the senior vervet, their emotional state of fear has been changed developmentally by social and individual learning. It has been altered from a mere reflexive response primed by genetic and neurological properties to an appropriate response to this paradigm scenario through their capacity to remember previous scenarios, expect various things from an alarm cry from a senior vervet, distinguish between senior and juvenile alarm cries, categorize kinds of predators and sounds, and associate one category of sounds with one

category of predators. This process and the developmental nature of emotional reactions generalizes to at least all mammals to the extent that they have the normal capacities for basic emotions and cognitive capacities for memory, expectation, categorization, and association. When mammals have the environmental opportunities to develop their natural emotional and cognitive capacities, they are able to develop more complex, meaningful, and discriminating emotional responses. Yet this does not exhaust the way that emotions are developed over time through social and environmental interaction. It is more germane to our purposes to take stock of the evidence for the developmental nature of empathic emotions, which are natural candidates for moral motivation in animals.

There is convergent evidence for the presence of empathy in at least all mammals; where empathy is understood broadly as having feelings, emotions, or thoughts more appropriate to another's perspective or situation. It is well established that organisms as diverse as rats, mice, elephants, chimpanzees, dogs, monkeys, and humans respond with empathic emotions to the perceived suffering of another in a way that motivates behavior that will alleviate, prevent, or lessen that suffering. (Preston and De Waal 3, Bekoff and Peirce 90-109) Preston and De Waal have recently attempted to explain this behavioral data with an explanation of the capacity of empathy that attempts to reconcile and integrate various emotional, cognitive, and conditioning views of empathy that have been proposed in the literature.

Their model, which they call the Perception-Action Model (PAM), is intended as an explanation for the myriad of ethological observations of empathic behaviors and experiments that have been done on animals of different species that exhibit empathic

behavior. In their extensively peer reviewed paper "Empathy: Its Ultimate and Proximate Bases," Preston and De Waal first review the studies have been performed on rats, monkeys, elephants, apes, and humans that indicate that empathy is a phenomenon that is widespread in the animal kingdom. Interestingly for our purposes, these studies also indicate that the strength of the empathic response increases developmentally for at least mammalian species. To see this, consider the following brief survey of the relatively sparse empirical literature on animal empathy, beginning first with Preston and De Waal's summary of the study of empathy in albino rats.

When an albino rat sees a conspecific distressed because he is suspended in the air by a harness, the rat will press a bar to lower the other rat to the ground (Rice and Gainer 1962 in Preston and De Waal pg 1). Rutte and Toborsky have also performed a study that suggested that rats that are helped by an unfamiliar rat are more likely to provide help to an unfamiliar and unrelated individual. (in Bekoff and Pierce 55) Similarly, Warneken and Hare have shown that chimpanzees will help humans even if there is no reward. (Bekoff and Peirce 75) Wechkin, Masserman, and Terris performed a study on rhesus monkeys indicating that they will not take food if doing so subjects another monkey to an electric shock. (Bekoff and Peirce 98-99) Finally, Church has also conducted a similar study on rats with similar results – rats also refused to eat if it meant another rat would be shocked. (Bekoff and Peirce 96)

There are countless anecdotes of elephants and chimpanzees modifying their behavior to help elderly or disabled members of their group move around or obtain food (Bekoff and Ian Douglas-Hamilton 97-98, 102-104). For instance, elephants will gently touch the injured part of another elephant with their trunk before helping them to move

around and acquire food, (Bekoff and Peirce 103) and chimpanzees will stay especially near and help sick or disabled members of their group and help them in idiosyncratic ways specific to the nature of the sickness or disability. (De Waal 1996, 51, 59, 80) It is also well established that both elephants and monkeys, when deprived of normal early mother-infant experiences where there is constant emotional feedback and close contact between mother and infant, have a decreased capacity for empathy and an increased capacity for violent, neurotic, or antisocial behavior (Bradshaw 106) and (De Waal 1996, 178-79). Interestingly, Novak has shown that monkeys who develop neurotic tics and are terrified of bodily contact due to their being raised in isolation can be rehabilitated simply by bringing them into contact with normal members of their species so that they learn that bodily contact and interaction with conspecifics can be desirable rather than frightening. (De Waal 1996, 178) Similar effects can be observed in deprived or abused human children and domestic pets, who likewise show an obvious increased propensity to antisocial, neurotic, and violent behaviors and a decreased capacity for empathy as a result of being deprived of an early normal caring relationship of constant empathic feedback from a caregiver.

Let me conclude this quick review of the animal altruism literature with the notable De Waal study involving normal stump tailed macaques who were given an opportunity to act as social tutors for rhesus monkeys. (De Waal 1996 178-180) This study indicated that (some) animals can actually learn to be more kind and empathic to others just by being around a species that exhibits more frequent kind and empathic behaviors. In this study a group of slightly older, easy going, and tolerant stump-tails (who statistically display reassurance and reconciliation behaviors more often than rhesus

monkeys) were mixed with a group of younger rhesus monkeys (who have a strict hierarchy and display three times less reconciliation behaviors after fights). During the study the rhesus monkeys actually permanently modified their species typical behavior to the same frequency of kind behaviors that their older more pacifistic tutors typically display. As a result, the rhesus group came to develop kinder and more peaceful dispositions than is typical of their species. By the end of the study the rhesus monkey's reassured each other and reconciled with each other with just as much frequency as the usually more peaceful Stump-tails.

Interestingly, this result is not explainable by mere imitation, because the rhesus did not adopt any of the behavioral patterns typical of stump-tails except for the friendly and reconciliatory behaviors. They did not, for instance, pick up the hold-bottom and teeth chattering behaviors of the stump-tails, but instead behaved exactly like typical rhesus monkeys except for a more pacifistic and friendly dispositions they learned from their tutors. (De Waal 178-180) Together these studies indicate that, just as with humans, animals have a capacity to feel empathy that becomes developed by others modeling empathic behavior, particularly the initial caregiver of the offspring. When an organism acts to help, make peace with, or reassure another on the basis of this capacity, it is parsimonious to conclude that they internally represent the welfare of the other through their empathic emotion, and help with the genuinely altruistic motive of alleviating or preventing the suffering of another. These studies also indicate that the strength and appropriateness of the empathic response of animals increases with increasing familiarity (previous experience with the object), similarity (perceived overlap between subject and object such as species, personality age, and gender), learning (implicit or explicit

teaching), past experience (with the situation of emotional arousal for the object), and salience (strength of perceptual signal). (Preston and De Waal 3)

Preston and De Waal maintain that they can identify and explain consistencies of empathic mechanisms across at least all mammalian species with the PAM explanation because "all empathic processes rely on a general perception-action design of the nervous system that has been postulated for over a century, is adaptive for myriad reasons, and exists across species." (ibid 2) According to this model, the proximate[59] mechanism of empathy is the process of the perception of an organism's emotional state that activates the empathic agent's corresponding representations of that emotional state. These representations in turn activate the somatic and autonomic responses in way that explains the basic behaviors of alarm, social facilitation, vicariousness of emotions, mother-infant responsiveness, and the modeling of competitors and predators; all of which are crucial for the evolutionary success of animals living in groups. (Preston and De Waal 1)

According to Preston and De Waal, the term empathy, because it is used to refer to a wide variety of cognitive and emotional processes, is best understood broadly as "*any process where the attended perception of the object's state generates a state in the subject that is more applicable to the object's state or situation than to the subject's own prior state or situation."* (Preston and De Wall 4) All forms of empathy involve a minimum level of emotional contagion and personal distress, they maintain, and can best be explained as occurring when:

---

[59] The difference between an ultimate and proximate cause, created by Earnest Mayr, is that "proximate causes govern the responses of the individual (and his organs) to immediate factors of the environment while ultimate causes are responsible for the evolution of the particular DNA code of information with which every individual of every species is endowed" (Mayr 1961, p. 1503 in Preston and De Waal pg. 2) Accordingly, I focus on the proximate causes of empathic behavior, the evidence for these causes, and the contribution of PAM to understanding these phenomena here.

> *attended perception of the object's state automatically activates the subject's representations of the state, situation, and object, and that activation of these representations automatically primes or generates the associated autonomic and somatic responses, unless inhibited.* (Preston and De Waal 4)

While restricting feelings of empathy to only those which involve personal distress is arbitrary, since for instance during play, grooming, and sex empathic transfer of positive emotions involves quite the opposite of personal distress, the PAM explanation can be adjusted to include such emotional states without losing its advantage of being able to explain how empathy shares with imitation the structure of automatic and spontaneous shared representation allowing for the identification with others based on physical similarity, shared experience, and social closeness. This model is also entirely compatible with the APV model of emotions I have advanced in this chapter; as the automatic and spontaneous shared representation causes a distinct bodily contribution of associated somatic response that in turn allow the organism to internally represent the value of the well-being of the target of empathy. The PAM model also explains how empathy and imitation have a hard-wired socio-affective basis in the same neural mechanism that motivates behavioral outcomes and allows for the development of more complex forms of empathy; which begins with emotional contagion, proceeds to sympathetic concern, and finally (in at least well-developed empathic primates) to perspective taking and targeted helping behaviors. (De Waal 135) Finally, the advantage of understanding empathy with the PAM explanation at the ultimate level is that this kind of explanation emphasizes that in group-living species friends, relatives, and conspecifics are organisms whose emotional states often requires a particular appropriate response. (Preston and De Waal 6) A highly adaptive nervous system organization that is preserved across species and that "responds automatically with empathy" to appropriate situations

has the added benefit of maximizing inclusive fitness and creating the appearance of reciprocal altruism. Group living animals need to respond to other individuals with a matching response, hence a response oriented nervous system that changed with ontogenetic effects such as familiarity, similarity, and shared experience is evolutionarily beneficial for such a species. (Preston and De Waal 6)

It is vitally important for our purposes to note that perception-action processes (that are preserved across species) facilitate the mother-offspring bond; where continuous emotional and physical contact between mother and infant organize the emotion regulating abilities of the infant and allow for empathic emotional competence later in life for that infant. Empathic mechanisms also allow for and guide the mother's behavior so that infant survives by eliciting appropriate caring behavior primarily through the empathic emotional responses of the mother to the infant. (Preston and De Waal 7-9) These mother-infant empathic behaviors have a profound effect on social competence and reproductive success and are plausibly regarded as the very first developmental emotional responses to paradigm scenarios that involve empathically understanding (via perception action processes) the emotional states of others.

To support this claim, recall that Panksepp has persuasively argued that there are distinct neurological circuits governing caring emotions and behaviors between mother and infant across all mammalian species. According to Panksepp, these underlying neurological structures, which he refers to as the CARE circuit, motivates maternal nurturing behavior in all mammals. (Panskepp 2004, pgs. 246-261) The presence of empathy between mother and infant and its vital role in emotional development is therefore evidentially supported in that is instantiated in a distinct neurological circuit

found in at least all mammalian species. Empathically modulated basic emotions and empathically felt pleasure and pain are a crucial part of a mammalian infant's first and perhaps most developmentally important social interaction with its mother. As has been noted by Darwin, McDougall, Plutchik, Harlow, and De Waal and Preston, among others, "The parent child relationship both relies upon and is necessary to develop the ability of individuals to be affected by the emotional state of others." (Preston and De Waal 7) The mother must give tailored care to the infant that relies upon a degree of emotional contagion, and the infant's emotional regulation abilities are coordinated by continuous emotional and physical contact between mother and infant. The capacity of the mother to respond to the distress of the infant; distress that they themselves feel via empathic mechanisms and that motivates them to act alleviate the source of the distress, is also the capacity for compassion as we have been understanding this emotion.

Recall that compassion is, as Lazarus suggests, merely the emotion of personal distress at another's suffering. The core relational theme is "*being moved by another's suffering and wanting to help"* by somehow reducing, eliminating, or alleviating that suffering. (Lazarus 288-289) As such, compassion does not require second order thoughts or linguistic judgments, but rather merely the capacity for personal distress at the suffering of another. The nature of this suffering may either be a pain or a negative emotional state that the object of empathy has, and the personal distress is the result of empathically feeling this pain or emotion. When this emotion also moves the organism to help (as in the case of a mother comforting, protecting, or feeding their infant) we are also justified from a behavioral perspective in calling this emotion compassion. The mother or caregiver's compassion involves the development of empathy, anticipation (of

the actions that will alleviate/prevent the suffering and of future suffering that can be prevented), and implicit and explicit memory of care-related paradigm scenarios the mother themselves engaged in when she was an infant (and that she has observed other mothers engaging in). All of these emotional and cognitive abilities interact to enable the mother or caregiver's capacity for empathic emotions like compassion for their infant. In this way that the mother is able to empathically feel the infant's distress, fear, or sadness and appropriately act to alleviate or prevent that suffering in a way that in turn develops the infant's capacity for empathic emotions and emotional regulation. The important philosophical point to be garnered from this empirical literature is that empathic emotions like compassion are developing and learned responses originating in the caregiver-infant relationships that increase with sophistication as the infant learns appropriate responses to social paradigm scenarios through empathic interactions with their mother, kin, and conspecifics.

Mothers (and other caregivers) typically display learned care behavior while the same developmentally alterable neurological circuits that underlie human caring maternal behavior are activated. This is strong evidence that animals not only have the same kinds of basic emotions as human beings, but also that they have compassionate emotions in virtue of having learned and empathically modulated aversions to the suffering of others. If compassion in this sense can be said to be morally motivating, than animals have morality. The way in which animals can be compassionate is not as cognitively sophisticated and discriminating as the compassion typically felt by adult humans, who can act not only from directly motivating moral emotions but also from explicitly learned and internally represented moral rules involving abstract reasons. Yet it is worth pointing

out that moral rules and abstract reasons may only have motivational pull, in the sense that they are things we care about and are motivated by, because of their ontogenetic beginning in the more primitive moral motivation of compassionate caring we share with all mammals.

Similarly, if we understand benevolence somewhat stipulatively as merely being moved by a loved one's actual or potential enjoyment and wanting to arouse or preserve such enjoyment, we are justified in ascribing a further potentially morally motivating emotion to animals. Consider the neurological evidence that during play and grooming activities, the underlying neurological circuits that Panksepp calls PLAY circuits involve the activation of brain opioid systems and empathic coordination in all mammals. (Panksepp 255) I think we are justified in calling the pleasant emotional state animals and humans experience during such activities benevolent, if we are careful to qualify our use of this term. By the term "benevolent," I do not mean universal love and the promotion of everyone's wellbeing equally as a result of that universal love, but rather only that the organism is moved to promote another's happiness. When close kin or companions play with or groom each other, and undergo the same physiological and neurological changes that underlie occurrent human happiness; this indicates an empathic appreciation of (and desire for) the other's happiness, joy, and pleasure when playing or grooming. It seems safe to say that empathic delight (including empathic joy, happiness, and pleasure) is just as possible and widespread in the mammalian kingdom as empathically modulated suffering (including empathic anxiety, fear and pain).

Radical changes in physiology and behavior, which are initially involuntary, preprogrammed evolutionary solutions to threats or benefits for the organism that do not

need to be learned, but quickly become part of more complex flexible, learned, and voluntary emotion states including empathic states, have an obvious impact on the organism's experience. The APV model explains the relation between the fine-grained empirical data for a set of common neurological circuits and emotional behaviors, and a common neurological capacity for empathy and empathic behaviors, and the question of whether an animal has a conscious empathic emotional state. A conscious empathic emotional state is the phenomenological registering of physiological states and behaviors as expressions of the value of another's well-being. When all three of the nonverbal ascription conditions are triangulated and the organism responds to environmental threats and benefits in a way that indicates a flexible and relatively perception independent understanding of the core relational theme, then we have sufficient evidence to conclude that an organism, whether human or animal, is in a basic or empathic emotional state of a particular kind. Which state they are in depends on the nature of the value represented by the physiological states contribution to phenomenology. When the emotional state is empathically modulated, the altruistic value of that the emotion is represented by the physiological states contribution to phenomenology through certain perception-action mechanisms whose activation is evidence of empathy.

I want to again draw attention to the emotional kinds that I am designating, somewhat stipulatively at this point, "benevolence" and "compassion." In later chapters, I will argue that when animals have emotions like these, as explained and justified by the APV model I have proposed, they also have morality, in a sense to specified. Yet at this point my arguments have merely established that animals should be ascribed at least a

subset of the emotions that we ascribe to human beings, and clarified the nature of those emotions and the evidence supporting such ascriptions.

A further issue, one I will address in my conclusion, is the *extent* to which animals should be considered moral beings in virtue of their empathic capacities. That is, while the arguments I will offer in this dissertation establish that animals are correctly ascribed moral motivation, there is a further level of moral understanding that some animals might have in rudimentary form. This is an understanding of morality itself, which I will argue can be had in virtue of empathically approving of another's empathic actions. I will suggest this understanding might indeed be possible for some animals in my conclusion, and indicate what studies have been done (and might be done) to indicate that this might be the case in at least chimpanzees, but for now I will remain neutral on the extent to which animals can understand moral norms and instead focus on establishing that, even lacking such an understanding, they are capable of moral motivation by internally representing and being motivated to promote the well-being of others (and alleviate the suffering of others).

I have advanced the developmental conception of APV account because I am convinced that it is the best explanation of the nature of basic and empathic emotions given the best available empirical information and the philosophical deficiencies of the standard reductionist accounts. Yet the key to understanding how animals have morality lies in explaining the relation between empathic emotions and morality, and this can be done somewhat independently of defending a particular account of emotions. As long as emotions express values and can be empathically motivated, and their underlying structure is such that neither of these capacities depends on complex linguistic,

abstracting, or introspecting abilities, then the relation between empathic emotions and morality can investigated independently of whether APV is ultimately the correct account of basic and empathic emotions. Accordingly, I will not assume that this account is correct in what follows. However, I will continue to endorse it as one possible and plausible explanation for how values are expressed in emotions and how animals come to have such emotions, and I will continue to argue for its advantages in the following explanation of the reasons we have for ascribing moral motivation to animals. Now that we have an explanation of the nature of the subset of the emotions that we can legitimately ascribe to animals and human beings, and clarified the evidence supporting such ascriptions, it remains to determine the relationship between emotions and morality. It is to this task I turn in the next chapter, which begins by critiquing a view of moral emotions that makes it highly unlikely that animals can be ascribed moral motivations.

Chapter 2: Strong Cognitivism and Moral Emotions in Animals

Summary

In this chapter I will criticize a contemporary kind of explanation of emotions I label "strong cognitivism" that attempts to reduce emotions to (or explains emotions purely in terms of) judgments, evaluations, or beliefs. Strong cognitivism makes it implausible, or at least highly problematic, to ascribe moral emotions to animals because according to strong cognitivism moral emotions will consist of one or more complex moral judgments. The main proponent of strong cognitivism that I focus my criticisms upon is B. A. Dixon, but I will also criticize the strong cognitivist views that have been advanced by Solomon and Nussbaum. I will defend the claim that these specific forms of strong cognitivism are untenable because strong cognitivism as a general kind of explanation cannot fully explain the nature of emotions.

This kind of explanation fails, I argue, first because it simply does not have the resources to explain a wide variety of mental states that are pretheoretically considered to be paradigmatic emotions such as strong fear, love, and empathic concern. Second, strong cognitivism cannot explain how emotions are value determining in a way that can override judgments. Third, it conflates the negative condition that an emotion cannot be had while an agent is aware of a judgment that contradicts a necessary implication of that emotion (I cannot be angry with John if I realize that John did not in fact wrong me) with what an emotion is. Fourth, this kind of account cannot adequately explain the phenomenology and physiology that accompanies emotional states. Finally, I argue that this general category of explanation imposes too strict a criterion for moral motivation in the human case. Actions motivated by an awareness that another is suffering and a desire

to alleviate that suffering are, when a human has them, clearly an emotional state that motivates actions we would pretheoretically characterize as moral even in the absence of explicit moral judgment. Strong cognitivism denies that this is possible, which is a reason to reject it as an explanation of moral emotions as well as more basic emotional states.

The term "strong cognitivism" as I shall be using it refers to any top-down view of emotions that reduces emotions to (or explains emotions purely in terms of) judgments, evaluations, or beliefs.[60] This can be contrasted with weak cognitivism, or any theory that holds that an emotion necessarily requires (but is not solely constituted or explained by) a belief (or set of beliefs) about the object of the emotion. The main proponent of strong cognitivism that I will focus my criticisms upon is B. A. Dixon who explicitly argues against the possibility of animals possessing morally laden emotion in her 2008 book *Animals, Emotion, and Morality*. Dixon accepts that morality can be a matter of being motivated by emotions with morally laden content, but challenges the idea that the content of moral emotions, when correctly analyzed, can be ascribed to nonhuman agents. In what follows, I will argue that Dixon's worries are misguided as they rely on a strong cognitivist view that makes the mistake of over-intellectuallizing the ontology of both emotions and moral motivation.

For Dixon, when an entity is motivated by a moral emotion, this emotion is essentially constituted by several complex judgments. (Dixon 66-69) The argument

[60] The proponents of this kind of view I shall be criticizing include Solomon, Nussbaum, and Dixon, but any view that explains emotions purely in terms of normative judgments will also be subject to the following criticisms.

Dixon gives against attributing moral emotions to animals thereby depends crucially on Dixon's defense of a strong cognitivist account of emotions (and how they become morally laden). Dixon adopts a strong cognitivist account that was originally proposed by Nussbaum.

It follows that Dixon's argument against the ascription of emotions with morally laden content to animals hinges on the plausibility of the Nussbaumian claim that such emotions necessarily involve the agent making a series of complex evaluative judgments. Any emotion, according to Dixon, is essentially constituted by several judgments about the overall situation and the target of the emotion. This is particularly true of moral emotions, which involve an understanding of the morally relevant features of the situation and a moral judgment that takes those features into account.

Before I explain the nature of these judgments in more detail, note the remarkable prima facie counter-intuitiveness of this overly-stringent view of moral emotions. This kind of strong cognitivism would rule out, for instance, *reflexive* (automatically initiated and unreflective) action out of (non-judgmental) empathic concern for the well-being of another as a mental state that is correctly characterized as a moral emotion. In fact, such a view would not even easily be able to explain why we (presumably rightfully) consider such a state to be an emotion. This common sense view of emotions seems completely at odds with strong cognitivism, and so the burden of providing convincing arguments and evidence for a revisionary account of morally laden emotions falls to strong cognitivists like Dixon.

In particular, Dixon must somehow explain why our common sense intuitions about an organism having a moral emotion are *wrong* in cases where the following characterizations of their actions, phenomenology, and bodily changes are true:

1. Actions that are usually described as moral are accompanied by the bodily changes typical of empathic distress that (presumably) register in the phenomenology of the organism,
2. These bodily changes focus the organism on the target of their emotion, and
3. The contribution of these bodily changes to the phenomenology of the organism motivate the organism to act to preserve or increase the welfare of that target.

While there may be no way to empirically discover the exact nature of the subjective contribution of bodily changes that accompany empathic distress in nonhuman animals, we are now justified in attributing value laden empathic states with certain functional and structural characteristics (the core relational themes) on the basis of the (previously argued for) behavioral and neurological evidence for structural continuity of the phenomenology of feelings of empathic distress in animals with extremely similar neurological and chemical states and behavior. Since these common altruistic actions motivated by empathic bodily changes are quite naturally described as moral in the human case, *even when no explicit judgment is involved*, Dixon must provide powerful reasons for us to reject this commonsense view that is supported by the arguments and evidence provided in Chapter One.

To understand why Dixon holds this prima facie implausible view of moral emotions, it will be helpful to consider her reasons for adopting a strong cognitivist account of emotions in general. According to what I will be referring to as the "Dixon-Nussbaum view," emotions in general (including moral emotions) are constituted by

judgments about the object of the emotion. The judgments that constitute the emotions must also capture, contain, or reflect a kind of valuing. (Dixon 66-69)

Dixon's argues that we must adopt such an account of moral emotions in order to explain why certain emotional states have moral content. For Dixon, it is not enough to observe or recount an animal's actions that bear a superficial resemblance to the moral actions of humans and then characterize these actions, on the basis of this superficial similarity, as movements that reveal an emotional state of grief, love, compassion, caring, and so on. Instead, she argues, we must first have a theory that explains what emotions are and how they come to have moral content, and only then will we be able to justifiably determine whether or not animals can have moral emotions.

Without a theory that explains what emotions are beside their characteristic expressions in (voluntary and involuntary) observable movements and how emotions like compassion become morally laden, we cannot determine whether or not animals can be correctly ascribed emotions that are comparable to or similar to human emotions. It is also necessary, Dixon argues, to first explain how moral emotions acquire moral content, and then provide an argument that whatever capacities are necessary for such a morally laden emotion are capacities legitimately attributable to animals. Otherwise, descriptions of animals acting in a way that superficially resembles humans acting morally can only function elliptically or metaphorically. That is, descriptions of animals as compassionate or caring can only function as descriptions that *would* be true if they applied to a *human rational agent* who had an understanding of the relevant values, but these descriptions are to be understood (lacking a theory explaining why animal emotions are morally laden) as

literally false of an animal who acts in an apparently compassionate or caring way. (Dixon 42, 63-66)

Although I disagree with Dixon on many points, here I think she is right on the mark. This dissertation is in some ways an answer to her challenge for a theorist of animal emotion to explain what animal emotions are, why animal emotions are similar to human emotions, how the emotions of *both* humans and non-human animals become morally laden, and what evidence there is to think that animals have the capacities that allow them to be motivated by such morally laden emotions. Dixon and I are in substantial agreement about the methodological point that substantial theoretical justification is necessary for correct attributions of morally laden emotions to animals (although we disagree about the need for substantial theoretical justification for ascribing animals *any* emotions). My criticisms in this chapter are about Dixon's specific explanation of what emotions are and how they become morally laden, not about the general need for such explanations. Let us examine, then, Dixon's preferred explanation of emotions and see if it can dissuade us of the commonsense intuitions we have about moral emotions.

For Dixon, if emotions are to be morally laden, they must be valued contributions to the good life; and a correct theory of morally laden emotions must reflect this fact. (Dixon 42) Again, so far I am in agreement with Dixon – emotions are, as we saw in Chapter One, both value presenting and themselves valuable. My disagreement with Dixon begins with her claim that a correct and adequate theory reflecting the emotions' contributions to the good life is given (perhaps only) by Nussbaum's eudaimonistic theory of emotions. This claim could have been supported by an argument to prefer

Nussbaum's strong cognitivist account over other accounts of emotions, yet Dixon simply neglects (at least in her *Animals, Emotions, and Morality*) to consider other theories of emotion except for an all too brief mention of the rather simplistic view that emotions might be mere bodily reactions, which she (correctly) dismisses as inadequate in so far as mere bodily reactions are in general not morally laden states. (Dixon 65)

Dixon's argument depends crucially on Nussbaum's strong cognitivist account (or another strong cognitivist account along the same lines) being the best explanation of what emotions are and how they acquire moral content. Yet this means that the conspicuous lack of a survey of the (overwhelmingly numerous and diverse) theories of emotions that have been proposed and that purport to explain how emotion acquire normative and ethical content and simply assuming that Nussbaum's theory is both adequate and the best available explanation of emotions as value bearers represents a serious methodological flaw in Dixon's overall argument. This flaw is compounded by the fact that there are notorious problems with Nussbaum's theory of emotions, which has received much critical attention in the literature. I will reiterate some of those problems here, although my criticisms of Nussbaum and Dixon's particular account(s) of moral emotions should be understood as instrumental towards advancing a positive view of how animal and human emotions come to have moral content.

Nussbaum's theory, as characterized by Dixon, is committed to four central claims. The first, which is relatively uncontroversial, is that emotions are intentional; or in other words, emotions are about objects, persons, and states of affairs. Secondly, the intentional objects of emotions are interpreted by the subject. It follows that an accurate description of the emotion must describe the intentional object of the emotion as it is

experienced through the eyes of the subject. Thirdly, emotions are constituted by beliefs and judgments about the object of the emotion. And finally, the beliefs and judgments that constitute the emotions must also capture, contain, or reflect a kind of valuing. (Dixon 66-69)

The beliefs and judgments that are necessary for an emotion according the Nussbaum's account as characterized by Dixon are about:

1. the object of the emotion,
2. the relation between agent and the object of their emotion, and
3. the object of their emotions relation to the overall goals of the agent.

Emotions for Nussbaum are "judgments in which people acknowledge the great importance, for their own flourishing, of things that they do not fully control" (Nussbaum 90) or, in simpler terms, emotions "are acknowledgements of our goals and their status." (Nussbaum 135) Emotions-as-judgments are not necessarily self-reflective or linguistic, but are object directed intentional states that themselves contain evaluations of the world and states of affairs in it. (Nussbaum 126-127, 136) Finally, on the Dixon-Nussbaum account, the following general condition is added to these three necessary conditions for any emotion to be correctly considered morally laden:

4. The agent is aware of the morally relevant features of the situation through judgments that are necessary and jointly sufficient constituents of a moral emotion.

According to the Dixon-Nussbaum account of moral emotions, in order for animals to be correctly described as being motivated by moral emotions they must have an understanding of the values that those emotions necessarily imply, and this understanding must occur through judgments or evaluations that occur to the animal. For instance, in order to have the emotions correctly described as compassionate, caring,

trusting, loyal, and loving, animals must understand and be sensitive to the value of integrity, friendship, caring, trust, and the welfare of others; and make certain judgments about the relationship between the object of the emotion and those values.

It is because of the complexity of the judgments and evaluations 1-4 themselves, the cognitive abilities entailed by an awareness and understanding of these judgments, and the values they are about, that it is extremely unlikely on a strong cognitivist account like the Nussbaum-Dixon view that animals can be morally motivated by such an emotions (or have emotions that are comparable to human emotions in general). Importantly, given the truth of this particular version of strong cognitivism, it is also extremely unlikely that a large class of basic emotional states; such as reflexive fear, empathic distress, and lustful or playful attraction, are correctly considered to be emotions (at least not until a judgment gets involved *after* the bodily reaction and emotional phenomenology takes place). As we shall see, this is a serious and persistent problem for the Dixon-Nussbaum account and for strong cognitivism in general.

The Nussbaum-Dixon view of emotions falls under the broad category of cognitivism. Lyons defines "a cognitive theory of emotions [as one] that makes some aspect of thought, usually a belief, central to the concept of emotion and, at least in some cognitive theories, essential to distinguishing one emotion from another." (Lyons 33) Strong cognitivism, as I have been characterizing it, is sub-category of cognitivism, and refers to *reductive* cognitivist views of emotion. This can be contrasted with somatic accounts of emotions, like the James-Lange view that emotions are forms of awareness of bodily states, and non-reductive cognitivist views, like Lyons' view that emotions are physiological changes caused by evaluations. Because cognitivism is minimally

committed to holding that a thought or belief is a necessary element of emotions, it can be reductive or non-reductive. *Strong cognitivism*, however, refers only to reductive cognitivist explanations of emotions. Taking this general characterization as our foundation for understanding how the Dixon-Nussbaum view fits into various approaches to explaining emotions, we can elaborate on the connections that thought and emotion might have in a cognitive theory that makes it central to the concept of emotion.

The first possibility is that an emotion might simply require a belief about an object. This type of cognitive theory distinguishes emotions from mere physiological feelings, but because it is non-reductive falls short of what Dixon has in mind. Let us characterize this minimal requirement, which I earlier referred to as weak cognitivism, as follows:

> CB: an emotion necessarily requires a belief (or set of beliefs) about the object of the emotion.

This non-reductive requirement can be strengthened into a constitutive claim, in which case:

> CBN: an emotion contains a belief (or set of beliefs) about the object of an emotion as a necessary component.

Finally, this requirement, in turn, might be further strengthened into a necessary and sufficient claim:

> CBC: an emotion is constituted by, and only by, a belief (or set of beliefs) about the object of the emotion.

Since nothing in CBC specifies what kinds of beliefs about an object qualify as an emotion, it must fall short of distinguishing emotions from other kinds of beliefs. And to the extent that emotions require more than just any belief about an object (for instance a belief about the object's position) this simplified cognitive account is incomplete as an account of what an emotion is. Of course, no one really holds a view like CBC, and I do not intend to attack a straw man here. Yet this simplified cognitivist account is useful in so far as we can see how the first or second belief requirements; CB and CBN, might be strengthened into more plausible cognitivist account of emotions.

The cognitivist theories that have been proposed, for instance by Solomon, Nussbaum, and various psychologists usually strengthen CBC in one of the two following ways:

> BE: An emotion is constituted by and only by a factual belief(s) and an evaluation(s) (or a set of beliefs and a set of evaluations), or
>
> BJ: An emotion is constituted by and only by a factual belief(s) and a normative judgment(s) (or a set of beliefs and a set of normative judgments).

Although this is somewhat stipulative on my part, let us consider a judgment and an evaluation to be different kinds of mental states and implying different kinds of capacities. A judgment implies, at least for our purposes, that a process of reflective reasoning has taken place and a decision among (at least some) conceptualized alternatives is reached. An evaluation, however, can simply be a conscious normative preference that does not imply a reasoning process between conceptualized alternatives.

This distinction between evaluation and judgment implies that while every judgment necessarily implies an evaluation, the converse is not true. The evaluation "that

this is good," is distinct from the judgment "that this is better than that." Even though the judgment implies an evaluation, the evaluation need not imply a judgment – something can merely appear to be good when an organism perceives something good-ly (as an adverbialist would put it), and in virtue of this perception-as the organism can evaluate. If we accept this stipulation, then BJ is the strongest possible formulation of the cognitive evaluative view of emotions. It is also the characterizing commitment of any form of strong cognitivism, including the views of Nussbaum, Dixon, and Solomon. In what follows I will explain each form of strong cognitivism committed to BJ and the problems for these views, with the aim of eventually showing how strong cognitivism in general is an inadequate approach to explaining emotions.

Solomon, like Nussbaum, notoriously argued that emotions are rational normative judgments. Solomon holds that emotions can be purposive, and that emotions are (in humans) somehow conceptually tied to normative judgments about the intentional object of that emotion (whether or not the subject is aware of the normative judgment). On Solomon's view, the reason that emotions are held to be a certain kind of judgment is based in his argument that when an emotion's intentional object is independent of its cause, one cannot come to know this without diffusing or negating the emotional state. (Solomon 832) For instance, imagine that I am angry about a bad review. If I come to believe my anger is completely caused by lack of sleep and not by a bad review as I initially thought it was, this belief seems to entail that I cease to be angry about the bad review. Solomon takes this implication to show that if one is to be angry, one must also (at least implicitly) have the judgment that the state of affairs at which one's anger is directed is unjust or wrong in some way. (Solomon 830) However, it is fallacious to hold

that the alleged necessary entailment between emotions and rational normative judgments about the intentional object of the emotion is a sufficient reason to conclude that emotions simply are rational normative judgments.

Additionally, the reductionist view Solomon has put on the table fails to capture or address a wide variety of emotional states. Briefly, these states include emotions that are previous to and influential of our normative judgments (such as empathic states), non-propositional emotions (love or anger with a person, and not a state of affairs), and (as Solomon admits) global emotions or moods like depression and joy that resist and persist through rational criticism. Even disregarding the problem cases of depression, dread, and joy (it may be arguable that these are moods, not emotions, and that there is some principled difference between the two), consider the violence we would have to do to Solomon's picture of emotions to accommodate empathic states.

Empathic emotions function in many ways that are counter to (or independent of) rational judgments. For instance, it is not irrational to walk away from a stranger in need, and being motivated to help a stranger in need by empathic emotions is not dependent on the judgment that it would be a good thing to help a stranger. Rather, feeling empathy for someone in need is more primitive than the judgment that it would be a good thing for you to help them, and one can have the latter without the former or the former without the latter. You can, for example, think it would be a good thing for society to help a homeless person, and a good thing for the homeless person if you helped them, but not feel the motivating emotional state of empathy for them. Alternatively, you can have the judgment that the homeless are unworthy of emotional consideration, that it would be ultimately worse for them if you helped (they'll just spend it on liquor!) and yet feel

empathy at the sight of a homeless person suffering and be motivated by this empathy to hand out a few coins. These considerations show, I think, that the judgment that it would be a good thing (for you, for society, or for the stranger) to help a person in need comes apart from the feeling of empathy for that stranger. Minimally, I think we have established that even if the state of empathy with another's suffering carries with it a rational normative judgment that it would be good to alleviate that suffering, it certainly does not seem to be identical to this judgment.

A further problem for Solomon is that in the case of many emotions, the object of the emotion is not a fact or state of affairs that can be captured in a propositional attitude in the way that Solomon suggests. For instance, feeling love for a particular person doesn't translate into feeling love about such and such a state of affairs. Feeling love *that* a person is beautiful, young, intelligent, and so on; seems to make little sense (one doesn't simply love that a person is beautiful, rather one loves a beautiful person and perhaps admires or appreciates their beauty) and in any case be on the wrong track, since one may love a person whether or not they continue to have any of the these propositionally expressed properties. The same considerations apply to anger, fear, disgust, friendliness (and many, many other social emotions) that one feels with a person instead of a state of affairs. When the object of love is a person, and not a state of affairs, the emotion picks out its object in a way that more directly attaches to the person than love about any particular state of affairs or even the collection of some subset of states of affairs involving that person. Solomon's view is silent on this property of emotions to be (quite often, in my opinion) non-propositional, and there does not seem to be any clear-

cut way to describe non-propositional emotions that are about other organisms in propositional terms.

Even if these objections can be satisfactorily answered, Solomon's view is also more fundamentally flawed, as I have previously mentioned, because it depends on an invalid inference. It is simply logically fallacious to conclude that emotions are identical to the normative rational judgment to which they are conceptually tied from the claim that any emotion about an intentional object necessarily entails a normative judgment. Necessary entailment and identity are two very different logical relations, and a cool judgment that a state of affairs is unjust or dangerous does not entail that the emotion anger or fear is felt about these states of affairs (the entailment does not hold, as it would have to if it were an identity relation, from judgment to emotion even if it does hold from emotion to judgment).

Another mistake is made by Solomon when he infers the constitutive claim that judging someone has wronged me is necessary for me to be angry from the mere negative condition that I cannot be angry with someone if I also think that the target of my anger has not wronged me. Even if we ignore, for the moment, cases of irrational anger, it is still perfectly possible that someone can feel anger and yet not make the positive judgment that they have been wronged by someone. This will undoubtedly occur when agents who are simply not be aware of this necessarily entailment become unreflectively angry.

Finally, Solomon's attempt to distinguish cool factual judgments from hot emotional judgments by the latter's rashness, immediacy, myopic nature, and appropriateness to a situation that is counter to our rational purposes, ultimately fails

because emotions can arise in the absence of any or all of these conditions. Anger at the injustice of slavery and racism can be arrived at slowly and methodically as one researches and learns of the tragic course of history and human nature, and this judgment is not rash, immediate, myopic, or appropriate to an obstacle-like situation. However, this judgment is for many who come to learn of this injustice of the hot emotional variety and not the cool judgmental variety. Emotions are differentiated from judgments by the fact that they are accompanied by a motivating feeling that is not entailed by a mere normative judgment. The invalidity, extensional inadequacy, and glaring counter examples to Solomon's form of strong cognitivism indicates that we should reject this particular version of strong cognitivism, and begin to suspect that the general commitment of strong cognitivism to BJ makes this form of explanation untenable.

Let us turn now to the strong cognitivist approach of Nussbaum that Dixon uses to argue against the possibility of animals having moral emotions. Nussbaum's account is similar to Solomon's in many ways; and so to a great extent it faces the same or similar theoretical difficulties. She differentiates her view from that of Solomon in so far as her account does not consider emotions as necessarily involving value positing that is willed and subjective. She remains agnostic on whether values are willed or subjective, and instead "presents the valuation nature of our appraisals from the internal viewpoint of the person having the emotional experience." (Nussbaum 22-23)

However, the difference between emotions being explained as eudaimonistic judgments from the internal viewpoint of the person having the emotional experience and emotions being accounted for as willed subjective judgments about the world is not a difference that will insulate Nussbaum from the general extensional inadequacy that

plagued Solomon's view. In any case, there are more specific theoretical difficulties that Nussbaum's view faces, and so let us take a particular example, the emotion of grief, to further explicate the details of the view.

Recall that for the Nussbaum-Dixon view emotions are "judgments in which people acknowledge the great importance, for their own flourishing, of things that they do not fully control." (Nussbaum 90) The beliefs, judgments and evaluations that are necessary for an emotion are about:

1. the object of the emotion,
2. the relation between agent and the object of their emotion, and
3. the overall goals and flourishing of the agent.

Yet it is not enough that an emotion simply track or express these judgments when described from a third person point of view, rather these judgments must occur to the agent and the agent must be aware of making these evaluations about the object of the emotion. For example, grief at the loss of a good friend is intentionally directed in the sense that it is about the loss of a good friend. How the loss of a friend is represented is something that must be understood as interpreted through the eyes of the subject. This state of grief must also involve some fairly complex cognitive-evaluative states. For instance, grief about the loss of a friend must involve beliefs about what a friend is and the judgments that friendship is a good thing and that it is important to one's overall goals and flourishing. Finally, there must be an acknowledgement that losing a friend is a terrible loss to one's overall goals and flourishing. That is, the state of grief must necessarily involve judgments that reflect a deep understanding of the value of lasting friendships, the value of the trust, caring, and intimacy that are involved in such a relationship, and the seriousness of losing these values. (Nussbaum 49-50)

This conception of grief can be challenged in several ways. First, it makes it highly problematic, and perhaps almost always theoretically unjustified, to ascribe grief to animals. The well-known cases of elephant grief at loss of a matriarch and the profound and sometimes fatally crippling grief of young chimpanzees at the loss of their mother are two well researched and documented cases of what many ethologists characterize as animal grief. It is unlikely that these animals are forming judgments that reflect a deep understanding of the value of lasting friendships and caring family members and the seriousness of losing these entities and relationships. It follows that on the Dixon-Nussbaum account we cannot say that such animals are grieving. It is unclear if Dixon and Nussbaum accept this problematic implication, as they hold that

> Some emotions will prove altogether unavailable to many animals, to the extent that the sort of thinking underlying them proves unavailable: [including] grief, to the extent to which [it] requires causal and temporal judgments. (Nussbaum 146-147)

On the Dixon-Nussbaum view grief does require judgments about values and temporal judgments about the permanence of the loss (or else the loss cannot be understood as serious to one's life), and so if pressed they will have to deny that animal grief is possible. At the very least, it becomes highly problematic to explain how animals often seem to be experiencing such devastating grief that they are literally incapacitated by their emotional state.

However, refuting the Nussbaum-Dixon view is more complicated than simply citing such cases as counterexamples, since it is always open to Dixon and Nussbaum to deny that these really are cases of grief. Accordingly, in what follows I will criticize the account of grief defended by Nussbaum with the aim of demonstrating the general

inadequacy of her overly cognitive account of emotions. For Nussbaum, what makes an emotion a state of grief is not any particular physiological sensation, but rather the urgent judgment that "a central and very important person to my life and my goals and projects is gone." (Nussbaum 126) This judgment can be, and must be in the case of many nonhuman animals for which grief is possible, not a product of reflective self-awareness, but rather (something more like) a vague way of intending an object and marking it to conscious experience as having a certain importance, urgency, or salience. (Nussbaum 126, 129)

Strangely, Nussbaum claims that this way of intending an object need not be linguistically formable. Yet it seems highly problematic to claim that a nonlinguistic, non-reflective state of vaguely intending an object has the content that an incredibly important aspect of one's life is gone. Notice that the APV account of emotions can explain how, by being aware of the background contribution of bodily states and involuntary behavior as representing the value of the intended object, emotions that are nonlinguistic can express values. However, Nussbaum gives no explanation for how judgments that are necessary and sufficient for emotions can occur to non-linguistic mammals vaguely intending the object of their grief.

Recall that for Nussbaum and Dixon, if an emotion is to be correctly ascribed to a being, that emotion must be described as it appears to the subject, or through the agent's own subjective point of view. In order for an animal (or infant) to have such an emotion it must also somehow make subjectively accessible appraisals about its life goals, the overall value of the intentional object of the emotion, and the relevant surrounding normative concerns. These judgments are somehow passively revealed to the animal (or

infant) without them being reflectively aware of any of these things and without any of these judgments having a linguistic structure. It seems that either the mechanism for how these judgments occur to animals and infants that have emotional states is utterly mysterious, or else Nussbaum must deny that animals and infants experience any emotions at all. Neither option is very plausible.

Yet let us abstract from the problem for strong cognitivism that there is no viable way to account for the way that these judgments are structured without reflection or linguistic competence being elements of the structure. Even granting that these judgments are possible and correctly described as judgments; we must also determine if it is really necessary to being in a state of grief that a grieving being makes a collection of judgments about their overall life goals. It seems more likely that grief can occur as a somewhat non-reflective visceral feeling accompanied or caused by memories of the lost one. In fact, this is the very way that Nussbaum herself initially describes a state of grief that she felt towards the loss of her Mother. In cases of grief where one feels as if the "nail of the world," has ripped a gaping whole (to use Nussbaum's own visceral description), reflective judgments about life goals are often the furthest thing from one's mind, which instead tends to dwell on the deceased in helpless sorrow.

The strong cognitivist account also seems to be deficient as a general account of what emotions that run counter to rational normative judgments. For consider how you might have a friend who you know to be a bad influence but who you nevertheless deeply care about and would be grieved if you lost them. It is even possible in such a case to have a friend who you know does not reciprocate the friendship, and yet whom you feel a strong emotional bond towards despite having complete knowledge that this friend is bad

for your long term goals. You might have such an emotion even if you realize that this kind of friendship is not valuable like a reciprocal friendship would be. Still it is possible you would mourn the loss of such a friend to the extent that you deeply cared about them not because of their prudential value, but rather as a result of a strong empathic connection.

A more damaging kind of relationship follows this pattern and is worth mentioning as a counterexample. In tragic patterns of domestic violence a person sometimes stays with their spouse out of irrational but heartfelt attachment. What is happening in such cases seems to be a prior and value determining emotional state is influencing the agent, and the power of this emotional state makes the agent simply want to be around the harmful person they feel attachment towards. That is, a prior emotional attachment guides the askew value judgments of the participants in such relationships. This prior emotional state is value determining, independent of rationality, and deeply affects the abused spouse's ability to be persuaded by the eudaimonistic judgments that Nussbaum incorrectly identifies as emotions.

In such cases, Nussbaum would no doubt maintain that the agents do not sincerely apprehend the judgments that their relationship is not ideal, is bad for their long term goals, and is bad for them; yet what can such an apprehension amount to for a Nussbaumian view? Either it is yet another cognitive judgment or it is whatever else must be added to a judgment to make it an emotion. If it is the latter, than emotions are not reducible to judgments at all, and the idea that judgments partly constitute emotions is cast in doubt (for whatever is added would be sufficient without the judgment). If it is another meta-judgment or meta-appraisal that constitutes sincerity, then the issue of what

makes a judgment of the importance of something to one's life goals sincerely apprehended merely gets deferred to the issue of whether the meta-judgment is sincere. For one might have the meta-judgment and not sincerely apprehend it just as one has the lower order judgment without sincerity. Then either a further judgment is required, generating an infinite regress of judgments confirming lower order judgments, or an emotion not identical to a judgment constitutes the sincere apprehension of that judgment. Notice that the regress is vicious because the sincerity of the first judgment depends on infinitely more confirming judgments, none of which is sufficient for sincerity. The alternative is to hold that sincerity of the meta-judgment is a matter of feeling a certain emotion that confirms the judgment, or being in some mental state that is not equivalent to the judgment. Either way, the motivation for explaining the first order emotion/judgment in terms of further judgments about value and life goals is undermined, and emotions quite clearly do not reduce to sets of judgments of this kind.

It also seems that people can sincerely assent to these evaluative propositions, even eudaimonistic ones, and yet feel no emotion. For example, this assent is quite independent from the feeling of grief. Otherwise it should be impossible for a psychopath incapable of empathy or caring to coldly understand the proposition that his Mother is dead, and assent to a series of eudaimonistic judgments that someone who was extremely important to his life goals is lost. Or consider the possibility of an alien being who received immense pleasure every time a serious loss happened in his life. Such a being would also be able to assent to the eudaimonistic judgments concerning loss, and feel great pleasure and no cognitive or visceral pain that seems necessary for a true (or

'sincere) state of grief. It seems, then, that assent to eudaimonistic judgments, even if dynamic and about an urgent goal, are simply not enough for an emotional state.

While we have been focusing on a specific emotion, the Nussbaum-Dixon account of emotions can also be criticized on rather general grounds as an inadequate as a theory of emotions for its inability to account for the following five categories and properties of emotions:

1. Irrational emotions,
2. Emotions that are value determining and prior to judgment,
3. Certain physiological facts about emotions and neurological research on what physiological states and behaviors emotions like fear entail,
4. The phenomenology of emotions, and
5. Unreflective empathic emotions.

It is extremely problematic to explain, on the strong cognitivist strategy that Nussbaum employs, how some emotions, such as phobias, are completely irrational. I may feel an irrational fear of heights even though I sincerely judge that I am safe behind numerous guard rails. Yet the emotion may persist and irrationally compel me to descend in a way that overrides my sincere judgments about my safety. Again, an irrational attachment (I hesitate to use the term love) to an abusive significant other or bad friend are also well-known cases of what seems clearly to be a persistent and irrational emotional disposition overriding sincere judgments. These cases seem to indicate that judgments are something that occur further down the causal stream from emotions, at least in these instances of particularly strong irrational emotions.

These emotions also seem also to be value determining. That is, these irrational emotions focus on a specific person or state of affairs and, without the influence of rational and reflective judgment, present the value of that person or state of affairs

passively to the agent, who is then compelled to accept the negative (as in the case of an irrational fear) or positive (as in the case of irrational love) value of the intentional object despite their better judgment.

As we have seen, there are also well known neurological circuits that are homologous in all mammals that underlie states of fear, including irrational fears and phobias in humans. The behavioral and neurological similarities of such a wide variety of animals, and the fact that many of these animals are not capable of any kind of assent to judgments, indicates that fear is more a matter of the physiology of organisms and what physiological reactions indicate about the intentional object of the emotion to the agent.

These physiological states are left out entirely on the strong cognitivist picture, or added as an afterthought (as perhaps necessary, although no reason for their necessity is given). Yet the feeling of trembling, the sickening sinking feeling in ones stomach, the cold sweat and paralysis that accompanies an irrational fear of heights when one judges oneself to be perfectly safe, all contribute to the phenomenology of the relevant emotion in a way that an adequate theory of emotions must explain. Emotions are not just propositional judgments about an intentional object, even if these judgments are dynamic and eudaimonistic. Rather, there is a significant contribution from the bodily states and how they combine in the agent's phenomenology to present the value of the intentional object.[61]

[61] A possible response from Nussbaum here would be to claim that judgments have a phenomenological character that captures the apparently visceral nature of emotional reactions; however, the hypothetical case of Cog Aliens and the actual case of psychopaths are beings for which these judgments can occur without the phenomenology characteristic of emotions. This, I submit, is precisely because of a lack of somatic involvement (via the awareness-with relation) in the mental states of such beings.

This contribution cannot be separated from the emotion, for it is what makes the emotion urgent and motivating in a way that often overrides our best eudaimonistic judgments. The reason that emotions motivate is that they are expressed through motivating bodily reactions (whether or not we are aware of those reactions), and these reactions motivate us to do something. Emotions affect what we do and what we believe through bodily agitations understood as pleasurable or unpleasant bodily changes that contribute as vehicles expressing the value of the object of the emotion.

This is why, I submit, that there is vast and growing literature on emotions that occur before the neurological mechanisms underlying higher thought and that influence those mechanisms, often in ways that subjects are unaware of. These influences are referred as affect priming and/or unconscious emotional bias, and are thought to not only sometimes precede and influence judgment but also to sometimes persist even after we have contradicting thoughts. Fear, for instance, can persist long after we are cognitively convinced there is nothing to fear, which is precisely why fear can be irrational.

Finally, the case of unreflective empathy, such as the immediate empathy I might feel for a starving stranger with an expression of utter despair, is deeply problematic for the Nussbaum-Dixon strong cognitivist explanation. While such an explanation is committed to emotions being eudaimonistic judgments (or an assent to such judgments), you need not judge that it is a good thing for you to help a stranger that you feel empathy for to be emotionally motivated to help. Instead, there is an unreflective empathic transfer of emotions that is previous to any such judgment.

As I argued when criticizing Solomon's strong cognitivist account, feeling empathy for someone in need is more primitive than and independent of the judgment

that it would be a good thing for you to help them. It is perfectly possible to think it would be a good thing for society to help a starving stranger, and a good thing for the starving stranger if you helped them, as you coldly ignore them because you lack any motivating emotional state of empathy for them. And of course, you can have the judgment that the homeless are unworthy of emotional consideration, that it would be worse for them if you gave them money that they would unwisely spend, and yet feel empathy at the suffering of a starving stranger and be motivated by this empathy to give them some amount of money. The judgment that it would be a good thing (for you, for society, or for the stranger) to help a person in need is independent of the emotional feeling of empathy for that stranger.

In light of the criticisms of the strong cognitivist explanations of Nussbaum, Dixon, and Solomon, let us consider again the reason that Dixon offers that animals are unlikely to have morally laden emotions. Her reasoning, recall, was that in order to have morally laden emotions, one must be able to make several complex judgments about the situation and the relation between the agent and the target of the moral emotion. In particular, in order to have a moral emotion, one must form judgments or have beliefs about the following four entities:

1. the object of the emotion,
2. the relation between agent and the object of their emotion,
3. the overall goals and flourishing of the agent, and
4. The morally relevant features of the situation.

To take a specific example, consider that for Dixon and Nussbaum what makes an emotion like compassion morally laden is that it is an emotion that is occasioned by an

awareness of another person's undeserved misfortune. That is, compassion, when analyzed with respect to its morally laden content, implies three necessary conditions:

1. The agent must understand the proportionality of the misfortune such that the misfortune is judged or evaluated as a serious loss for the sufferer. This requires adopting the point of view of the sufferer and seeing the loss through their eyes.
2. The agent must understand that the suffering is undeserved, either because the agent is not at fault or because the suffering is disproportionate to the faults of the agent.
3. And finally, the agent must be able to appreciate that the circumstances causing the suffering are circumstances that would cause suffering to the compassionate person if their situations were reversed. (Dixon 66-68)

By contrast, a raw physiological feeling of kindness, unlike compassion, cannot have moral content because it does not involve this cluster of judgments directed at the appropriate kinds of intentional objects.

Yet when we commonsensically regard a human action compassionate, such as a mother comforting a scared child, or the 9/11 firefighter respondents rushing to save those trapped in the Twin Towers, we do not require that these compassionate agents understand and form the above necessary and sufficient judgments in order to be regarded as motivated by compassion. Neither do we require a human to be able to list the judgments that they *must* make according to the Dixon-Nussbaum view when they, for instance, feel immediate compassion for a suffering injured child (lest we confine compassion to *very* reflective philosophers). Rather, the behavior of a human in these instances is naturally regarded to be compassionate because it is motivated by an emotion we can describe as heartfelt warmth and concern for the wellbeing of others. This emotion moves the compassionate agent to help another who is suffering, yet it does not reduce to these value judgments. It follows that even if we grant the highly problematic

Dixon-Nussbaum view of emotions, the explanation of how emotions gain moral content is inadequate because it cannot explain paradigm instances of human compassion. But we need not grant Dixon so much, as we have now seen that the strong cognitivist view of emotions that allows her to deny that animals can have moral emotions is independently implausible.

The same criticism against strong cognitivism's overly stringent ascription conditions of emotions like grief therefore also apply to this overly stringent account of compassion. Just as one can be overwhelmed by thoughts of a loved one and the phenomenological mental anguish one feels at this loss in a way that precedes and precludes eudaimonistic judgments of any sort, empathic concern for the well-being of another that moves an agent to relieve or prevent their suffering can precede and preclude judgments about whether the agent is at fault for their suffering or how serious the suffering is to the agent. We have seen how empathic motivation and rational judgment are independent of one another, and since through empathy the value of the other's well-being is a motivating concern for the empathic agent, these judgments (that may or may not be motivating) are unnecessary and beside the point. The difference between sincere and insincere compassionate judgments will also pose a problem for this view, since whatever makes the judgments sincere is either a further judgment which can itself be called into question (generating an infinite regress), or something else (an emotion) must be added to moral judgments in order to make them sincere moral judgments. Finally, it is not available to Dixon to argue that, even if Nussbaum's view of compassion is untenable, another strong cognitive account might be correct. The theoretical problems

we have seen with Dixon, Nussbaum, and Solomon are not specific to those particular accounts but generalize to the strong cognitivist strategy in general.

Because strong cognitivism is committed to a BJ reduction of emotions, it simply does not have the resources to explain a wide variety of mental states that are uncontroversially considered to be emotions such as strong (or irrational) fear, love, and empathic concern. Also, it cannot explain how emotions are value determining in a way that can override whatever judgments that the particular form of strong cognitivism conflates with emotions. As long as there are conflicts of emotions and judgments, as there will be in the case of irrational emotions, a reduction is untenable. Third, this kind of account cannot explain the phenomenology and physiology that accompanies emotional states, as these factors are not necessary for judgments in general. Finally, strong cognitivism in any form imposes too strict a criterion for moral motivation in the human case. Actions motivated by an awareness that another is suffering and a desire to alleviate that suffering are, when a human has them, clearly an emotional state that motivates actions we pre-theoretically would characterize as moral even in the absence of explicit moral judgment. Unless we have prior reasons to think that emotions motivating moral actions must involve such judgments, then such actions count as counterexamples to the Dixon-Nussbaum picture of moral motivation. The only prior reason that is offered by the Dixon-Nussbaum picture seems to be that, in general, emotions can only be explained in terms of a set of constituent judgments. However, this reason has been thoroughly dismissed.

Importantly, these problems occurred for the specific versions of strong cognitivism critiqued here precisely because of the common element of attempting to

reduce emotions to judgments or evaluations of some kind. Since this general commitment characterizes strong cognitivism, we ought to reject any form of this problematic explanatory strategy. And if we reject strong cognitivism, we need no longer be troubled by the fact that, if it were true, it would make ascribing animals moral emotions highly problematic or impossible.

I conclude that, because strong cognitivism cannot offer a compelling and complete explanation of the nature of emotions in general, there is little reason to think that moral emotions involve the kinds of complex judgments that strong cognitivism holds compose all emotions. Accordingly, strong cognitivism as a general strategy of explanation and the implications that it has against attributing animals moral emotions ought to be abandoned in favor of a view that does not equate emotions with judgments. It remains, however, to explain how emotions acquire moral content if not by the series of judgments that the Dixon-Nussbaum view proposes.

We ought to be comforted rather than dismayed that this deficient and inadequate account of emotions directly contradicts the APV theory of basic and empathic emotions I have proposed. The way that this account of basic emotions can be elaborated to include moral emotions is by combining this account with an explanation of what is added to basic emotions (or any emotions) to make them morally motivating if not the kind of moral judgments that the Dixon-Nussbaum view proposes. This explanation must also not imply conceptual or cognitive abilities that are beyond the ken of nonhuman animals, but here I think we needn't be too worried. There are, as I will argue, at least two ways in which animals can be justifiably ascribed emotions that are morally motivating. But first we must determine what human morality is, and how it relates to

human emotions, before we can determine whether animals are also moral creatures. There is a Kantian tradition in ethics that denies that even human emotions can be morally motivating, which would in turn entail that animals cannot be morally motivated on the basis of empathic emotions such as compassion or benevolence. Accordingly, in the next chapter I will criticize this Kantian tradition that holds that human morality is a matter of critical reflection and rational self-government, focusing in particular on Korsgaard who has explicitly denied the possibility of animal's being motivated by moral emotions.

Chapter 3: Morality without Critical Self-Reflection in Humans and Animals

Summary

A traditional view of morality, most notably and recently defended Korsgaard, holds that moral action necessarily requires the capacity for rational self-reflection and normative self-government of one's actions and purposes or motives for acting. As Korsgaard and others have noted, this puts moral behavior beyond the ken of nonhuman animals. In this chapter I challenge this assumption and the adequacy of the general metaethical picture that requires rational reflection and normative self-government for moral behavior. I argue that when (some) social emotions motivate altruistic actions, this is sufficient for moral behavior in humans and animals.

If it is true that emotions can be regarded as having moral content either because of their empathic properties or the role they must play in human moral behavior, then we have good prima facie reasons to think that a subset of animal emotions can be correctly specified as moral. Yet if self-reflective moral judgments are necessary to moral motivation, this rules out the possibility of certain unreflective empathic emotions being sufficient for moral behavior.

Unfortunately, rational reflection is commonly thought of as a necessary requirement or common starting point for moral behavior.[62] The kind of rational reflection that the traditional Kantian conception of morality holds is necessary for moral behavior is (the capacity for) critical scrutiny of one's motives and goals, and the ability to determine which of those motives and goals are morally justified and which are not.

[62] This claim is explicitly defended by Korsgaard (2010) pg. 6.

On this view, the capacity to reflect on, and then endorse or reject, one's motives and goals is what determines that agents are morally responsible for their actions, motives, and character. Importantly, this traditional view of ethical behavior holds that without (the capacity for) such critical scrutiny, agents cannot be praised or blamed for their motives, goals, and behavior, and hence are not properly regarded as moral agents.[63]

As the following arguments will show, this view of moral behavior is in direct conflict with our commonsense intuitions about what qualifies as moral action for human agents and with ethological characterization of the apparently morally laden emotions of animals. For example, it is intuitive to suppose that a mother, whether human or nonhuman, can unreflectively act with compassion and caring concern for their young when they comfort them and defend them against threats. When a female human, wolf, elephant, or chimpanzee comforts their frightened offspring, it seems right to characterize their emotional state as compassionate and expressing caring concern for the safety and well-being of their young. These emotional states are naturally regarded as moral emotional states. Similarly, we might regard a human mother's actions when she comforts and defends her infant as compassionate and brave even if the following psychological description were true of her:

1. Agent M does not critically reflect on her motives or goals.

[63] It is worthwhile to note, at the outset of this Chapter, that critical self-reflection and normative self-government may not be necessary for morality in all cases of human moral behavior. Hoffman and Slote have argued, for instance, that it is enough to think about the harm that one's actions would cause to ensure a primitive kind of guilt that doesn't require or imply moral judgment. An action that is resisted because of this rudimentary sense of guilt is, on this view, an instance of moral behavior; as is an action that is undertaken because of the anticipated benefit for another. A child feels such rudimentary guilt when their attention is drawn to the harm that their actions cause in others via a process of induction, where a parent or caregiver elicits in the child the feeling of guilt by attempting to get them to imagine how the agent wronged by that child feels as a result of the child's action (Slote *The Ethics of Care and Empathy* 15). The self-consciousness that a child feels about their own actions is not, in this case, about the moral status of their actions but the harm that their actions have caused another. Self-reflection and critical scrutiny of one's actions and motives need not involve moral concepts like wrongness, which is an important initial point against the traditional Kantian ethical conception that Korsgaard advances.

2. Agent M acts instinctually, in the sense that she does not choose to act in one way over other considered alternatives, but simply responds to the immediate needs of her child.

3. The motivation for Agent M's actions is her fear for her child's welfare and her empathically motivated desire to alleviate her child's distress.

4. This motivation is not arrived at through a process of reasoning resulting in a reason to act, but rather a reason or motive to act is contained in the content of her emotional state of empathic distress.

This characterization, if true of the agent, qualifies the behavior a mother to be moral irrespective of whether the mother is human or nonhuman, or so I shall claim.

This claim has been explicitly challenged by ethicists committed to the traditional view by two separate but related arguments. The first, perhaps most systematically argued for by Dixon but also implicit in the view of Nussbaum and other strong cognitivists, was that when we analyze why emotions such as empathic distress and compassion are correctly considered to be moral emotions, we discover that the emotions themselves involve capacities (like the capacity to make linguistic judgments, judgments about one's life goals, judgments about desert etc.) that are beyond the abilities of nonhuman animals. Since I have already argued that strong cognitivism is a view of emotions that is highly contentious and problematic, we need not concern ourselves with this first challenge here. However, it remains possible that even if animals can be motivated by the emotions I have called compassion and benevolence, where the former is understood as being moved to prevent or alleviate the suffering of another and the latter is understood as being moved to promote the happiness or well-being of the other, these emotions are still not sufficient for moral motivation. This brings us to the second challenge.

Essentially, this challenge is that even though an emotion like compassion might very well motivate a nonhuman agent to behave in a way that we would call moral if acted out by a human with reflective capacities; a nonhuman agent and their behavior are not properly regarded as moral because the agent lacks the capacity to reflect on her goals and her reasons to endorse or reject those goals. Since a human mother has this capacity, even though it is not active at the time of her compassionate action, she qualifies as a moral agent. A nonhuman animal, however, (probably) lacks this capacity altogether, and so cannot be correctly described as acting morally even if the same psychological description is true of that animal. To overcome this second challenge for my view I must argue that self-reflective and self-governing capacities are not necessary for moral behavior. I will address my arguments against the second challenge to Korsgaard, who has presented the most direct and ingenious arguments for the traditional conception of morality.

Before I explain Korsgaard's arguments to the contrary, it is necessary to take stock of the emotional and empathic capacities of animals that give us prima facie reason to think that animals can be morally motivated by empathic emotions. Recall that the empathic development in nonhuman animals is a continuous process of imbibing the empathic behaviors of caregivers and close conspecifics and learning to empathize with ever-greater sophistication through the implicit and explicit teaching of empathic behaviors by caregivers and conspecifics.[64] It is vital to understand that nonhuman empathy is not an unconscious reflexive capacity in animals, but one that develops in

[64] Preston and De Waal offer a good summary of the cross-species references for the development of empathy, (rather than its operation as a simple on-off switch) which increases in sophistication and strength with the empathizing subject's implicit and explicit learning, familiarity with the situation, and familiarity with the target of empathy itself. (Preston and De Waal pgs 3, 7)

sophistication and flexibility (as to what kinds of behaviors it motivates in response to different contexts) through the normal process of social interaction in group living mammals. This development does not require self-consciousness or an organism's reflection on the motives and consequences of actions, but rather occurs as a result of empathic states becoming increasingly more sophisticated as the organism's general capacity for emotional contagion interacts with their increasingly sophisticated ability to distinguish and react to the emotions of others. This is why, as we have seen, de Waal proposes that we understand this general mammalian capacity as "increasing with familiarity (subject's previous experience with object), similarity (perceived overlap between subject and object, e.g., species, personality, age, gender), learning (explicit or implicit teaching), past experience (with situation of distress), and salience (strength of perceptual signal, e.g., louder, closer, more realistic, etc." (Preston and de Waal 3) Empathy increases in strength and sophistication as the organism develops both cognitively and emotionally so that it can internally represent the complex emotional states of others and be motivated by these complex empathic emotional states (that may require cognitive abilities like memory, association, expectation, and categorization) to promote the well-being of others and alleviate their suffering.

It is of particular importance to note that empathy increases with past experience with the object and situation and with implicit or explicit teaching. Studies have been done on apes, monkeys, rats, and human children by a variety of researchers that verify this finding (Preston and De Waal, 3), indicating that empathy develops across species in a flexible and developing way that is responsive to not just the examples of empathic

behavior by caregivers and conspecifics, but also to the organism's individual learning that occurs with repeated experience with the paradigm scenario.

It is initially through the mother-infant (or caregiver-infant) bond that "Continuous and coordinated emotional and physical contact between the mother and infant are thought to organize the emotion regulation abilities of the infant, which determine the emotional competence of the individual (e.g., Brazelton et al. 1974; Deboer & Boxer 1979; Gable & Isabella 1992; Levine 1990; Stern 1974; 1977)." (Preston and De Waal 7) This emotional regulation happens as the infant gets more accurate at interpreting the mother's actions and expressions, which are "mapped onto existing representations of the infant and generate actions and expressions in response." (Preston and De Waal 8) While a main source of empathic development in humans is induction, which requires self-consciousness on the part of the child being elicited by the parent, it is not the only source of refining and developing empathy in humans or animals. It is important to note that empathy in animals does not, because it lacks self-consciousness, thereby become a rigid unresponsive and non-developing instinct, but is refined in subtle ways through experience and learning.

Finally, this empirical evidence is, I have argued, best explained by the APV model of basic and empathic emotional states. According to this model, an animal internally represents core relational themes through the background contribution of their empathic bodily reactions to phenomenology. The core relational themes of interest to the present arguments against animals being morally motivated are those of compassion, which involves internally representing the suffering or pain of another as psychologically negative in a way that motivates the organism to alleviate or prevent that suffering, and

benevolence, which involves internally representing the delight or pleasure of another as psychologically positive in a way that motivates the organism to act to promote or sustain that state of delight or pleasure. The content of these empathic emotional states represents the well-being of others, and because these states develop in discriminatory power (with respect to the intentional object, the wellbeing of that object, and the appropriate actions to promote their wellbeing) as the organism naturally becomes socialized by parents, caregivers, and conspecifics, the way that an animal represents this content and the helping or kind behavior these content bearing emotions motivate become increasingly sophisticated.

Despite this empirical evidence that animals act empathically with the content of their motivating emotional state representing the well-being of another, Korsgaard argues that is not the content of the intention of an action that makes that action moral or immoral, even if the content involves the welfare of another.[65] To persuade her reader that the content is beside the point, she begins her argument by explaining that to say something acts with a purpose is a description that covers a wide range activity, including even functionally describable movement. She explains that machines, organs, and plants act with an intention in this very weak sense, because the purpose of their various activities can be described by a function arrived at through natural selection or through the derivative intentionality of human creators.[66] With animals such as cockroaches and spiders, however, Korsgaard holds that a different kind of intentional description is appropriate.

---

[65] The full argument can be found in in Korsgaard's "Morality and the distinctiveness of human action" beginning on p 107.
[66] Ibid 107-108

For the rigid, stereotyped, and largely tropistic behavior of (many) insects, it becomes appropriate to describe purposeful behavior as an animal intending to perform a certain activity that is guided by that animal's perceptions. However, even though the movements are guided by the animal's perceptions, there is no need to say that the purpose of those movements (getting away from a swatting hand, or attaining food from prey) is before the animals mind.[67] Korsgaard does not elaborate on why she thinks it is explanatorily irrelevant to posit purposes that are internally represented, but we may charitably support this claim with the argument that the stereotyped and tropistic movements of a cockroach or spider are so rigidly determined by perception that there is no evolutionary or biological need for, and hence no evolutionary or biological reason to posit, an intermediate mental representation of a purpose that the spider or cockroach intends to accomplish.

Yet with intelligent animals, Korsgaard maintains, there is contrastingly no reason not to posit purposes and desires that are before the animal's mind in a correct explanation of their purposeful behavior. However, by being before the animal's mind she means not that they experience objects of awareness (in the sense of being aware *that*) that are desirable or to be desired, but rather that they are aware of the objects of their desire as desirable or "to-be-sought." The difference between such an animal and a rational animal is that a rational animal is aware not just of objects as desirable but also aware that they desire such objects. However, she concedes that there is likely a gradual continuum between purposeful behavior in the former sense of functionally organized

[67] Ibid 108

systems guided by perceptions and purposeful behavior in the latter sense of awareness of some goal that is intended and before the animal's mind.[68]

This level of intentionality, where an animal is aware of its purposes and capable of even deliberating about how to accomplish those purposes, is not yet the level of intentionality Korsgaard holds to be necessary for moral agency (and by implication, moral behavior). This is because even if an animal is aware of their purposes and capable of instrumental reasoning about those purposes; still an animal's purposes are not chosen, but are rather given to it by its emotions and instinctual desires. The deeper level of intentionality that animals lack entails the assessment of purposes themselves, in the sense of considering whether your wanting a particular goal justifies, or gives you a reason to pursue, actions towards that purpose.[69]

A purpose and the action towards that purpose are morally justified, at least for Kantians, if and only if undertaking the action for that purpose can be universalized without contradiction. For Kant, an agent capable of this kind of deliberation can reject an action and the purpose for which that action would be done not because of an emotion, but rather because the action, when done for that purpose, is judged to be wrong. Rational adult humans, according to Kant (and Korsgaard), are capable of denying their most urgent and natural desires in order to refrain from doing an action that is deemed wrong in this way.[70]

However, while it is true that normal adult human beings have the capacity to avoid such an action, we are left wondering how this capacity and the desire for acting morally is strengthened through moral education. In other words, it is left mysterious

[68] Ibid 108
[69] Ibid 110
[70] Ibid 110-112

how one becomes the type of being that can be motivated to develop a sense of universally right and wrong actions and then act from a conception of universal justice against their natural instincts. As Sentimentalists such as Slote and Hoffman have noted, Kantian accounts of moral education, such as Kohlberg's cognitive-developmental model, have notorious problems with explaining how children are supposed to become motivated to take the interests and claims of others into account and be willing to sacrifice, compromise, or negotiate their own claims and interest instead of simply understanding the perspective of others as a basis for egoistic manipulation. (Hoffman 131) This is because Kantians neglect the role of affect (empathic affect is precisely the basis for the child's choosing to take others' interests into account) and overemphasize the rational cognitive process of decenterizing the child's capacity for judgment.

According to the Kantian picture of moral education, this decenterization occurs as the child progresses from being able to understand and form judgments about externally determined consequences of their actions, to understanding conventional moral rules that take into account group welfare, to finally being able to understand and make judgments from an autonomous and principled level on the basis of universal moral rules. The driving motive for a child to go through this process according to such theorists seems to be exposure to moral information at a higher moral level that then generates a cognitive disequilibrium that the child desires to resolve by integrating this information with their own point of view. (Hoffman 129) Essentially, the problem for such theories is that this new information, and new capacity for perspective taking, precisely because it does not involve prosocial affect, is perfectly consistent with the child using perspective taking information for egoistic rather than prosocial ends. (Hoffman 131)

To see the force of this criticism, consider how Korsgaard illustrates the capacity for denying urgent natural desires in order to avoid acting on a maxim that cannot be universalized without contradiction. Using an example from Kant, she argues that a man can accept death rather than give false testimony against another man. Such a man can do so even though his egoistic desires and instincts toward self-preservation, and, we might add (modifying the example to make it more realistic and to begin to illustrate the weaknesses of this view of morality), his desires to preserve and care for his family and those that depend on him motivate him in the opposite direction. According to the Kantian conception of morality, even though he is motivated by emotions and desires that are about his own welfare and the welfare of those that depend on him, these emotions and desires can and should be denied because rational reflection reveals that false testimony cannot be universalized without contradiction. As a result, if he is a moral man, the judgment that lying is wrong will allow him to refrain from acting out of his deeply held emotions and desires and instead do his rational duty.[71]

Yet there is no plausible educational process by which such a man would come to do this, even assuming that this is the moral choice in these circumstances. First, the instinct to not act on maxims that are not universalizable cannot be the result of affect that one might attribute to a child (or an animal). We are invited to assume, by Korsgaard, that agents somehow shape their own morality by choosing to self-cultivate in order to become moral people. Yet lacking a desire to be moral that is motivated by empathic sensitivity to the welfare of others, it is mysterious how agents get motivated to even start the process of moral cultivation. A more sensible picture of moral cultivation is that just as animals depend on their caregivers and kin to become more empathic and

[71] Ibid 111

prosocial, we depend on others (primarily our parents, peers, and teachers) to be shaped into more empathic and caring creatures. The motivation to become more caring comes from without, and this is precisely what the Kantian picture of moral self-cultivation cannot easily explain.

According to Korsgaard, since we are capable of refraining from undertaking an action motivated by such emotions, then there is a corresponding sense in which we can be said to have adopted a purpose when we do decide that an action is justified when done for that purpose. However, if the reason we are motivated to develop a sense of justice and be motivated by this sense of justice to refrain from an action is itself the result of emotional states, such as caring concern, compassion, or empathic distress at the thought of the consequences of our actions, then the import of this claim as a reason to deny that animals have morality is significantly undermined. If these empathic emotions lead us to respect others' autonomy and care about their well-being so that we then seek to understand and refrain from violating universal moral principles, then the same kinds of emotional capacities that lead us to develop and use the capacity for reflective judgment are those that lead animals to act in altruistic ways.

Despite this basic fact about motivation for developing and refining judgmental capacities to assess and respect the interests of others, Korsgaard claims that ultimately the reason that we are intentional (and moral) agents in the deepest sense of intentionality is that our purposes, while they might be suggested by emotions and desires, are not determined by them because we can reject an action and its purpose if we judge that acting for that purpose is unjust. Rational agents do not merely have good and bad intentions, but rather assess not only the means to accomplish those goals, but also the

goals themselves.[72] For Korsgaard rational animals are ones that have a level of self-consciousness such that they are aware of their purposes as possible reasons for acting, and as such are not just aware of emotionally presented objects. Rational human beings, for example, are aware of more than just the fact that the objects of their emotions are presented as desirable or undesirable, or as fearful or loved. Rather, reasoning creatures are aware *that* they desire something, *that* they love something, and *that* they fear something. Partly because they have the capacity to be self-consciously aware of their own mental states, reasoning creatures are always capable of asking themselves whether they *should* love, desire, or fear the object of their emotion. It is the capacity to question whether we should love, desire, and so on, that makes humans a kind of being capable of having actions that are distinctly moral or immoral.[73] In Korsgaard's conception of morality, this is the source of our capacity for evil and our capacity for good, whereas animals are simply beyond moral judgments.[74]

To summarize, this argument rules out animal morality in the following way:

1. Acting for purposes that are before your mind, even if those purposes are about the welfare of others, is not sufficient for moral behavior since you are not responsible for acting for those purposes unless you are also be capable of endorsing or rejecting those purposes.

2. In order to be a moral agent capable of moral action, one must be capable of adopting or rejecting purposes that are suggested by emotion and desire.

3. Having this capacity entails that:

---

[72] Ibid 112

[73] It is worth noting again that this capacity does not entail the motivation to constantly ask these self-critical questions. What seems likely to motivate such a capacity is either prudence or caring concern for others, both of which are motivational capacities that we can ascribe to animals.

[74] Ibid 118

a. You are aware of and able to reflect on your actions and purposes and consider whether you should do them, by considering whether they are universalizable without contradiction, and

b. You are able to refrain from your instinctually given emotions and desires, even those related to self-preservation, in order to act only for purposes that are universalizable.

4. The capacity to reflect on your purposes, consider whether they are justified (or universalizable), and to reject or endorse your purposes even if they are suggested by the strongest of instinctual emotions and desires, is a capacity that we do not likely share with other animals.

A lot of pressure can be put on the first premise of this argument, since it cannot be the (constantly and continuously) active reflecting and endorsing of purposes suggested by emotion that qualifies an agent's behavior as morally relevant. Korsgaard must (or at least should) avoid requiring rational reflection about every action before that action is considered moral, since this would mean that certain paradigmatically moral actions, such as acting because one is overtaken by an emotion like empathy or sympathy, would not qualify as moral unless one also simultaneously or beforehand reflected on whether actions for purposes like this are universalizable. Such a stringent requirement for moral behavior would mean, for instance, the helping of a starving child because one compassionately desires to alleviate her suffering is not moral unless one before-hand or simultaneously considered whether one's actions are universalizable or rationally justified.[75] To avoid this implausible result, Korsgaard must hold that it is an

[75] Bernard Williams' well known example of a man who has 'one thought too many' when consulting a universal principle to decide whether it is morally permissible to save his drowning wife instead of a drowning stranger is a similar instance of this difficulty. See his *Moral Luck* (1981) for the full objection. The example has been thought by many to show that pre-reflectively motivated action out of love and caring concern for another is far more ethical than rational deliberation on such dire occasions. Michael Stocker has also given a memorable and much remarked upon argument for the psychological and theoretical schizophrenia between motive and value that occurs when consequentialism, deontology, and egoism leave the friendly or loving relationship and its actors out of moral deliberation. (1997, 66-67) It is in this line of thought that Stocker argues that a person who visits their sick friend in the hospital solely because of their commitment to a utilitarian or deontological principle displays a theoretical and psychical

agent having the *capacity* for rejecting emotionally driven actions and purposes that entails that those actions and purposes are, when not rejected, rightly said to be adopted in the sense that the agent is responsible for those purposes. When a being has such a capacity, then, and only then, are their actions rightly characterized as moral or immoral.

In this case such actions are considered moral because they have not been rejected for rational reasons and because, were the agent to reflect on them, the agent would come to realize they could be universalized without contradiction. This counterfactual account of moral responsibility has some implausible consequences. It is quite strange to suppose that I would be morally responsible for every action whose purpose I do not explicitly accept or reject in virtue of being an agent that could possibly do so. For this leaves open the possibility that I never once reflect on my purposes, and yet somehow act in benevolent and compassionate ways purely as a result of my good upbringing and fortunate circumstances. In such cases, it is being an agent that has an unrealized capacity that makes my actions moral and an animal's actions (with a similar upbringing and good fortune) somehow beyond morality. This makes the capacity for a certain kind of reflection seem either extraneous to moral characterization or else it leaves its role in morality, as well as the motivation for being moral via self-scrutiny, ultimately mysterious. It seems far more plausible to say that the moral status of my action is not determined by a capacity for a certain kind of reflection that I may (or may not) have at

---

split between their (misplaced) reason for acting – the principle – and what they should or do value – their sick friend. (1997, 74) Additionally, Nel Noddings has argued that it is often ethically commendable for a parent to break certain moral principles when their aim is to help the child understand that they are "infinitely more important than a rule." (2003, 53) Thus a utilitarian rule forbidding ice cream before dinner is justifiably occasionally broken because of the parent-child relationship's ethically trumping value. And finally, Michael Slote has drawn attention to the absurdity of a mother who, by maintaining a deontological or consequentialist critical vigilance about all her emotions, deliberates as to whether it is ethical to love her newborn child. (2007, 78-79) I will a similar example later in my own criticisms of the Kantian conception of morality.

some past time (or some future time) have engaged in. Rather, the moral status of my action seems to be entirely determined by the moral content of my motivation.

By way of illustrating how this is an unsatisfactory account of moral action, imagine that I am a reflective agent about everything except whether or not I should be a loving father. Imagine further that I perform actions characteristic of this role, which are, let us grant for the sake of argument, driven by emotions that I cannot possibly reject or deny. Suppose this is so even if the actions and purposes driven by compassionate emotions turned out to be universalizable. Let us call this situation one in which I am reflectively blind to the justification conditions for a particular range of purposes and actions.

Korsgaard's account does not have the resources to explain this situation as one in which my actions as a loving and responsible father are moral, which is strikingly similar to the relational dynamic of many loving families. My reflective blindness as to whether I should act compassionately toward my child either entails that, in this sphere of activity, my actions cannot be moral; or if they are so, it is because of a reflective judgment capacity *entirely blind* (and hence irrelevant) to this sphere of activity. Remember, we are supposing that I cannot help but choose to have such purposes or perform such actions, nor can I reflect on whether I should have those purposes. Why then, would the capacity to accept and reject *other* actions and purposes entail that I am responsible for the actions and purposes that I am reflectively blind to and that I am are driven to perform by emotions that lead me to be a good father? The kind of father that I am is, we are assuming for the sake of argument, completely responsible, loving, and fair to my children, doing everything possible to guarantee their safety, moral development, and

flourishing. The only difference between a Korsgaardian ideal father and me is that I am not a father capable of reflecting on whether I should be motivated to care about my children in these ways.

To see more clearly why reflective self-scrutiny is not necessary for moral motivation, suppose now that I am completely reflectively blind as to whether I should or should not be concerned with any moral norms or virtuous traits, including fairness, justice, compassion, benevolence, and so on. Suppose instead I just happen to be raised so that I am unreflectively preoccupied with acting benevolently, fairly, justly, and compassionately. I can, in this case, only reflectively endorse or reject egoistic purposes. Here again it seems that the capacity to reflectively arbitrate self-interested desires does not thereby make me a moral being, because that capacity is reflectively blind to the sphere of moral concerns. Again, the moral or amoral characterization of my actions according to my capacity for reflection seems entirely arbitrary. As the range of my reflective blindness expands it becomes increasingly implausible that the capacity to reflectively self-govern is what makes a being one capable of moral actions. Finally, we can imagine a scenario suggested by Rowlands, where a being behaves perfectly morally and enjoys doing so, but has no reflective capacity whatsoever. While we might hesitate in assigning praise and blame to such a creature, this is a conceptually distinct question from whether or not the actions of such a being are correctly characterized as moral.[76]

In the case where I am a reflectively blind but nevertheless a loving and responsible father, it is also worth pointing out that if I were able to reflect on whether I should act compassionately towards my child, to the extent that I was constantly

[76] I will elaborate more on Rowlands' example below. It is found in his forthcoming *Can Animals Be Moral* (2012).

preoccupied with determining whether such actions and purposes are universalizable (and therefore justified according to the Kantian conception), we would suspect that I was in fact not the best kind of father. I would be constantly preoccupied with reflecting on whether it is in fact my parental duty to act lovingly to my child as I performed the same set of actions that I would if I were reflectively blind to this sphere of activity and instead purely driven by caring and compassion.[77] Suppose my child were to choose whether I should act in this way or out of genuine care for her well-being in a way that is beyond reflective self-scrutiny. If my child's alternative were to choose that I should act without any emotion whatsoever being the *determining factor of my actions*, and instead act on the basis of (constant and continuous) rational reflection about my parental duty, it takes no stretch of the imagination to guess what type of father my child would prefer. Nor is it likely that a passionate and caring father, even though he is reflectively blind to such purposes, is morally inferior to the rational duty driven parent.

Paradoxically, while it makes sense to say that I am not personally responsible for having compassion that is beyond reflection and rational arbitration towards my child, it is also right to say that my compassionate actions are moral. The counter-intuitiveness of this split between moral responsibility and moral action can be explained away by introducing a subtle distinction, independently noted by both Rowlands and Waller, between two conceptually distinct questions:[78]

1. What determines whether agents should be praised or blamed for their actions, motives, and character?

2. What determines whether actions and behavior are correctly describable as moral?

---

[77] Michael Slote offers a similar argument in his *The Ethics of Care and Empathy* (2007, 78-79).
[78] Similar distinctions can be found in Waller's "What rationality adds to animal morality" pgs. 347-348 and Rowlands' *Can Animals Be Moral* (Forthcoming 2012).

The answer to the second question might not involve critical rational reflection and self-scrutiny of one's purposes, actions, and motivation. However, it is extremely problematic to account for moral motivation without moral responsibility, especially in the absence of critical self-scrutiny. This is because, as Rowlands notes, it seems that if beings cannot control whether or not they are in a morally motivating state via rational reflection and self-scrutiny, then they cannot be held morally responsible for being or not being in a morally motivating state.[79] Yet if they cannot be held morally responsible, in what sense can they be morally motivated? While the moral motivation of the subject and the praise or blame we attach to this motivation (or lack thereof) are conceptually distinct properties, it is hard to see how an agent can be morally motivated if they cannot be praised or blamed as an agent who is morally responsible for their acts and their motivation through their critical scrutiny and rational reflection.[80] Accordingly, I will explain in the next chapter in great detail just how moral normativity, on my view, is possible without moral responsibility.

While whether agents should be praised or blamed may well turn on whether the agent has the capacity for such self-scrutiny, Waller and Rowlands have separately offered different arguments for thinking that, unlike personal responsibility for one's

---

[79] Rowlands, *Can Animals Be Moral?* Chap. 6

[80] The task of explaining how moral motivation is possible in the absence of critical self-scrutiny and praise and blame on the basis of this capacity is undertaken in great detail by Rowlands in his *Can Animals Be Moral.* His tactic is to explain how moral normativity for a *moral subject,* (a being that is sometimes motivated to act for moral reasons) is possible on the basis of moral sensitivity and how moral responsibility is possible through understanding, rather than what he argues is an elusive and problematic notion of control through critical self-scrutiny that is assumed by Kantian and Aristotelian frameworks of moral responsibility. See in particular pgs. 89-93 for the distinction between moral agents and moral patients and Chapters 3,4, and 10 for his full account of moral motivation for moral subjects. I will take a different strategy here, and first argue for the possibility of moral motivation without praise and blame through the counterexamples offered by Rowlands and Waller, respectively, and then give an account of the nature of this moral motivation that does not assume praise, blame, or critical self-scrutiny in the next chapter.

purposes and actions; moral behavior can be characterized *as moral* without explaining it terms of (or assuming) rational reflection. Yet before I explain the reasons supporting such a claim, we must first consider another line of argument that Korsgaard has advanced against the idea that non-reflective agents can act morally that seems to be somewhat sensitive to this distinction between (questions of) moral responsibility and moral action.

In her "Reflections on the Evolution of Morality," Korsgaard argues for the same conclusion in a slightly different way. There she argues that biological accounts of the evolution of morality are seriously deficient as explanatory strategies because it is unclear how they can explain the emergence of normative self-government, or in other words the capacity to be motivated to do something by the thought that you ought to do it.[81]

The argument she offers begins with the assumption that morality is a manifestation of reason, in the sense that moral laws are themselves principles of reason. She notes that this assumption is shared by both Kantians and rational intuitionists,[82] but fails to mention that this assumption is contested and in many cases outright rejected by sentimentalists, care ethicists, and emotivists; to name just a few categories of ethical theories opposed to this assumption.

In any case, Korsgaard is quite right to acknowledge that acting on the basis of a moral principle is different from acting out of altruistic, empathic, or caring emotions. She goes on to claim that to the extent that evolutionary accounts of morality seek to explain the social concerns and interests reflected in behavioral and psychological

[81] Pgs. 2-3 "Reflection on the evolution of morality"
[82] Ibid p. 3-4

dispositions for altruism, empathy, or caring, they are unsatisfactory explanations because they explain something other than morality.[83]

Here it seems Korsgaard is somewhat aware of the difference between the conceptually distinct questions of whether agents are acting out of emotions whose content reflects social well-being, and whether agents are responsible for their purposes and actions that they choose to endorse or reject on the basis of reflection on moral principles. Yet she insists that only behavior from agents who have the latter capacity qualifies as moral behavior because morality, she claims, is only a matter of being able to take responsibility for our character and ourselves by acting under laws that we make for our own conduct. If this conception of what morality entails is correct, this means that morality is not just a relation to others but also one of normative self-government towards oneself.[84]

Now, as Korsgaard notes there are certain historical Sentimentalist attempts, for instance by Hume and Smith, to explain what is added to social instincts in order to account for human morality. For Hume, what was added to social instincts to result in human morality was our approval or disapproval of instincts that were already there. Approval and disapproval required, according to Hume, that one is somehow aware of the motives of others, since motives are the proper targets of moral approval or disapproval.[85] Approval and disapproval also required, for Hume, that agents are able to abstract and reason from an impartial perspective when considering what their actions are

[83] Ibid p. 7
[84] Ibid pp. 3-4
[85] Ibid pp. 8-9

aiming to achieve and whether that goal is good from an impartial perspective.[86] So we come to understand and be motivated by moral norms for Hume when we

1. Sympathize with both the victims and beneficiaries of others from an impartial perspective,
2. Decide on the basis of this sympathetic reflection (when it is performed from an impartial perspective) what we approve or disapprove of, and
3. Are then motivated to perform an action or refrain from an action because we have sympathy with what (we imagine) impartial moral judges would approve or disapprove of.

Because of this sympathetic connection to imagined moral judges we also act according to what we ourselves would approve of and refraining from doing things that we ourselves would disapprove of. It follows that for Hume, as Korsgaard interprets him, the reason to avoid wrongdoing is to avoid self-hate and the disapproval of others, and the reason to do good actions is to attain self-love and the approval of others. She then criticizes Hume's view since it entails that we would avoid doing whatever (we or) others would disapprove of, even if the grounds of their disapproval is not wrongdoing. Morality then inherits an unacceptable social arbitrariness from this Humean sentimentalist account.[87]

I do not wish to defend Hume's view or get into an interpretive debate about Hume's sentimentalism here. In fact, what I am about to say in response to Korsgaard's criticism of Hume departs significantly from what I take Hume's view to be. Instead, I am going to challenge this charge of (implied) arbitrariness, which I think is a little too quick.

---

[86] It is unclear, however, firstly whether anyone is ever able to achieve an impartial perspective free from all the partial motivation of empathy, and secondly how such a perspective would function to morally motivate an organism.

[87] Ibid pp. 8-9

There may be a necessary connection between moral wrongness and what one disproves of from an impartial perspective. Similarly, there may be a necessary connection between and moral rightness and what one approves of from an impartial perspective. Because of these possibilities, there is much more riding on what impartiality amounts to then Korsgaard seems to realize. If impartiality means, for instance, taking everyone's interests into account equally, there may be a way to avoid the implication of cultural or biological arbitrariness that is overlooked by Korsgaard (as well as Hume and Smith, since this departs significantly from their respective views). Assuming it is possible to impartially approve and be motivated by such approval; one example of how impartiality can function objectively in an approval-based theory of morality is *impartially approving of maximizing the satisfaction others' interests*. If actions are approved of because of these blatantly utilitarian concerns rather than arbitrary social norms, then when something is being approved of it is being approved of precisely because it is (objectively) moral rather than socially or culturally favorable. Likewise, if this supplementation of the view is granted, we also have reason to posit that something is being disapproved of from an impartial perspective because it is immoral (in so far as it neglects or harms the interests of others), rather than being disapproved of on the basis of some contingent connection to social disapproval.

There is, however, a more general problem Korsgaard identifies with both Sentimentalist and Darwinian evolutionary accounts of morality. The problem has to do with the fact that, as she characterizes such accounts, they all begin with social instincts and behaviors, and then attempting to add something else to such traits in order to arrive at a morality that is continuous with and conceptually linked to (human) morally

motivating emotions. To see this, consider how Korsgaard argues that despite the differences between the accounts of morality offered by Hume, Smith, and Darwin; all three are unable to properly explain the nature of morality.

While for Hume morality is a matter of approving or disapproving, where these sentiments are forms of approval and disapproval based on the sympathetic connection of the moral judger to the victims or beneficiaries of a moral or immoral action, for Smith approval is simply a form of sympathy, and a lack of sympathy is disapproval. On Korsgaard's reading of Smith, we approve and disapprove of motives on the basis of their propriety (for instance we disapprove of disproportionate anger at a small slight) and their utility (for instance we disapprove of disproportionate anger that leads to a cycle of violent vengeance). We approve when we sympathize with those that are benefited by benevolence because the motive and the action it motivates is useful to the beneficiary, and by sympathizing with the usefulness we also approve of that action.[88]

Yet for Smith we can also judge that feelings of disapproval and approval are proper, or fitting, and when we do so we are judging whether something is blameworthy or praiseworthy. These kinds of judgments are thought of by Smith as given by an internalized other, who is unbiased and unimpeded by our more myopic and selfish desires. We use such an internalized judger to decide whether actions and the motives behind them are blameworthy, hence appearing to solve the problem for Hume that it was the avoidance of self-hate and social disapproval that was the motive for avoiding wrongdoing.[89]

---

[88] Ibid p 10
[89] Ibid p 10

By contrast, Darwin's evolutionary account of the origin of morality holds that there are two competing types of states. These are our appetites, which were short lived, and amoral desires that often override the second type of state, which is a continuous social instinct. When we developed the memory and the capacity for a theory of mind to remember and identify these two types of competing mental states, Darwin maintains, we also came to realize and remember that when our social instincts were overridden by our appetites it was disagreeable to us because a short lived inferior desire was indulged over a more lasting desire. For Darwin it is the awareness of this incongruity that is, or leads to, the moral emotions of remorse or regret.[90]

According to Korsgaard, the common problems that prevent all three of these accounts from being able to explain what is added to social instincts to make moral beings is that they have no resources to explain the objective nature of morality and the proper motivation for moral action. In Darwin's theory, there is no explanation of why the constantly felt instincts are the right ones to act on. Any instincts that were constant and steady in their influence would become authoritative over any instincts whose influence was occurrent and short lived, regardless of the content of those instincts. Similarly, she claims, Smith cannot explain why the motives and responses with which others can sympathize are supposed to be the right ones to act on.[91]

More specifically, the criticisms that Korsgaard is making to all three accounts are as follows:

1. If morality is explained in terms of the unpleasant or painful feelings of disapproval with oneself or others, or displeasure due to the small volatile emotions trumping the steady and persistent social emotions,

[90] Ibid pp 11-13
[91] Ibid pp 13-15

2. Then such an account of morality is too contingently related to moral rightness and wrongness to explain the objective nature of moral norms, and
3. In such accounts, the motivation for the agents is the emotion, which is merely a painful or pleasant mental feeling.
4. It follows that it is left unexplained why we avoid actions that are wrong because we regard them as wrongful, rather than because a certain negative emotion happens to attach to it, and finally,
5. It is also left unexplained why our negative emotions are about wrong actions, rather than any other actions, since given the contingency of these emotions, there is no guarantee that the emotions of self-disproval, displeasure, or one's conflicts of desires are about (or even track) morally wrong actions.

For Korsgaard, whatever makes our actions right or wrong must also be the source of our motivation for performing or refraining from those actions. Whatever non-contingent and non-accidental property does make an action right or wrong must be, Korsgaard claims;

1. What the agent is aware of when they perform a moral act, and

2. What motivates the agent to do that action instead of some intervening mental mechanism like an internal mental pain or feeling of displeasure.

This argument is flawed in many respects. First, since she equates morally laden emotions like sympathy and empathy to a painful or pleasant feeling that is arbitrarily determined by biology, culture, and social favor, she drastically oversimplifies the structure of morally laden emotions and sentiments. The content of an emotional state can be about a complex state of affairs including the historical and current value-laden relations between the self and other agents, even in the absence of capacities for linguistic judgments. It is probably this fact that has led many psychologists and philosophers to endorse some form of cognitivism about emotions, such that emotions are minimally characterized in most theories as intentional (about a state of affairs), value-laden

(expressing the value of that state of affairs), and psychologically dependent on cognitive abilities including memory, categorization, and expectation. For instance, for an animal or human to fear something that has happened before, the organism must remember the state of affairs that led to undesirable circumstances in the past, categorize (or recognize on the basis of some more basic association mechanism) the intentional object of the emotion as the same kind of state of affairs that has happened before, and expect the same undesirable result to occur. Empathic emotions are generally recognized as internally representing emotions or feelings more appropriate to another animal's situation, and are widely regarded to be present in all mammals. As I have argued in earlier chapters, it is plausible to think these emotions represents to an animal the value of another's well-being, and can thereby function to motivate an animal to help another on the basis of a complex altruistic emotional motive.

Partly because of Korsgaard's deficient characterization of emotions and their content, she neglects the possibility in her argument that wrongness can be represented in the content of the emotion. If the emotion was about morally relevant features of the scenario (where precisely what these kinds of features are must, for now, be left vague), then the emotion could motivate us to perform certain actions through its morally laden content, rather than because of a pleasure or pain we experience. Correspondingly, moral wrongness that is expressed, reflected, or tracked by being contained in the content of negative emotions could be both the reason why we refrain from certain actions and what makes those actions wrong.

An emotional state with a certain empathic content, such as the emotional state of fearing another's suffering at our hands by (empathically anticipating the suffering and)

desiring that they not suffer, could express just such a reason. What motivates a person to avoid the action in this case is not just a contingent mental pain, but rather an emotion that is a complex intentional state about morally relevant features of the environment. In this example, empathic distress is not just a mental pain determined by social contingencies of approval and disapproval, but rather a negative emotion about the morally relevant property of the suffering of another that involves feeling (or anticipating) emotions more appropriate to the other's situation. Such an empathic emotion is not only motiving, it is motivating in a directed way such that the action is done for the right reasons. I will have much more to say about the relation between empathic emotions and morality, but for now it will suffice to point out that emotions can track and express moral content directly in a way that Korsgaard has either overlooked or ignored.

This answers Korsgaard's first argument against a Sentimentalist view of animal and human morality as sharing essential properties. Because emotional content can express and track moral wrongness (where the nature of moral wrongness must be left vague at this point save to say it is property that can be expressed and tracked without self-reflective capacities), it does not follow that for a Sentimentalist (and evolutionarily informed view of morality) the motivation for a moral action cannot be the rightness or wrongness of the act. My answer to her second objection, that the wrongness of, for instance, unnecessary suffering is not explained as an objective property by the negative emotion, even with this more accurate characterization of empathic emotions, is more complex. It requires a metaethical account of what moral rightness and wrongness is that explains how certain empathic emotions are conceptually connected to objective

morality, and although I will defend an account along these lines that has been recently put forward by Michael Slote in his *Moral Sentimentalism*, my more immediate critical concern is to dismantle this Kantian picture of morality being a matter of rational reflection and normative self-government that rules out emotions being sufficient for morality in humans and animals.

To this end, consider that there are many theory laden ways of specifying just why it is wrong to cause unnecessary suffering to another being. Such an action might be wrong because it is causes overall aggregate suffering to increase, or because it violates a rule that maximizes happiness, or even because its maxim cannot be universalized without contradiction; to name just a few ways of accounting for the wrongness of such an action. The point is that the property that makes an action wrong (which is potentially specified by whatever turns out to be the correct moral theory) is the unnecessary suffering that is expressed in the negative morally laden emotion we feel towards the action. Such an emotion presents this unnecessary suffering in a painful and motivating way to the agent, and does so independently of however the wrongness of the suffering is ultimately explained by ethical theory. While there is no agreement as to what constitutes a correct moral theory, all of the various competing accounts agree that causing the unnecessary suffering to another is morally wrong. If this property is the target of an emotion such that it motivates an agent to act to cease the unnecessary suffering, then we have a coherent (and non-reflective) explanation about how the wrongness of the situation, and the rightness of intervening, are the emotionally given and motivating reasons for the agent. On my view, these normative qualities are expressed and tracked in

the emotions themselves, and do not require mental acts of self-reflection or abstract rule following.

At this point a Kantian like Korsgaard might be tempted to insist that awareness of the rightness and wrongness of token situations and actions cannot be reducible to an awareness of suffering or flourishing at the hands of others. Rather, awareness of moral properties must be the result of self-reflection on one's motives and accepting or rejecting them on the basis of whether the actions and their ends can be universalized. Only then, Korsgaard might insist, can agents be morally responsible for motives and purposes suggested by their emotions, and only then can they act for moral reasons rather than acting from mere intermediate mental states like emotions.

I think that this general picture of moral motivation can be challenged by its failure to capture a wide variety of spontaneous moral action that is done on the basis of complex and immediate emotional states like compassion and bravery. An example suggested by Waller will suffice to illustrate this point.[92] Saving an infant might be done from a spontaneous emotion of empathic distress, but the emotion can have the factual content that the infant is in danger and the evaluative content that one *should* save the infant.[93] Acts of heroism, compassion, or benevolence, even if they are spontaneous, unreflective, and motivated by our emotions, are quite commonly regarded to be moral activities. This characterization is correct even if such actions are not done for the sake of one's duty as revealed by reflection and abstraction and even if not done with the thought *that* one *should* save the person. Finally, I would argue, this characterization is

---

[92] Waller uses the example of an animal protecting and comforting its young in his "What rationality adds to morality" p. 345

[93] Rowlands suggests the distinction between an emotions factual and evaluative content in his *Can Animals Be Moral?* I will explain this distinction in more detail shortly.

correct even if the agent cannot question whether it is right that one desires to save the person.

Something more needs to be said about how an emotion can track and express moral properties. Fortunately, Rowlands has provided convincing arguments allow us to make sense of the idea of an emotion at least tracking a moral evaluation, even if the subject is not aware of the evaluation in a reflective way.[94]

To understand his arguments to this effect, recall the suggestion that when properly characterized an emotion has both factual and evaluative content. If I believe that you have wronged my friends by yelling at and threatening them, then my emotion has factual content about your actions towards my friends, and evaluative content about the fact that harming my friends is wrong. Yet my anger at your unjustly treating my friends could be misguided, for instance because you were desperately trying to get them to stop harassing your wife. Since yelling and threatening to protect a loved one is not something that anger at injustice is appropriate to, then my moral anger is misguided. My emotion of anger is misguided because it is based on an erroneous assumption; the assumption that the yelling and threatening were unnecessary and undeserved. In general, possessing an emotion, Rowlands argues, entails that there is a certain evaluative proposition (that the emotion is not reducible to, and that the agent need not be aware of) that must be true in order for the emotion not to be misguided. In this way, we can make sense of animals and young children entertaining emotions that track *moral* propositions, when the truth of a moral (evaluative) propositions guarantees, via a reliable asymmetric

[94] The following arguments are derived from similar arguments in Rowlands' forthcoming *Can Animals Be Moral?* In a later chapter, I will argue that an empathy based approach to morality allows us to explain how an emotion can not only track, but also *express* a moral property.

relation,[95] the nonmisguided status of that emotion. In this way Rowlands argues that an organism is morally motivated by an emotion when that emotion tracks a true moral evaluation. More specifically, he states that

> 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, then p is true. (Rowlands *Can Animals Be Moral?* pg.69)

This more nuanced and accurate characterization of emotions thereby provides an explanation of how an emotion can track *moral* evaluations even if the subject is not (or cannot be) reflectively aware of those moral evaluations through a reflective judgment. Of course, more needs to be said about why an agent's emotional state, if they are not aware of its moral tracking properties, qualifies as morally laden. I will merely suggest, at this point, that empathy, or having and being motivated by emotions and feelings more appropriate to another's situation, is both operative in such states and conceptually connected with morality.

In a similar vein, Waller suggests that although a moral act must be one that is done for the right reasons, this does not imply that a moral act must be one that involves reasoning. For Waller, acting for moral reasons often means merely acting on the basis of a moral emotion with the right content. And this can be done without reflective reasoning about the factive and evaluative content of that emotion and whether it is justified from a reflective perspective. Just as we can act upon moral intentions we do not reflectively adopt, so too can animals. The intent to rescue a child, if it comes spontaneously to me out of an unreflective empathic emotional reflex, might not be

[95] The nature of this reliable asymmetric relation is specified on pg 51 in my discussion of Rowlands' account of belief ascription to nonhuman animals.

something I can explain verbally at the moment, or even ever, but having the intent is comparatively simple.[96]

Here again we can see the implausibility of Korsgaard's claim that in order for any agent to have morality they must be able reflect on their actions and purposes and decide, on the basis of whether or not they violate universal moral principles, to endorse or reject such actions and purposes. While in order to be *praised or blamed*, the capacity for reflectively endorsing or rejecting one's purposes and actions may indeed be necessary, the above considerations indicate that moral actions in general do not require that the agent has the requisite reflective capacities to be appropriate targets of praise or blame. I will conclude with two counterexamples that show how explicit attention to the distinction between (reflective) moral responsibility and moral motivation disarms Korsgaard's arguments that the capacity for rational reflection is a necessary requirement for a moral action.

The first example, from Waller, involves imaging a woman called Joyce who acts from the motive of wanting to help the unfortunate and end their suffering.[97] Because Joyce does such acts regularly out of a steady altruistic and caring disposition of character, and fully intends to do such actions, her actions are correctly described as moral. Yet as Waller points out, why Joyce happens to be compassionate, generous and virtuous is a different question. She might be genetically predisposed to be this way, or she might have had a good upbringing, or both. Alternatively, she might have consciously adopted principles of duty upon philosophical reflection, she might have become virtuous through a rigorous Aristotelian program of self-reflective habit forming, or she might

[96] See Waller's "What rationality adds to animal morality" pp. 347-348
[97] This paragraph is a paraphrase of an example Waller gives on pp. 346-348

have been religiously inspired. But the question of whether Joyce is responsible for her character and her motives is conceptually distinct from the question of whether she has a good character and acts from good motives. Whatever the causal story of her virtues and motives, she is correctly described as having a virtuous character. She acts from virtuous motives because she acts from the right kinds of morally laden sentiments that track and express the right reasons for acting.

Our judgments about whether Joyce is responsible for her good character will vary depending on whether she was simply morally lucky in having such virtues or whether she played some active role in developing them. But being responsible for your character, by reflecting on your desires and making reflective decisions that are informed by those higher order reflections about the person you wish to be can only determine whether you are responsible for your moral character. This question of responsibility is different from the question of whether you have a good moral character or are performing moral actions. Moral responsibility over your character and motives may necessarily require higher order rational reflection, but morality does not. (Waller 348) Still, it is highly problematic to explain what motivation amounts to in the absence of higher order scrutiny and endorsement or rejection of moral motives seem to be, in virtue of their normative nature, motives we *should* endorse or reject.

Fortunately, Rowlands has offered a different kind of counterexample to the necessity of rational self-scrutiny and self-correction for an agent to be a moral being that explains moral motivation in terms of moral sensitivity rather than critical self-reflection.[98] To do so, he asks us to consider two (logically) possible people, Myshkin

[98] The following paragraph present a (perhaps unjustly concise) summary of Rowlands' example based arguments in *Can Animals Be Moral?* Chapters 5-7 (forthcoming 2012).

and Marlow, both of whom continuously and dependably act from moral motives. Myshkin performs good acts motivated out of moral sentiments that are appropriate to the situation, and acts out of good reasons. However, those reasons are inaccessible to Myshkin. Due to a unique cognitive limitation, Myshkin has no awareness of the good reasons for his moral behavior. Rowlands characterizes the moral sensitivity that Myshkin has to features of his environment in terms of a moral module, whose processes are inaccessible to Myshkin. More specifically, Myshkin is characterized by the following five properties:

> (M3) (1) Myshkin performs actions that are good, and (2) Myshkin's motivation for performing these actions consists in feelings or sentiments that are the morally appropriate ones to have in the circumstances, and (3) Myshkin has these sentiments and so performs these actions in these circumstances because of the operations of his "moral module," which connects perceptions of the morally salient features of a situation with appropriate emotional responses in a reliable way, and (4) Myshkin is unaware of the operations occurring in his "moral module" and so is (5) unable to critically scrutinize the *deliverances* of this module. (Rowlands, *Can Animals Be Moral?* 146)

Marlow, however, who also performs good actions that are motivated by moral sentiments and also acts for morally good reasons, is able to understand those moral reasons, calmly compare them with alternative reasons for acting and alternative courses of action, and (I will add to Rowlands example) has the ability to decide which purposes are universalizable and which are not and choose to act according to those deliberations. Crucially, they both perform the same acts, with the same (first order) motivation, and for the same reasons.

On what grounds could we justifiably say that Marlow acts morally while Myshkin does not? For someone like Korsgaard, the answer would be that that unlike Myshkin, Marlow can choose to act for certain purposes and deny the moral sentiments

that pull Myshkin one way or the other. In other words, for any motivation, Marlow can have a judgment about that motivation and, if that motivation and the action it motivates are not moral according to this judgment, refrain from doing that action. Now the question arises, what is epistemic status of this judgment about his motivation? Is it something over which Marlow has complete control, that is entirely unaffected by situational, contextual, and social factors? There is no reason to suppose, Rowlands argues, that this judgment is immune to the same kind of emotion driven motivational forces pushing and pulling Myshkin. (Rowlands *Can Animals Be Moral,* pgs.186-187)

To take an example from the situationism literature, suppose Marlow was part of the Milligram experiments, and had to make a judgment between two moral reasons for acting, calmly evaluate two moral sentiments corresponding to those reasons, and decide between two possible actions. Marlow could, on the one hand choose to administer a fake electrical shock that he was convinced would be fatal to another person. He is reflectively aware of the sentiments of trust, loyalty, and the obedience he feels to the authoritative (fake) doctor telling him to administer the shock. He is also reflectively aware of, let us posit, of the moral reason that this is (he falsely believes) an important medical test that will result in a good state of affairs in the long run. These sentiments and reasons are guiding Marlow's choice, but because he can form reflective judgments about his moral reasons, other considerations of which he is reflectively aware are competing with these reasons and sentiments.

The competing moral reasons and sentiments of compassion for the victim, the belief that causing suffering and death is wrong, and the awareness of the moral reason to prevent such suffering, provide competing motivation for Marlow's judgment about

which moral reasons, sentiments, and actions to endorse. Marlow is, at the meta-level of reflective judgment, swaying in wind of competing emotions and reasons that will determine his choice, just as Myshkin is at the mercy of the first order emotions and reasons guiding his actions. We can even imagine that Marlow is aware that administering the shock would be violating the categorical imperative; but then again so would violating his agreement with the doctor to follow all directions in this experimental test.

It is possible, even likely, that Marlow like many of the rational and reflective subjects that actually took place in this kind of experiment would administer the shock because of the surprising and disturbing effect of situational influence on moral judgment, even though he has the capacity and time to make a reflective moral judgment. The point that I want to make here is that second order deliberation, including considerations of the universalization of actions and purposes, are mental states that are subject to the same types of influences that determine first order emotions, motives, and sentiments.

Myshkin and Joyce both have strange causal histories explaining their virtues and emotions, such that they are not (and cannot be) responsible for developing the morally good characters they happen to have. Still, there is no reason to refrain from characterizing their behavior as moral simply because higher order reflection on their motives and character does not (and cannot) occur when they help the sick out of compassion, when they affectionately comfort the helpless, and when they selflessly defend the threatened. While we may not hold them responsible for having a good character, still we are justified in regarding their actions as moral.

But what remains to be said here is something more specific about how emotions and motivations can be characterized as moral without the agent being reflectively responsible for having such sentiments or motivations. We can give clear content to the idea of ascribing morality to an agent to which praise and blame are appropriate, but lacking this, what could distinguish a being that acts morally from one that acts accidentally?

This difficulty can be clearly illustrated with a sports example. There is a difference between a basketball player who, driven by a strong competitive emotion, deliberatively lowers his shoulder and pushes me out of the way in order to get the rebound, and a player who, driven by a different kind of strong emotion, deliberately lowers his shoulder and pushes me in order to permanently injure me. This difference exists even if both actions have the same consequence and both are done in the throes of a passion that does not allow for reflective reasoning. If we cannot explain why the latter action is immoral in terms of reflective reasons, the challenge is to explain how emotions and intent come to be conceptually connected to morality when properly analyzed. This is a very different kind of project than the traditional self-reflective or normative self-government based attempts to account for moral action, but one that is both possible and promising. I will take up this project of explaining the relation between certain emotions and moral motivation in the next chapter.

A final echoing concern from Korsgaard is that some story must be told about how a moral emotion or sentiment can be characterized so as to explain the apparently a priori and objective natural of moral norms. This concern is especially pressing given that the attempts by Hume, Smith, and Darwin all fail to provide an adequate explanation

of the objectiveness and a priori character of moral norms. Here I have been arguing that acting out of moral emotions can be sufficient for moral behavior even in the absence of reflective capacities for normative self-government, but I have not explained how a correct analysis of moral emotions can capture this objective and a priori quality of moral norms.

As I noted at the beginning of this chapter, many of the emotions that I have defended as sufficient for motivating moral behavior, such as sympathy, empathy, caring, compassion, and benevolence, are attributable to animals. My overall project is to defend the view that we are justified in ascribing morally laden emotions to animals, and although I have made significant progress by undermining the assumption that morality requires rational reflective judgment upon one's purposes and actions, or morally laden emotions that essentially involve capacities that it would be implausible to attribute to animals, it remains to argue for a view of moral emotions that accounts for the apparently objective and a priori nature of morality and explains how animals have the capacities for moral motivation.

Chapter 4: What Makes Certain Emotions Moral?

Summary

When (some) social emotions motivate altruistic actions, this is sufficient for apparently moral behavior in humans and animals. However, even if there is a common pattern of emotionally motivated cooperative, altruistic, kind, and helping behaviors that are found in all mammals, this is compatible with some mammals being motivated by mere reflexive emotions, egoistic goals, or partial, contingent, and relativistic emotional values that seem to bear little resemblance to the apparently objective and a priori nature of human morality. Accordingly, in this chapter I give two plausible explanations for how we might be justified in ascribing moral motivation to animals exhibiting such emotionally motivated altruistic behavior.

I begin by presenting some worries about the apparently contingent and relativistic nature of an empathy or altruistic emotion based morality for humans and animals. These worries led Bekoff and Peirce to embrace a species and situation relative morality in the case of animals, and Prinz to defend a relativistic conception of human morality. I will argue that the problems for relativism in the human case motivate the task of explaining how empathic emotions that have relativism implying limitations might nevertheless be part of an objective conception of morality.

In part one of this Chapter I advance an argument suggested by Rowlands that some emotions that motivate mammalian altruistic behavior are correctly classified as moral because they reliably track or correspond to moral propositions that, if the emotion is not to be misguided, must be true. This explanation, I argue, has the advantage of being neutral as to what account of morality is ultimately correct, and can be beneficially

supplemented by the APV account of empathically modulated emotions. By combining the APV account with the tracking conditions suggested by Rowlands, we can explain why we are justified in ascribing moral property tracking emotions whose structure is such that it internally represents the well-being of others as valuable.

In Part Two of this Chapter I show how an empathy based morality might have objective and a priori character for humans and still be ascribable to animals on the level of moral motivation. To accomplish this I first argue for a specific metaethical theory: Michael Slote's empathy based account of morality. When this Sentimentalist account is supplemented with the APV account of empathically modulated basic emotions that track moral propositions, we are able to explain how certain emotions function to morally motivate both animals and humans in a way that avoids making morality unacceptably relative or contingent.

Part One: Is Moral Relativism Entailed by Morally Motivating Empathic Emotions?

For this Chapter, I will make three assumptions that have been extensively defended in earlier Chapters. First, I will assume that some nonhuman animals have emotions, and that I have explained what emotions they are likely to have with as much empirical and philosophical precision as the subject matter will allow. I will also assume that direct arguments from strong cognitivism[99] against the possibility of moral motivation in nonhuman animals are implausible in so far as they are based on an

[99] To remind the reader, strong cognitivism refers to any view of emotions that holds emotions are constituted or solely explained by judgments.

indefensible theory of emotions. Finally I will assume that moral motivation does not require critical self-reflection and normative self-government.

Given these defensible assumptions it is possible to explain why some emotions, especially ones that we are justified in attributing to animals, are correctly *classified* as *moral*. As we have seen, the Sentimentalists and Darwin have offered flawed explanations of the relation between emotions and morality. More recently, Prinz, Rowlands, and Slote have separately defended contemporary views explaining what they take to be the conceptual relation between (apparently moral) emotional states and morality. I will begin this Chapter by explaining Prinz's view, which is somewhat dependent on his flawed conception of emotions as embodied appraisals, and which has unfortunate relativistic consequences. While moral relativism is embraced by theorists like Prinz (in the case of human morality) and Bekoff and Peirce (in the case of nonhuman morality), I will argue that the implications of relativism are reasons to embrace more objective approaches to the relation between emotions and morality. Two particular theories stand out: Mark Rowlands' moral proposition tracking approach that explains in a theory-neutral way how emotions are related to moral propositions, and Michael Slote's empathy based metaethical explanation of why empathic emotions are properly classified as moral. When supplemented by the APV account of empathic emotions, I believe these approaches are compatible and offer two plausible and complimentary explanations for why some animal emotions are correctly classified as moral. Yet first let us take stock of the properties of moral norms that are commonly and pretheoretically thought to characterize human morality.

Moral norms seem to be objective, motivating, impartial, universal, and a priori. They are objective both in the sense that moral claims seem to be true or false and the sense that moral claims are not made true by either one's own preferences or the standards of one's culture. Hence it is wrong to kill innocent Jews, whether you are an enthusiastic Nazi or a morally sensitive American, and what makes that claim true is something objective.

Moral norms also seem to be universal in the sense that a moral reason for one person is a moral reason for any person in that same scenario. If it is wrong for me to steal a car for my own convenience, it is wrong for anyone to steal a car for this reason. Similarly this universal norm ought apply impartially to anyone regardless of the particular and partial personal relationships one has; hence it is wrong to steal a car from *anyone*, rather than it being wrong to steal from just those that are near and dear to me.

Additionally, moral norms seem to be a priori, as it would be strange to think that we must empirically discover and confirm moral claims like the fact that torturing babies is morally wrong, rather than to think that torturing babies is morally wrong independently of any empirical experience. Finally, moral norms also have motivational force, in these sense that if I really understand that cruelty is wrong I am motivated to avoid being cruel and to prevent the cruel actions of others. I offer these characteristic properties of moral norms not as the final say on the nature of morality, but rather as pretheoretic intuitions that informed ethicists have about the nature of morality. A theory of morality ought to be able to explain, or explain away, these pretheoretical intuitions as well as correspond to our common sense moral judgments about obvious cases of moral or immoral behavior. More importantly for our purposes, if animals have morality the

relation between their empathic moral motivation and moral norms with these characteristic properties must be adequately explained.

It will be helpful to see how a theory or morality based in emotions or empathy becomes problematic when it contradicts some of these pretheoretic intuitions. This will motivate the explanation I will ultimately give of nonhuman morality that attempts to explain these intuitions rather than explain them away. To this end, consider that in contrast to many of these pretheoretic intuitions, Prinz holds that sentiments, which are constructed through biocultural interactions, constitute relativistic moral rules that produce moral facts. (Prinz 307) By "biocultural interactions," Prinz means to indicate that evolved biological norms (such as norms governing reciprocal altruism, kindness to kin and group members, rank, and sexual partner selection) are schematic guidelines for human moral norms; and it is from the interaction of flexible constraints with cultural norms that gives rise to human morality. (Prinz 259) Biological guidelines get filled out and modified through, and only through, cultural interactions that rewrite and sometimes override them. (Prinz 286, 246-254) For Prinz, the biologically prepared behavioral dispositions of kindness, fairness, and reciprocity are "culturally malleable and insufficient to guide our behavior without cultural elaboration." (Prinz 277) Culture converts these behaviors, by grounding them in moral emotions, into moral norms. We are emotionally conditioned, for instance, to withdraw love from children who behave badly and this in turn results in those children feeling badly about their bad behavior. (Prinz 271) Culture also takes biologically based norms that are stereotyped and limited (kindness to kin) and alters them into specific norms about what we should do for whom. (Prinz 277) The problem for this view is that it is a contingent fact (that sometimes does

not obtain) about our culture that our biological norms happen to be altered in a way that corresponds with commonsense moral norms.

Prinz in fact embraces this implication of cultural relativism and cites some (widely shared) extra-moral values that might guide us to ever more prosocial moral sentiments in order explain away our commonsense conception of morality as non-relativistic. The extra-moral values he gives include consistency of values, thriving/flourishing, increased social cohesion, increased welfare, having beliefs based on factual knowledge, ease of carrying out moral rules, increased generality, universality, emergence from noble historical circumstances rather than ignoble circumstances, and consistency with premoral biological norms (Prinz 291-292). However, as Prinz himself points out, these commonly shared values have no weight in cultures that do not value them. As he puts it: "If we did not value these things, they would not be seen as advantageous when weighing moral rules." Such shared moral standards vary in importance (and presence) across cultures and times. Bearing this in mind, consider some of Prinz's evidence for descriptive (and, he hopes, also ethical) relativism. (Prinz 209-210)

Cultures have existed (and in some shameful cases continue to exist) where there are valued norms constructed through biocultural interactions allowing for and encouraging slavery, racism, sexism, extreme diets and undernourishment of women, cannibalism, opportunistic rape, genocide, and female genitalia mutilation. The status of such norms, for Prinz, is that they constitute sentiments that produce moral facts. That is, Prinz is committed to a view that holds that it is a moral fact for these cultures that these practices are morally justified. This consequence makes his account of morality and its

relation to sentiments extremely implausible. For on such a view not only are our commonsense judgments about particular cases of clearly moral and immoral behavior contradicted (it is moral to murder, enslave, rape, and so on in these cultures), but morality in general is relative to culture and contingent in the sense that moral claims made true by contingent facts about one's culture rather than being objectively true or false. Moreover, moral norms can no longer be considered universalizable, and so a moral reason for our culture not to enslave a race or genocide a group is not a moral reason for a different culture that allows for and encourages slavery or genocide. Moral norms are also not a priori, but discovered and verified through the process of biocultural interaction for any given individual. In short, except for the motivational character of moral norms Prinz's relativistic theory cannot explain any of the characteristic properties of moral norms. Yet this flaw (from the perspective of our pretheoretic intuitions about moral norms) is not nearly as serious as Prinz's theory's implausible contradiction of our commonsense judgments about clear cases of moral or immoral behavior.

His account does, however, provide an excellent example of how problematic it is to ascribe morality on the basis of biological and cultural values that are represented in emotional concerns motivating apparently moral behavior. If the kind of morality that we ascribe to humans or animals on this basis so little resembles the characteristic properties of moral norms, it is unclear that what we are ascribing is in fact morality rather than merely motivating emotional values that are contingent and relative to biology and culture.

This problem is compounded for the case of animal morality in so far as there is great variation in the biological and species-specific norms that are implicit in animal

emotions motivating apparently moral behavior. As we shall see in the following explication and criticism of Peirce and Bekoff's species-relative account of animal morality, if moral norms are held to relative not only to human culture, but also to the particular species and the group dynamic of particular aggregates of that species, these norms so little resemble our normal conception of human morality that it becomes even more implausible to regard the behavior of animals to be moral than it is to regard Prinz's relativistic conception of human morality to be correct.

The suite of interrelated other-regarding and prosocial behaviors such as altruism, empathy, and cooperation that a variety of nonhuman species exhibit has led Peirce and Bekoff to claim that animals have a "species-relative and situational morality" (Bekoff and Pierce 12, 148). The reason that they claim these examples of empathic behaviors are species relative is that the expressions of prosocial behaviors and the expectations animals have for the behaviors of others in a particular group of a particular species will vary depending on the typical behaviors of members of that species and that typical behaviors of that group within that species. However, embracing this kind of relativism (even if it is intended to be descriptive of animal morality and to not imply anything about human morality as Peirce and Bekoff claim) cannot help to answer the question of whether animals have morality rather than behaviors that merely resemble human moral behaviors. That is, although it is likely that animals exhibit patterns of behavior that "promotes harmonious co-existence by avoiding harm to others and providing others with help," the philosophically interesting question is whether those behaviors are motivated by moral concerns that are legitimately described as moral given our intuitions about

(any) morality being objective, a priori, motivating, and impartial. (Bekoff and Peirce 148)

It is not enough to describe how the apparently moral behavior of animals superficially resembles the moral behavior of humans; we must also offer reasons explaining why we should think that the emotions that motivate this behavior internally represent moral concerns to the organism. In order to do this, we must understand what these moral concerns are in a way that does justice to our pretheoretic intuitions about the characteristic properties of morality. It follows that if we are to say that animals act from moral concerns, those concerns must be more than just species relative and situation relative norms, because we would not regard such norms to be anything like moral norms as we pretheoretically understand them. To accept that morality is relative not only to species but to groups within a species is to embrace an even more problematic relativism than the relativism entailed by Prinz view. There is, however, a way of understanding an animal's moral concerns that allows for these concerns to be conceptually related to objective, universal, impartial, and a priori moral norms. This is the moral proposition tracking approach that has been defended by Rowlands.

Rowlands has argued that certain emotions are morally motivating because they reliably track moral reasons.[100] That is, for Rowlands, an emotion or sentiment is morally laden just in case it has a content that tracks a true moral claim. Ascribing any content to an animal's emotional state is controversial given the notorious arguments of Stitch and Davidson against the legitimacy of ascribing beliefs to animals and the more recent arguments of Chatter and Heyes against the notion of ascribing concepts to

[100] The following is derived from Rowlands forthcoming *Can Animals Be Moral?*

animals.[101] Recall that to address these challenges, Rowlands proposed a theoretical apparatus for relativizing content ascriptions. According to Rowlands, when we ascribe a belief such as "the squirrel is in the tree" to an animal (such as a dog barking at the foot of a tree), we can conceive of this ascription as *tracking* the content of the animal's belief, where tracking is understood as the following relation between propositions:

> (*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**. (Rowlands*, Can Animals Be Moral?* pg. 58)

These propositions are anchored to the context of the human's or animal's interrelated web of beliefs that holistically determine the meaning of the contents of *p* and *p**, respectively; so that when we us the human context-anchored proposition [H:*p*] to explain the behavior an animal, we are essentially providing an explanation of the animal's behavior that bears a conceptual relation to the anchored belief that the animal actually believes: the animal context-anchored proposition [C:*p**]. Importantly, when we do so, we are not ascribing to the animal the linguistic proposition *p*. The actual proposition *p** entertained by the animal nevertheless bears a conceptual relation towards *p*, because when we explain the behavior of an animal by saying that they believe *p*, we are actually ascribing to them *some* anchored belief [C:*p**] whose truth is guaranteed via a reliable asymmetric relation by the truth of the anchored proposition [H:*p*]. As I explained in Chapter One, the reliable asymmetric relation is best captured by universal quantification over the concepts expressed by the subject terms of propositions [H:*p*] and [C:*p**], respectively; where for any subject concept X partially constituting anchored

[101] See Chater and Heyes criticisms in their Animal Concepts: Content and Discontent in Mind and Language (1994), and Davidson (1975, 1982) and Stich (1978).

proposition [H:*p*] and for any subject concept Y that is concept X's de-anchored counterpart and that partially constitutes anchored proposition [C:*p**], if something is an X then it is a Y. (ibid 62-3) This theoretical apparatus allows us to legitimately use anchored human beliefs to explain the behavior of animals as occurring because they entertain some belief [C: *p**] that is anchored to the context of their holistically determined web of beliefs and whose truth is guaranteed by the truth of [H: *p*] via a reliable asymmetric relation.

As we have seen, Rowlands explains the ascription of moral motivation to animals along the same lines, where we can ascribe to an animal a morally motivating emotion just in case it is legitimate to explain their behavior as motivated by an emotion that tracks, via an reliable asymmetrical relation, a true moral (evaluative) proposition. An emotion tracks an evaluative proposition just in case that proposition must be true if the emotion is not to be misguided. An emotion is misguided just in case it is based on an assumption of entitlement that is erroneous. (ibid 68) For example, if I believe that Jones has morally wronged me when he does not lend me money, whereas he is in fact justified in not doing so because I have never returned to him any money that he has lent to me, then my emotion of indignation towards him is misguided because it entails the erroneous evaluative proposition that he has wronged me by refusing to lend me money. This relation allows us to specify how emotions that we can ascribe to animals can be morally laden as follows:

> 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, then p is true. (ibid pg.69)

By making use of this notion of the tracking relation between the contents of human emotions and animal emotions, and the notion of an emotion being misguided when its evaluative content entails an erroneous evaluative proposition, Rowlands provides us with a way of explaining animal behavior as morally motivated through the de-relativized ascription of emotions that track moral propositions. This does not entail, Rowlands is careful to express, that moral emotions are reducible to moral evaluative propositions. Instead, the possession of a moral emotion tracks a moral proposition such that for any emotion there is a particular moral evaluative proposition that must be true in order for the emotion to not be misguided, and there is a reliable asymmetric dependence between the content of the moral emotion and the moral proposition.

If we are to justifiably ascribe a moral emotion such as empathic compassion, then there is a moral proposition such as: "this animal's suffering at the hands of another is immoral," that must be true if this emotion is not to be misguided. That is, the moral emotion of empathic compassion is accurately ascribed just in case the content of that emotion has an accurate factual content (an animal actually is suffering at the hands of another) and the evaluative content is not misguided (it is correct that it is wrong for an animal to make another animal suffer unnecessarily). The content of the moral proposition "this animals suffering at the hands of another is wrong" is a correct moral claim independently of whether or not the animal motivated by empathic compassion is aware of the moral norms that ensure this emotion is not misguided.

Importantly, the tracking relation that guarantees a moral emotion can be legitimately ascribed to a nonhuman animal independently of whatever turns out to be the correct analysis of what makes moral propositions objectively true or false. As long as

there is *something* that makes moral propositions true or false, then an emotion being appropriate or misguided depends on whether the moral proposition it implies is true. Hence Rowlands states:

> Different moral theories, embodying different accounts of well-being, will provide different answers with regard to what grounds these evaluations. The hedonistic utilitarian, for example, will judge that a situation is a good one to the extent it elevates the overall amount of happiness in the world, and bad to the extent that it diminishes overall happiness. For the hedonic utilitarian, elevation and diminution of overall happiness are, accordingly, the respective good and bad-making features. The sort of *capabilities approach* developed by Martha Nussbaum and others will begin with a concept of *flourishing*: of what it is for a creature of a certain type to flourish, and then understand the good- and bad-making features of a situation as ones that, respectively, promote or suppress flourishing. (ibid 223)

However, Rowlands' proposal that there is a way to explain how an emotion tracks moral propositions without being self-reflective or involving complex judgments that the subject of the emotion is aware of, is not (on its own) intended to explain the internally represented values implicit in moral emotions that are crucial to discovering whether or not an animal is morally motivated from an internal perspective. To accomplish this, Rowlands uses the example of Myshkin, a morally sensitive yet non-reflective being intended to represent the possibility of an animal who is morally motivated. Recall that in the previous chapter Myshkin was introduced as a being who was sensitive, in virtue of a moral module (which is not intended to be a psychologically realistic mechanism) to the morally relevant features of his environment, but was not aware of the processes of this module. In Rowlands words:

> Myshkin's sensitivity is directed toward objectively good- and bad-making features of situations. It is this that makes his emotions—the experiential expressions of this sensitivity—the sort of things that can be normatively assessed. That is, his emotional responses to situations are ones that can be judged as correct or incorrect. In effect, Myshkin's

> emotions *track* the good and bad-making features of situations, and this tracking is the sort of thing that can be successful or otherwise … [I]t is an objective moral fact that the features in question make the situations in which they are present good ones or bad ones. The correctness or incorrectness of Myshkin's emotional response is a matter of whether it accords with the moral facts. It is this possibility of accord or discord that underwrites the normative status of Myshkin's motivations. (ibid 228)

Myshkin, who we are to imagine as a place-holder for potentially morally motivated animals, has reliable emotional responses (or sentiments) that track good and bad-making features of the environment, and this tracking is normatively assessable by whether the implied evaluative moral positions of Myshkin's emotions accord with the moral facts. While I endorse Rowlands' use of this tracking relation of normatively assessable emotions (that are the result of a reliable "moral module" that is sensitive to the good and bad-making features of Myshkin's environment), I will argue that the exact nature of the "moral module" mechanism that enables Myshkin moral sensitivity requires further explanation; an explanation that can be provided by the APV account of empathic emotions.

Regarding this mechanism of moral sensitivity, Rowlands writes:

> This normative sensitivity has not come about by accident but is, rather, grounded in the operations of a reliable mechanism or "moral module" that links perception, emotion, and action. The outputs, or *deliverances* , of his "moral module" take the form of emotions or sentiments, and these motivate him to think and act in ways that are (morally) appropriate to the exigencies of the situation. However, having no access to the operations of his "moral module," Myshkin is unable to subject these deliverances of the module to critical scrutiny. (ibid 153)

However, a reliable psychological mechanism of moral sensitivity that Myshkin is reflectively blind to is compatible with Myshkin acting for hedonically egoistic motives such as the avoidance of mental pain (that happens to be caused by witnessing the suffering of others) and the pursuit of mental pleasure (that happens to be caused by

witnessing the flourishing of others), which are states that might be delivered by this moral module. Yet as Rowlands notes, assuming that because Myshkin or the animal moral subjects he is intended to represent act for purely egoist motives when they act to avoid emotional pain (sadness, compassion, etc.) *about* the misfortune of others and emotional pleasure (happiness, benevolence, etc.) *about* the flourishing of others would be a mistake. (ibid 225-7) As long as the pain that an animal feels at the suffering of another animal is felt *because of their awareness and internal representation of that suffering*, we have little reason to think of their motivation as selfishly or hedonically egoistic. Accordingly, an explanation of the nature of (and the evidence we have for) the actual psychological mechanism that results in this non-egoistic intentional structure is needed in order to understand how the moral module-based sensitivity of the animal agents that Myshkin stands for actually internally represent the well-being of another. This further explanation is required precisely because it is possible, though unlikely, that the actual deliverances of whatever psychological mechanisms constitute animals' moral modules are mere emotionally contagious internal pleasures and pains that motivate the animal to respond with apparently moral behavior. As a result, the animals that Myshkin stands for might act only to acquire internal pleasures and avoid internal pains, rather than to promote and protect the well-being of another, as Rowlands maintains must happen (and does happen in the hypothetical case of Myshkin) in order for the agent to be morally sensitive. This is why we must inquire into the nature of the psychologically realistic mechanism of *empathy*, and its effect on the relation between emotions and morality, in order to determine whether not only the deliverances of a moral module can be normatively assessed, but also the *actual animal agents* that are represented by

Myshkin. We must, therefore, examine the evidence for, and the nature of, actual states of *empathic emotions,* and how and why these emotions entail that the well-being of others is internally represented so that it functions as the animal's motivation for moral behavior.

Rowlands' moral proposition tracking approach, when combined with his notion of moral sensitivity via a moral module, does allow us to ascribe to animals emotions that are morally motivating without making morality relative to species or group dynamic. This is a significant contribution to the project of clarifying the evidence and reasons that we have for ascribing animals moral emotions. Whatever makes moral propositions true on this tracking/moral sensitivity approach can be something objective, universal, and a priori, and the fact that moral facts are considered objective is not dependent on any particular moral theory being correct (although it is dependent on some objective moral theory being correct). Still, this approach leaves unexplained the actual psychological process of moral sensitivity that results in the internally represented values implicit in moral emotions, which is crucial to understanding how an animal is morally motivated from an internal perspective. Moreover, because moral sensitivity is explained in terms of a possible agent, Myshkin, whose responses to the morally salient features of a scenario are the result of a "moral module" that is not intended to be psychologically realistic, the question of what actual psychological mechanism allows animals to approximate the abilities of Myshkin still looms large.[102] Rowlands' significant contribution to the moral status of animals is to have argued, successfully in my opinion, that it is *possible* that (some) animals are morally motivated (to the extent they

[102] "I am using the idea of a "moral module" in a very loose, psychologically unrealistic, sense—as a way of designating the mechanisms that connect perceptions of morally salient features of situations with emotional responses to those features." (Rowlands *Can Animals Be Moral* pg. 150)

approximate Myshkin's moral sensitivity). However, we must still inquire into the nature of empathy, which is (I will argue) the psychologically realistic mechanism that enables (some) animals to be morally sensitive, and explore its relation to morality. It still remains, therefore, to determine what the empirical evidence suggests about the nature of the actual mechanism of moral sensitivity in animals and to explain the role of this psychological mechanism in morality. Fortunately, the APV account of emotions that I have argued for, when supplemented by an explanation of the role of empathic modulation in moral motivation, can provide us with a psychologically realistic model of moral sensitivity and inform us of the likely phenomenological structure of morally motivating emotions in a way that rules out (or at least makes entirely implausible) explanations of the apparently moral behavior of animals as performed because of hedonically egoistic (the avoidance of mental pain and the pursuit of mental pleasure) or relativistic motives. The way that the APV account of basic and empathic emotions can supplement the moral sensitivity approach that Rowlands defends begins with the idea that having morally laden emotions can be explained in terms of the degree to which the organism's actions are (or would be) motivated by empathy.

We can formalize this suggestion as the following conditions. In order for an animal or any agent to possess a moral emotion the following minimal cognitive and affective conditions must true of that agent:

1. The organism has a peripheral awareness of locally contributing sensations of physiological states not merely as states of the body, but as *vehicles of value,* or in other words, as contributory representations *indicating* the *psychological value* that the intentional object has to the organism,

2. The organism's global phenomenological state includes a *non-mediated* focal awareness of an intentional object *having the value that is indicated by their peripheral awareness* of sensations of physiological states, and

3. Empathy, or having emotions, thoughts, or feelings more appropriate to the situation of the other organism, is a motivating component of the psychological value that the intentional object has to the organism through the phenomenological experience of a physiological disturbance or involuntary behavior.

According to this account of moral emotions, when an emotion is empathically modulated the value-laden contribution of the physiological state represents the value of the well-being of the other to the agent who entertains the empathic emotional state. The value of the other thereby constitutes a moral reason to act that is internally represented to the organism through the physiological contribution of their bodily changes to the overall phenomenology of their emotion. In the case of compassion, this is a negative empathic emotional state that represents the negative value of the other's suffering as a moral reason to alleviate or prevent that suffering. In the case of benevolence, this is a positive emotional state that represents the value of the other's happiness and constitutes a moral reason to promote this happiness. With this supplementation of the moral sensitivity approach, we can legitimately ascribe to animals emotions that track a moral proposition which must be (objectively) true *and* that have the phenomenological structure of internally representing to the organism the value of another's well-being. When these conditions are met, we are justified in ascribing moral motivation to animals.

This is not proposed as an account of all morally laden emotions, but ones that are both relatively simple (not requiring language or theory based concepts) and attributable to species with the relevant capacities (that meet the conditions of Rowlands' translation

schema). Now we have a way of specifying the phenomenological structure of morally motivating emotions based on the empirical evidence that was systematically laid out in Chapter One: when emotions are empathically modulated and serve as the motivation of an animal's compassionate (in the sense of being moved to alleviate or prevent the suffering of another) or benevolent (in the sense of being moved to promote or contribute to the happiness of another) behavior they are appropriately described as morally motivating. They are morally motivating because they represent the welfare of another internally to the animal as valuable through the contribution of physiological states to the background phenomenology of an empathic emotion, and because the content of their emotion implies a moral proposition that must be objectively true if the emotion is not to be misguided.

Part Two: The Role of Empathy in Moral Motivation

To make this account of moral motivation more plausible, I will now more clearly explicate the concept of empathy and its role in morality. In doing so I will address a central problem for this view of moral motivation, which is that empathic emotions seem to be inherently partial. Predators must often act in ways that display a lack of empathy towards their prey, and even within species there are instances of infanticide, rape, and deadly violence to organisms that fall outside (and occasionally those that fall inside) of the circle of close conspecifics and kin. We seem forced to conclude from such cases either that these animals act from immoral motives or else hold that a kind of species and group relativism is true of animal morality.

However, as I earlier argued, moral praise and blame and moral motivation are conceptually distinct issues. It may be that we correctly regard actions as morally or immorally motivated even if the agent cannot be praised or blamed for having those motives. I will argue that it is indeed plausible to regard animal actions as motivated by moral or immoral concerns on the basis of whether or not those actions display a presence or lack of empathic concern for the other. As an intriguing and seemingly counterintuitive result, some kinds of predation may be classified as immorally motivated (to the extent that an animal acts in a way that expresses a lack of empathy towards its prey), but not as morally blameworthy.

To arrive at this intriguing result, I will first explain and argue for a contemporary sentimentalist account of morality that is based on explaining the nature of morality purely in terms of empathic concern. This will serve three purposes. First, we will have one possible and plausible explanation for what makes moral propositions objectively true or false that we are, via Rowlands tracking apparatus, able to use to ascribe moral emotions to animals. Second, we will be able to state more precisely the relation between human moral agents capable of both moral judgments and morally motivating emotions, and animals who are only capable of the latter. Finally, we will have a psychologically realistic explanation of the mechanism of moral sensitivity that specifies the phenomenological structure of an animal's mental state when they act out of caring concern for the well-being of another.

Michael Slote's has defended a Sentimentalist view that bases morality in empathy which, I will argue, suggests a way that animals might be morally motivated that does not imply moral relativism or require thoughts about moral norms for morally

motivated actions. This is possible because Slote's account explains both *moral meaning* and *moral motivation* in terms of empathy. In order to understand this view, let me first explain what the concept of empathy entails for morality.

Empathy is usually characterized as having feelings or thoughts that are somehow more appropriate to another's situation.[103] (Slote, 17) This kind of mental state can be distinguished from sympathy, which is feeling bad for another, but not necessarily actually feeling what that other person is feeling. (Slote, 5) For instance, I can sympathetically feel bad for someone who is embarrassed after committing a social faux pas, while not myself empathically feeling their embarrassment. However, if I am empathically engaged with that person, then I actually feel empathic embarrassment, or embarrassment more appropriate to their situation or point of view (even though I maintain a sense of identity that separates me from the other).

Empathy can be divided into two varieties. First, empathy may be the result of a cognitive activity that is consciously and deliberatively initiated by the agent. In such cases, empathy is the result of an act of the will. This act is a cognitive choice to try to understand or imagine the viewpoint of another or a group, and the result of this cognitive act is having emotions, thoughts, or feelings more appropriate to that viewpoint or situation. Hoffman and Slote refer to this kind of empathy as projective empathy.

The second type of empathy is referred to as associative empathy. It is not the result of an activity of the will or a cognitive choice, but rather a passive contagion of the emotions of another or a group. Whether the term empathy refers to associative or

[103] See Slote pg 79, footnote 15, where Slote explicitly states he is speaking loosely of empathy, or the definitions given in Hoffman, and Preston and De Waal, respectively, that characterize empathy in general terms. Presumably they do so to capture what is in common between the more cognitive forms of empathy such as perspective taking and the paradigm cases of emotional contagion.

projective empathy, the essential element that makes these mental states empathic is the presence of thoughts, emotions, or feelings in an agent that are more appropriate to another's situation or point of view.

The concept of empathy is a cornerstone of Slote's ethical and metaethical view of the justification and meaning of moral claims. The account of morality that Slote defends holds that both being moral and moral evaluation itself can be explained in terms of having empathic concern for others. In order to understand how Slote is able to use the concept of empathy to explain morality in a way that avoids the threat of species and group relativism, it is necessary to first understand the research Hoffman has conducted and systematized concerning empathic development in children.

For Hoffman, attaining the final stage of empathic development, where a child acquires the full and mature sense of empathy that Slote's view will make use of, involves the child first experiencing what Hoffman calls inductions or encounters of inductive discipline. In these disciplinary encounters the parent (or authority figure) makes the child explicitly aware of the harm he or she has done to another by exercising the child's empathic imagination. So a parent might say "Imagine how sad you would be if someone stole your teddy bear," or "Look how awful you've made Ann feel by calling her names," and thereby induce the child to empathically imagine the other's situation in a way that arouses the emotional displacement and cognitive engrossment that the Sentimentalist tradition (including Slote and myself) holds to be central to moral behavior. (Hoffman 142-44)

This kind of corrective induction is internalized into feelings of guilt and discomfort at another's suffering. When empathic sensitivity is cultivated to the extent

that the child has feelings more appropriate to another's situation involuntarily aroused (and without loss of identity or a merging of personalities), we can say that the fullest sense of empathy is acquired by the child. (Slote *The Ethics of Care and Empathy* 13-15) As I will explain in more detail below, this full sense of empathy includes the ability to empathize with groups and classes of people such as the hungry or the poor by imagining what it is like for an average member of these groups to experience disadvantage. (ibid 29)

On the level of normative ethics Slote argues that we can understand our ordinary moral judgments, including those we make about utility, rights, and autonomy, in terms of fully developed empathy. He argues in *The Ethics of Care and Empathy* that "all, or almost all, the moral distinctions we intuitively or commonsensically want to make can be understood in terms of – or at least correlated with – distinctions of empathy." (ibid 4)

This systematic empathy-based account of the nature of morality that Slote has presented fits in a larger tradition of Sentimentalist ethics, and inherits the advantages of such an approach while avoiding the well-known pitfalls associated with such accounts. Moral Sentimentalism, as general approaches to ethics, is often characterized in critical sometimes dismissive fashion as being partialistic, limited in the range of ethical behavior it can explain, committed to a form of subjectivism or ethical relativism, and unable to explain the (apparent) objective and a priori character of moral norms. Yet before I explain why I think Slote's account overcomes all of these characteristic difficulties of Sentimentalism, let me first explain why I find Slote's view advantageous over other approaches to morality. After independently motivated this account and explaining how it is able to account for morality in a way that does justice to our

pretheoretic intuitions about the a priori, objective, and universal character of moral norms, I will show how this account also entails that animals are morally motivated in virtue of their empathic emotions.

Acting from caring concern is something we commonsensically regard to be morally appropriate in close relationships and situations calling for immediate and unreflective action. Often, these kinds of caring actions are more appropriate to such situations than acting out of a reflective understanding of universalizable moral rules. Consider, for instance, Bernard Williams' well known example of a man who has 'one thought too many' when consulting a universal principle to decide whether it is morally permissible to save his drowning wife instead of a drowning stranger. (Williams 1981) This example has been thought by many to show that in some situations acting with the pre-reflective motive of caring concern for another is morally appropriate, and that acting after rational deliberation with the motive of following universalizable rules is, in such situations, morally inappropriate. Michael Stocker has similarly argued that a person who visits their sick friend in the hospital solely because of their commitment to a utilitarian or deontological principle displays a theoretical and psychical split between their (misplaced) reason for acting – the principle – and what they should or do value – their sick friend. (Stocker 74) Waller gives an additional example of this kind of moral motivation in close and personal relationships when he argues that a mother who risks her life to save her child, and does so spontaneously and without hesitation or deliberation,

seems to be more virtuous than a mother who performed the same action only after deliberative consideration of their ethical duty. (Waller 344)[104]

These examples have a logical structure in common. This structure is that in instances requiring immediate reactions, particularly in close and personal relationships, *reflective moral reasoning* is not necessary for moral *action*. The extent to which self-cultivation through reflection on moral principles is required to develop psychological dispositions to have moral emotions is another issue. As I previously argued, this might explain why we morally praise the agent for developing the character they have that allows them to act with caring instincts, but self-cultivation is not necessary for characterizing actions as moral or immoral. This is because the causal history of the agent's character is irrelevant to explaining why the occurrent action should be characterized as moral. Rather, a morally laden emotional reaction seems to be sufficient for moral action (an action we morally approve of because it is motivated to promote the well-being of another and/or alleviate the suffering of another). The more natural explanation of the moral character of these kinds of actions (since they do not involve reflection on principles, and need not involve even the capacity to do so) is to hold that they are moral because the emotional motives for those actions reflect or exhibit caring concern. This is the first advantage of the Sentimentalist approach to accounting for morality in terms of (dispositions to have) morally laden emotions.

Michael Slote's empathy based approach inherits the general advantage of Sentimentalism by being able to explain moral actions in these close and personal circumstances as being moral in character on the basis of the empathic emotional concern

[104] Michael Slote has also drawn attention to the inappropriateness of a mother who, by maintaining a deontological or consequentialist critical vigilance about all her emotions, deliberates as to whether it is ethical to love her newborn child. (2007, 78-79)

that functions as the motive for these actions. However, it also able to avoid certain difficulties that Sentimentalist accounts notoriously share. To this end, consider that the immediate needs of another are what are being empathically engaged with in these examples in a way that might seem to exclude considering the interests of others beyond the person who is the direct target of empathy. This creates a problem for Sentimentalist approaches to morality.

This problem is that any ethics based on empathic moral sentiments must explain, or explain away, the partialistic quality of many empathic emotions. For instance, while it is empathic and ethical that a husband preferentially save his wife instead of a stranger, it also seems to be empathic and unethical if a husband preferentially favor his wife over a more qualified applicant for a job position. Worse still, basing morality on empathic emotions seems to lead to moral absurdity when applied without qualification to unjust associations and groups. Agents within unjust groups may well display empathic concern for members within the group, and yet willfully propagate the violent trampling of outsiders' rights. The existence and clear immorality of these organizations (persecutory religious movements, racist organizations, criminal fraternities, etc.) is a clear problem for any ethical approach that accepts or implies a form of partialism. The problem is exacerbated for animals who act empathically for members of their species and group, but are utterly oblivious to (and sometimes seem to enjoy) the suffering of other species or group members. When predator animals hunt, there are good evolutionary reasons and obvious behavioral evidence for the thesis that predators do not have an empathic concern for the well-being of their prey. Such behaviors are undoubtedly what led Bekoff and Peirce to hold that animals have 'species-relative' morality. However, as

Prinz's flawed view of morality demonstrated, we would consider even *humans* to be acting morally if they only exhibited empathy within a very narrow range of organisms (family members, race members, etc.) and exhibited outright cruel behaviors to other humans who fell outside of their group. A relativistic and partial morality is no morality at all.

To illustrate this difficulty more clearly, consider how the notable care-ethicists Nel Noddings grapples with these issues by maintaining that "a certain Ms. A, who sides with her racist family members, despite her moral disagreements with them, cannot be criticized from the care perspective. (Noddings 109-112 in Herr 475) Of course, Noddings had much more to say about such cases, condoning a of breaking ties with the racist family members when their damage to the "ethical ideal" estranges Ms. A from them or when they harm another "cared about" individual so much that this weaker (less partial) connection to the harmed person grows and supersedes the diminished caring relation Ms. A has to her family members. (ibid, 110) Still, we must consider whether it is really practically possible for agents like Ms. A, agents who are completely engrossed with unjust individuals, to be empathically sensitive to those that the unjust individuals hate and hurt.

Slote offers a solution to the relativism and partialism problem for Sentimentalism by arguing that empathic concern can explain, or at least correlate with, our judgments about the clear immorality of the partialism of unjust groups. To this effect, Slote asks us to consider whether the "neo-Natzi march in Skokie, Illinois (which never actually occurred) should have been permitted, despite all the suffering it would or could have caused to the Holocaust survivors who lived in that town." (2007, 68) Imagine a situation

in which this march was about to occur, but the neo-Natzi leader's granddaughter – we can call her Zoey (for brevity) – hides his false teeth because she knows this small deception will make her grandfather too embarrassed to go out in public. If we assume Zoey knows this action will cause the whole demonstration to fall apart, is her action morally permissible? Kantian liberalism, Slote argues, is forced to uphold the so called sanctity of the human right for autonomy, and deny that it is morally permissible for Zoey to interfere with her grandfather's right to make his own decisions.

By contrast, we might say that consequentialist theories would superficially agree with care ethics in holding that Zoey's act is morally permissible, but for the wrong reasons. That is, while Kantian liberalism will disapprove of Zoey's act because of its failure to respect and act in accordance with the right of autonomy, consequentialism will focus on the psychological harm, abstracted from any particular agent in which this harm is located, that the grandfather's act would cause. Yet this focus fails to give the grandfather, and the value he places on his right for autonomy, and the value of the relationship between Zoey and her grandfather, the appropriate weight in the moral status of the Zoey's act. Consequentialism and deontology similarly fail in a more ethically profound way by holding that the only determining factor in the morality of the act is whether it accords with abstract principles or prevents harm. From a care ethical perspective, it is not autonomy, but the grandfather's valuing his autonomy, and not harm, but the harm it would cause any particular one of the holocaust survivors, that are the factors which should enter in the determination of the ethicality of the act. These factors can be explained in terms of the proper empathic reaction of an agent that is sensitive to and weighs up the harm they would cause to the Grandfather by interfering

with his autonomy and the harm the Grandfather's actions would cause any particular holocaust survivor. Zoey acts rightly by hiding her Grandfather's false teeth because she respects her grandfather's autonomy out of an empathic understanding of that basic human good and its value to him, and yet still decides to interfere for the sake of the Holocaust survivors because of her overriding empathic discomfort at the psychological harm her Grandfather would cause an average member of this group.

This example also nicely illustrates the potential spatiotemporal and cultural flexibility of fully developed empathic concern. Zoey has (or could have) no direct contact with any particular Holocaust member, but instead imagines what it would be like for an average member, and then has an empathic reaction to the harm it would cause that imaginatively constructed member. It is on the basis of this cognitive feat that Zoey acts most ethically. Importantly, we are still able to say in this case that Zoey acts out of direct empathic concern for a cared for (imagined) other. It follows that close and personal contact and idiosyncratic knowledge of the other are not in fact necessary conditions for caring or empathic behavior.

The key to explaining why it is unethical to treat those outside of your sphere of care in ways that neglect their psychological well-being is to understand moral motivation in terms of a *fully developed* sense of empathy that takes the interests of all involved into account (though not necessarily equally). This full sense of empathy includes the ability to empathize with groups and classes of people such as the hungry or the poor by imagining what it is like for an average member of these groups to experience disadvantage. (Slote 29) Empathy can function in a way that transcends the partialism that is implied if we think of empathy only as emotional contagion between members of a

cared for group. It follows that a Sentimentalist morality explained in terms of empathic emotions does not entail that morality is relative to or limited by the circle of cared for others; whether that be one's family, culture, or species. A fully developed sense of empathy is one that is sensitive to the well-being of any others whose interests might be neglected and harmed rather than being restricted to emotional contagion in close and personal relationships.

While this fully developed sense of empathy capable of empathizing with groups is a capacity that most, if not all, nonhuman animals will lack, this does not preclude them from acting from moral or immoral motives. This is because having moral or immoral motives only requires acting in a way that exhibits or reflects the presence of empathic concern or a lack of empathic concern for those who are affected by their actions. We can judge their actions from the standpoint of theorists as approximating what a fully empathic agent would do without requiring that they themselves are fully empathic in order to be morally motivated. To the extent that their motivating emotions represent an empathic concern for the well-being of another, they act morally.

An interesting problem arises for my view if we imagine a scenario where an animal acts empathically towards his sister in a way that neglects the interests and emotions of the rest of his family. What seems right to say in such cases is that to the extent that the animal harms others by neglecting their interests, they do not act morally towards those who are empathically neglected (and are perhaps immorally motivated, although they should not be praised or blamed for this motivation). All animals (including humans) will have natural limitations to the extent to which they can empathize. These limitations determine the sphere of social concern within which they

can act empathically given their ability to empathize with increasingly diverse groups and organisms. Because we can understand morality in terms of what a fully empathic individual can do, we can explain why these limitations can result in non-moral or immorally motivated actions towards others outside the sphere of social concern. Just as we can say that the Grandfather is not acting morally by neglecting the emotions and interests of Holocaust survivors in the Skokie example, we also say that an ape who neglects kin, conspecifics, or other organisms that fall outside of his sphere of social concern is not acting morally towards those who are neglected by his actions. We should, however, regard the kind actions of the Grandfather towards his daughter and the kind actions of the ape towards his sister as empathic and morally motivated when those actions do not also exhibit a lack of empathy towards others.

Because we already have ample reason to think that animals are capable of such emotions, we are justified in ascribing moral motivation to animals. I will now further motivate this Sentimentalist account by explaining the arguments Slote offers for holding that empathic states are not only sufficient for moral motivation in humans (and by implication animals), but also necessary for understanding of the meaning of moral terms. That is, I will now explain Slote's arguments for how empathy functions, in a way to be specified, not only in moral motivation in humans and animals, but also in the very meaning of moral terms in a way that does justice to our conception of moral norms as being a priori and objective.

The central claim Slote makes in *Moral Sentimentalism,* where he advances these arguments, is that the "warm" feeling of empathy that we feel towards an agent who is acting empathically or kindly and warmly towards another being constitutes moral

approval, whereas the "chilling" feeling of empathy with an agent's cold-heartedness when they act cruelly (showing a lack of empathy) towards another constitutes moral disapproval. The way in which empathy is essential to understanding moral claims and moral judgments is that the warm or cold experience of empathy with empathy (or a lack of empathy) serves to fix the reference of moral terms and moral concepts like right, good, bad, and wrong. Slote holds that the reference of such moral concepts and terms is fixed in a way analogous to the reference fixing theory of the meaning of color terms that Kripke has proposed.

Very roughly, for Kripke the reference of color terms like "red" (or any terms referring to whatever it is that causes the color perceiving experience) is fixed by the experience of redness, which enables us to understand an essential element of what redness is and to refer, in thought and language, to whatever it is that causes that experience of seeing a certain color. Even though we discover a posteriori the properties that cause our experience of redness, the experience of redness enters into and is essential to the understanding of what redness is, such that it is "an a priori truth that (objective) red(ness) is whatever causes or has tended to cause, and is perceived by means of, our visual experience of red(ness)." (Slote, *Moral Sentimentalism,* 58-59)

Slote then argues that just as you cannot fully understand the color word 'red,' without having a red experience by which you fix the reference for yourself; so too you cannot fully understand moral terms without having second order empathy. (ibid 59) The mental states of sincere moral approval or disapproval that ground our understanding of the terms right and wrong are mental states enabled by and only by experiences of second order empathy. These moral states are literally constituted by warm, or cold empathic

reactions with empathic agents (or agents performing actions that exhibit a lack of empathy). To avoid the implication that it is a posteriori that caring is right just as it is a posterior that red objects reflect light of a certain wavelength, Slote argues that it is more than the subjective feeling of warmth fixes the reference of moral terms. Rather, a recognition or awareness of empathy is implicit in the feeling of warmth delivered by the empathic mechanisms, so that it is a priori that the reference of the concept "right" and its like are fixed by the feelings of warmth (or coldness) that is directed at agents. In this way it is a priori that goodness or rightness is whatever causes us to be warmed by the warmth and tenderness of agents who are acting empathically. (ibid 61) We feel moral approval by feeling empathically warmed by the actions of agents that themselves display or exhibit empathic motivation, and in this way approve of such agents as morally motivated. The causal relation between empathic agents and the welfare of the others that is represented internally to such agents are therefore actually part of the complex second order empathic state of an approving spectator that rigidly fixes the meaning of right and wrong. (ibid 67)

This, in any case, is the Sentimentalist picture of moral meaning that Slote argues for in great detail in *Moral Sentimentalism*. Importantly, for such an account we can explain the moral character of first order empathy, or the empathic motives that we correctly feel moral approval towards, as essentially moral whether or not the agent exhibiting first order empathy themselves feels (or is capable of feeling) moral approval or moral disapproval. It follows that, on this account, animals have moral motivation to the extent that their actions exhibit or reflect empathic concern that we correctly morally approve of by having empathy with their empathic motivations. I will have more to say

about the moral nature of animal empathic behavior below. Before I do so, however, let us examine some reasons that make this metaethical picture of moral meaning and moral motivation independently plausible.

The primary reason to accept this Sentimentalist view is it explains both why we regard certain emotional states to be morally motivating in the absence of moral judgment, and why when we do make moral judgments they seem to be objective, a priori, and motivating. While the prima facie impartial character of moral terms is explained away by this view, the example-based arguments against the morality of impartiality in close and personal relationship offered above support this move. We also should note that we can make sense of why it is unjust in some situations to partially favor those near and dear (such as giving a job to your under qualified wife over a qualified stranger) in so far as doing so does not exhibit or reflect the presence of a *fully developed* sense of empathy. Moral actions are, on this view, simply those that are motivated by empathic concern for the welfare of the other, and immoral actions are those that exhibit or express a lack of empathy.[105] Empathic emotional concern with the other is all that is required for moral motivation, and this does not require the second order empathy that is necessary for understanding the meaning of moral terms. Moral judgments which do require such an understanding are motivating in so far as the coldness of disproval draws us away from doing such acts, and the warmth of approval encourages us to do such acts. Yet it is objective and a priori that empathy with empathy fixes the meaning of moral concepts and terms. Additionally, the fact that empathic caring/concern for others corresponds to our ordinary moral judgments (as Slote argues in

[105] Hence predation is intriguingly immorally motivated but not morally blameworthy, which is a conceptually distinct issue involving, among other things, understanding one's motives as moral or immoral through second order reflection and critical self-scrutiny.

The Ethics of Care and Empathy) suggests that empathy enters into our understanding of and making of moral claims. (Slote 52) Finally, this account can explain "how and why psychopaths cannot make or full understand moral judgments." (Slote 54) They cannot do so because both moral motivation and the meaning of moral terms are inaccessible to organisms that lack empathy or the associative empathy required for being warmed by empathic warmth and chilled by cold-heartedness.

However, I believe that, like Rowlands' moral sensitivity model, Slote's account of moral motivation and moral meaning is subject to a worry that can be overcome by supplementing his account of with the APV account of empathic emotions. The worry originates in the ambiguity inherent in using to the term "empathy" to refer to both cognitive empathy and emotional contagion. Recall that empathic identification doesn't involve a felt loss of identity, but it does involve feelings or thoughts that are in some sense more "appropriate" to the situation of the person(s) empathized with than to the situation of the person empathizing. In general, as we become more aware of the future or hypothetical results of actions and events in the world, we learn to empathize not just with what people are actually feeling but with what they will feel or what they would feel if we did certain things of if certain things happened. It seems we even learn to empathize with their (situated) condition and not just with their hypothetical or actual reactions to it.

Slote maintains that there is a difference between empathy and sympathy, and that it is empathy that is relevant to making moral judgments and understanding moral language. He states: "empathy involves having the feelings of another (involuntarily) aroused in ourselves, as when we see another person in pain," whereas, "we can also

simply feel sorry for, bad for, the person who is pain and positively wish them well, and that is what we mean by sympathy." (Slote *Moral Sentimentalism*, 15) Following this rough characterization he defines empathy somewhat more precisely as involving feelings or thoughts that are in some sense more appropriate "to the situation of the person(s) empathized with than the situation of the person empathizing." (ibid 17) For example, a doctor can empathize with a hospital patient who is unaware of their terminal condition and currently happy on the basis of what they would feel (or will feel) if they had this knowledge. Yet suppose that unbeknownst to the doctor in this example, the terminally ill patient has an unusual fetish. Instead of dreading their death, they are excited by death (perhaps the mystery of what comes next, oblivion or some kind of afterlife, draws them with almost irresistible force). In fact, they have actually been longing for a terminal illness of some sort. While desiring one's death in this way is, of course, quite an inappropriate emotion to have, nevertheless the patient will be even more overjoyed then they currently are when they hear the normally dreadful news.

Here I think a telling puzzle arises regarding the accuracy conditions for empathy. It seems that in one sense empathy is successful, since the doctor is empathizing with what is appropriate to feel in these circumstances (the death fetish is as inappropriate as you can get!). Yet in another sense there is no contagion, or transfer of emotions, or involuntary arousal of the agent's emotions occurring. There is, in this sense, an utter failure of empathy on the part of the doctor. Yet if the doctor is empathizing with the situation of the patient but not with the patient, is empathy actually occurring, or is this rather a failure of empathy; perhaps a case of mere sympathy? Notice that if empathy with the potential or future feelings of the patient were to occur, where the doctor

cheerfully gives the news that the patient will soon die and rejoices with them, then an observing agent would probably feel the chill of moral disapproval.

If Slote insists that it is the situation that is properly said to be empathized with in the sense relevant to morality, and so empathy does occur, a second kind of counterexample threatens the view. If we turn from this farfetched example to cases closer to the interests of the feminist tradition to which Slote belongs, empathizing with the situation comes dangerously close to reflecting a lack of empathic understanding. For instance, consider the following example. A successful business woman feels bad when she reports her success to her less successful husband because of a culturally instilled prejudice that has been inculcated in her that women should not be more successful than their husbands in the workplace. If her husband can be said to be empathizing when he merely empathizes with her situation, without regard to her actual feelings, then empathy turns out have nothing to do with the actual feelings of the business woman.

Slote may reply to these worries by claiming that having empathy in both these scenarios requires that the agent do what a maximally empathic individual would do, who knows all the relevant psychological states of the agent(s) and the causes of those states, as well as what emotions are appropriate to the relevant circumstances (whether or not the emotional reactions of the agents is unusual). However, this move seems to be both circular (arguing that a notion of full empathy can give an answer to situations where there doesn't seem to be any way to decide whether empathy was successful or not is merely to assume that empathy can explain what it apparently cannot) and building in a fair amount of prior judgments, including the appropriateness of emotions, and the

correct way to empathize given this criteria of appropriateness, and the inappropriateness of the actual emotions.

When we imagine what it is like for someone else to be in a situation like that of the death-fetish patient or the successful business woman, the intentional object of our empathic deliberation is a fictional or merely possible imagined being (either ourselves in the person's situation or how we imagine it might be for the person) that is not identical to the actual agent, but used rather as a means to try to understand the agent that we value. When such an act of imagining occurs, the targeted person and the entity that is actually empathized with (the fictional or possible imagined agent) come apart. The only thing you are directly empathizing with in such cases is a fictional or merely possible entity or state of affairs. You then use this act of empathic imagining to understand the targeted agent's point of view, which is importantly different from direct emotional contagion.

If we supplement Slote's Sentimentalism with the APV account of emotions, then we can explain why empathy seems to fail in these cases.[106] In cases where a fictional or imagined person is empathically identified with, we merely have a case where the empathic emotional state is representing a hypothetical psychological value (of the target of empathy) to the agent through the background contribution of physiological states as they imagine what it would be like to be in the others' circumstances. In the same way fearing imagining creatures represents a value that would be accurate (and appropriate) if the object of fear were actual, empathically identifying with an imagined individual

[106] We can also have an explanation of how the feeling of warmth reflects the value of the well-being of another that is internally represented by the empathic agent who is target of these warm feelings. The feeling of warmth actually expresses to the agent a value, according to APV. It does so by being a bodily contribution to phenomenology that is taken by the agent as expressing the value of the intentional object – in this case, an empathically motivated agent.

presents the emotional values of that individual to the empathic agent as emotional values that would be empathically engaged with if the individual and their emotional values are accurately imagined. This nicely resolves the question of whether empathy succeeds or fails in the death fetish and business women counterexamples. In one sense, a kind of imaginative empathy does succeed; it simply succeeds with an inaccurate representation of relevant psychological values. The empathic agents in these cases (the doctor and the husband) have empathy with the imagined reaction of the targeted agent given the context and the typical and appropriate reaction to that context, which, in these examples, does not actually correspond to how the targeted agent actually feels. In another sense, empathy with the agents themselves fails, because they are not accurately represented to the agent that feels empathy with an imagined intentional object rather than the person themselves.

The APV account thereby allows us to explain how the psychological values that are internally represented in empathic emotional states might fail to correspond to the actual psychological values of the target of empathy. This supplementation also allows us to understand why we might feel empathically chilled at the doctor's reaction to their death fetish patient when the doctor *does* empathically engage with the patient and happily relates the news of the patient's terminal illness. When we feel a chill at the doctor's unusual reaction, we are not accurately representing the patient and the doctor's emotional values, but rather disapproving of what we imagine a doctor should feel about a typical patient in this context. Similarly, if we warmly approve of the husband's reaction that displays a lack of empathy with his wife, we are not doing so with an accurate picture of the way that the husband's empathy is conveying to him

psychological values that the wife does not in fact hold. Introducing the APV account thereby allows us to formulate these accuracy conditions for the values that empathic emotions convey so that we can resolve these paradoxes of empathic identification without introducing circularity or prior judgments.

With this supplementation of Slote's empathy based view of moral motivation and moral meaning in place, let us turn back to explaining how animals act morally when motivated by empathic emotions. I have offered the following formula for explaining how an animal can act for the sake of another's welfare and internally represent the other's welfare as valuable.

This formula is that in order for an animal or any agent to possess a moral emotion the following minimal cognitive and affective conditions must true of that agent:

1. The organism has a peripheral awareness of locally contributing sensations of physiological states not merely as states of the body, but as *vehicles of value,* or in other words, as contributory representations *indicating* the *psychological value* that the intentional object has to the organism, and

2. The organism's global phenomenological state includes a *non-mediated* focal awareness of an intentional object *having the value that is indicated by their peripheral awareness* of sensations of physiological states.

3. Empathy, or having emotions, thoughts, or feelings more *appropriate* to the situation of the other organism, is the (motivating) component of the psychological value that the intentional object has to the organism through the phenomenological experience of a physiological disturbance or involuntary behavior.

Using Slote's Sentimentalist and empathy based view of morality, I have argued that the problem of an animal's empathic states being partialistic and relative to their species and group does not mean that they are not morally motivated when acting on the

basis of empathic concern, since we can understand their actions as moral to the extent that they are motivated by empathy. This is possible because the objective, motivating, and a priori character of morality can be explained in terms of the actions and motivations of a fully empathic agent with all the relevant information about the psychological welfare of others. I have further argued that moral approval, or empathy with empathy, is properly felt of organisms whose empathic capacities are limited to partialistic concerns, as long as their actions do not exhibit a lack of empathy for other groups or organisms. Since there is a way to understand the objective and a priori nature of moral norms that grounds moral meaning and moral judgment in the empathic nature of moral approval and disapproval, there is also a way to understand why animals can act morally without implying that morality is relative to a species or group. To the extent that animals internally represent the interest of others via empathic mechanisms, and do not show a lack of empathy, they are morally motivated. Being morally motivated in this sense is an objective matter of acting in a way that we correctly feel moral approval towards.

Animals act morally just in case their physiological state's contribution to phenomenology represents the welfare of another as valuable and as a motive for acting, which occurs through emphatic mechanisms. We have convergent evidence for the presence of basic emotions and empathic capacities that develop in sophistication over time in at least all mammals, and so we are justified in concluding that at least all mammals can be morally (and immorally) motivated. Importantly, this conclusion does not depend on the particular account of moral sentimentalism argued for here being the correct account of morality. This is because the presence of empathically modulated

emotions that track moral propositions, that internally represent concern for the well-being of others, and that motivate organisms to act on behalf of others, is something that any theory of morality must take into account as prima facie evidence for moral motivation. I have defended Slote's Sentimentalist account here as an illustration of how the capacity for empathy can be used as the basis for an explanation of objective, universal, and a priori moral norms. It is one possible systematic explanation for how an animal's internal representation of the concerns of others is morally motivating, but it is not the only possible explanation. The tracking relation proposed by Rowlands, when supplemented by the APV account of empathic emotions, ensures that whatever account of morality turns out to be correct, we are justified in ascribing animals moral motivation. However, Slote's empathy based sentimentalism is compatible with and complementary to this tracking approach, as it explains what makes the moral propositions that animal emotions track objectively true in a way that does justice to our pretheoretic intuitions about moral norms while still explaining moral motivation in terms of empathic capacities that we share with animals.

Conclusion

Summary

I conclude by summarizing the arguments given in this dissertation against those who have given reasons to deny the possibility of animal emotions or animal morality and separating these critical arguments from the reasons I have for my positive account of moral motivation in animals. Then I argue that even if my positive account of animal emotions or animal moral motivation should turn out to be incorrect, the critical arguments I have advanced establish that there is no principled reason to deny that animals are morally motivated when they behave in ways that resemble human compassionate or benevolent actions. A further question that I address in this conclusion concerns the extent to which animals are, in virtue of their empathic capacities, moral beings that can not only be morally motivated but that also have a rudimentary understanding of moral norms. I examine this possibility and its empirical implications and then show how the positive account of moral emotions defended in this dissertation suggests some intriguing ethical implications for our responsibilities to animals.

In this dissertation I argued for the thesis that animals can be motivated by moral emotions. To do so, I first addressed some skeptical concerns implying that animals lacked the capacities for certain necessary elements of having morally laden emotional states. Then, after examining the evidence for animal emotions and what that evidence indicated about the likely phenomenological structure of emotions in animals, I defended a theory called the Awareness of Physiological Vehicle (APV) account of emotions that explained what animal emotions are and why we are justified in ascribing emotions of the

same kind to both humans and animals. Following this, I explained why certain emotions of *both* humans and non-human animals are morally laden by first refuting the skeptical doubts of Dixon and Korsgaard and then defending a theory of ascription conditions suggested by Rowlands that is neutral about the actual nature of morality. Finally, I argued for one specific account of morality and moral meaning; Slote's Moral Sentimentalism, as the best possible nonrelativistic explanation of the role of empathic emotions in moral motivation in humans and animals. By supplementing both the moral sensitivity approach proposed by Rowlands, and Slote's moral sentimentalism with the APV account of empathically modulated basic emotions, it was possible to explain not only why animals' empathic emotions track moral propositions and internally represent the welfare of others to the animal, but also to give a specific account of morality in terms of the empathic capacities that animals share, to a limited degree, with human beings.

Yet this positive account of how animals are morally motivated is independent of the arguments I used to dismiss the criticisms that have been raised to doubt the possibility of animal consciousness, or to doubt the possibility of animals possessing empathic emotions. Given that higher order thought is not necessary for consciousness, as long as we qualify our claims about animal concepts and animal consciousness so as to take into account the fact that phenomenal consciousness, while present in animals, is likely radically different for beings with radically different cognitive capacities, we are justified in ascribing consciousness to animals. Furthermore, since linguistic conceptual capacities are not necessary for having emotional concepts, including the core representational themes that are internally represented in basic emotions, we are justified in ascribing emotions that involve predicate and value-laden internal representations to

animals as long as we qualify such ascriptions by relativizing them to context using Rowlands' translation schema and thereby abstracting away from human language or theory based concepts. Finally, because basic emotions do not involve complex linguistic or self-reflective judgments, we can make claims about the identity of *kinds* of emotional states across species lines. This is because emotional states are not solely constituted by raw phenomenological feelings but are at least partially constituted by their shared structural and functional properties.

It follows from these critical arguments and qualifications that we are justified in ascribing value laden emotions to animals. The empirical investigations into the presence and development of empathic capacities in animals that I reviewed also provide strong evidence that we are justified in ascribing empathic emotions to animals. These empathic emotions develop in sophistication over time, convey the value of another's welfare to the empathic animal, and motivate apparently moral action. Korsgaard and Dixon doubted that such apparently moral action can truly be classified as moral because morality seems to require self-critical capacities and complex judgments about one's purposes and the value of others. However, I argued that these doubts were driven partially by an implausible overly stringent picture of moral motivation and partially by a conflation of the question of whether an organism can be praised or blamed with the conceptually distinct question of whether an organism can be morally motivated. I demonstrated that moral motivation does not require the capacity for critical self-scrutiny or making complex judgments about the morally relevant features of the scenario even in the human case. It follows that requiring this of animals is unjustifiably stringent as a requirement for moral motivation. This concludes the summary of the critical arguments

I have put forward in this dissertation, which are intended to stand independently of the positive view of empathic emotions and moral motivation and to shift the burden of proof to skeptics of animal morality.

My positive account of moral motivation is suggested by both sound philosophical argumentation and empirical evidence. It is that animals are morally to the extent that they are motived by an empathically modulated emotion such as compassion (empathically feeling and being moved to prevent the suffering of another) or benevolence (empathically feeling and being moved to promote the delight of another). Such an emotion conveys the value of the well-being of the other to the empathic animal through their physiological states value laden contribution to the background of their phenomenological state. The empirical evidence that suggests the APV account in conjunction with my arguments against other accounts of emotion is partly physiological, partly behavioral, and partly evolutionary. The APV account of emotions and empathically modulated states is, I have argued, the best explanation of what these three sources of empirical evidence about the structure of emotions (in at least mammalian species) suggest about the likely phenomenological properties of empathic emotions. This account also allows us to specify how animal's emotions might track evaluative moral positions in a way that preserves the intuition that an animal must internally represent the value of another's well-being in order to be morally motivated.

Rowlands' proposal was then argued for as a moral theory neutral way to ascribe animals morally motivating emotions. In sum, this proposal was that an animal is correctly ascribed a moral emotion just in case there is a moral proposition that, if true, guarantees that the animal's emotion is not misguided. This tracking account allowed me

to argue that whatever ultimately guarantees the truth of moral propositions – be it good consequences, virtuous motivation, or universal moral principles – we have good reason to ascribe moral motivation to empathically motivated animals. We can justifiably make such ascriptions in virtue of the animal's internally representing the well-being of another in an emotion that implies a moral proposition that must be true if that emotion is not to be misguided.

Finally, I argued for a specific account of how moral propositions get their meaning and justification conditions; a sentimentalist account that based moral motivation in the empathic capacities that we share with nonhuman animals. Slote's empathy based ethics and metaethics allowed me to show one possible and likely way that morality arises out of natural capacities we share with nonhuman animals and why we are justified in morally approving and disapproving (but not praising or blaming) animals' actions according to whether or not those actions reflect or exhibit the presence or lack of empathy. I argued that this account is independently plausible, and that it is compatible with and complementary to Rowlands moral proposition tracking account in so far as it gives an explanation for how moral propositions are justifiably considered true or false in way that avoids relativism. Then I argued that Slote's sentimentalist account is also usefully supplemented by my APV account of the phenomenological structure of empathically modulated emotions. This supplementation allows Slote's account to handle certain paradoxes that arise out of basing morality entirely on a (perhaps necessarily) vague conception of empathy.

If this sentimentalist APV account of moral motivation is correct, it would have an interesting, and, so far as I know, largely untested and unstudied empirical

consequence. This is because if some animals have empathy with empathy, then they also have the very same moral approval that grounds human moral judgments. That is, this account predicts that if animals approve of empathy exhibiting behavior and disapprove of behaviors that exhibit a lack of empathy it would constitute evidence that they also understand the meaning of moral goodness and wrongness.

I suspect that this empirical prediction would best be tested in relation to what I will call, for lack of a better term, "motherhood training" in young chimpanzees. When a chimpanzee mother allows curious younger females to hold and care for their offspring, I suspect that they will typically act in ways that exhibit a rudimentary sense of moral approval (empathy with empathy) of the empathic actions that the young female chimpanzee exhibits toward their offspring. The sentimentalist APV account likewise predicts that when a chimpanzee mother witnesses a young female chimpanzee acting violently or dangerously towards their offspring they will act in ways that are indicative of rudimentary moral disapproval. It is already known that chimpanzee mothers carefully supervise such encounters and intervene when improper actions are taken towards their offspring.[107] However, little attention has been devoted, as far as I know, to determining the specific emotional state that the chimpanzee mother is in as she supervises. If she does show empathy with the empathy of an adolescent female, this would constitute intriguing evidence of a rudimentary understanding of moral goodness according to the APV sentimentalism defended here.

A further area where this empirical prediction of APV sentimentalism might be studied in primates (and other species) is the peace-making behaviors primates display after conflicts; particularly with respect to the extent to which such behavior is

[107] See De Waal's *Good Natured* 1996.

encouraged by third parties and the lack of such behavior is discouraged by third parties. Again, it is well known that some animals, such as vervets, will attack the attackers of their kin and afterwards attempt to mend relationships with those attackers with reconciliatory behavior. (Chenny and Seyfarth in De Waal 1996 pg. 204) Chimpanzees often respond with celebratory behaviors when reconciliation occurs, and third party interventions between two violent or aggressive parties are also known to occur. (De Waal 205) This behavior is predicted and explained by the APV sentimentalist theory advanced here as the result of animals morally approving, by having empathy with empathy, of peacemaking and morally disapproving of violent behaviors that show a lack of empathy on the part of the aggressors.

Although both the mother-infant bond and the peacemaking behavior of chimpanzees has been (and continues to be) thoroughly experimentally researched, there has been little to no systematic attempt to test and interpret the possibility of second order empathic responses of mothers and third parties to conflicts. The APV sentimentalism advanced in this dissertation generates the novel and testable empirical predictions that some animals, to the extent that they can detect, understand, and have empathic approval of empathic motives, have an understanding of moral goodness. This theoretical fruitfulness is an explanatory virtue for a theory about a subject matter that is the proper domain of both philosophical and empirical inquiry.

Yet even if this positive account of *morality* should turn out to be incorrect, the fact that animals internally represent the welfare of others will stand as established on the basis of the APV account of *emotions* when combined with Rowlands' proposal that animals have moral motivation in virtue of having emotions that, if they are not to be

misguided, imply true moral propositions. This constitutes a moral theory neutral way of ascribing animals moral emotions that has some practical moral consequences for any particular theory of morality. The first of these consequences is that our obligations to animals extend far beyond the mere ensuring that they feel as little pain as possible as we use them as resources for the general good of human well-being. While, as I shall briefly explain, this is not a novel consequence, the APV account and Rowlands' tracking approach do suggest some new insights into the issue of animal experimentation for medical knowledge and human well-being.

Since Peter Singer's seminal utilitarian arguments (in Animal Liberation 1975) against the use of animals in factory farming, animal husbandry, blood sports, and so on; a variety of theorists have derived arguments for the ethical consideration of animals from alternative ethical perspectives. These include, but are not limited to, virtue ethics, deontology, care ethics, and contractualism.[108] There has been a growing consensus in the ethics community among those sensitive to the ethical import of the by now overwhelmingly obvious affirmative answer to Bentham's famous question of whether animals can suffer, that current factory farming practices, hunting and blood sports, animal use for leisure products, and cruel and unnecessary animal experimentation, are all unjustifiable from any sensible ethical perspective.

However, what remains an open question is whether the use of animals in experiments involving animal suffering or death (fear conditioning, pain induction research, vivisection, etc.) is justifiable when there are well known valuable gains for

[108] In particular, Mark Rowlands has argued from most of these ethical perspectives for the equal consideration of animal wellbeing in his *Animal Rights* (1988/2009). Before him, Regan advanced impactful, but potentially problematic, deontological arguments (in *The Case for Animal Rights*, 1983) for the intrinsic value of animal life that ethically required the abandonment of cruel animal uses including some medical research practices.

human medical knowledge and the well-being of human and animal life. This topic has sparked a new kind of debate, one where the emotions, flourishing, pleasure and pain, and general well-being of both animals and humans is taken into consideration. Most theorists and scientists writing on under what conditions (if any) the use of animals in experiments involving animal suffering or death is justifiable when there are well known gains for medical knowledge and the well-being of human and animal life take one of two dichotomous attitudes. These attitudes towards medical experimentation on animals are driven by the implicitly or explicitly endorsement one of the two following justified (but seemingly incompatible) reasons:

1) Because the results of scientific research both A) historically and B) currently (when considered collectively), involving experimentation on animals that would be unethical to perform on humans have allowed professionals in the medical field to understand, save, and dramatically improve the lives of humans, we ought to favor using animals for some medical and scientific experiments (regardless of their suffering), as long as these practices are regulated to guarantee as little suffering as possible.

2) A) Any animal's life and welfare is valuable both to and for that animal and, in virtue of this value and all it entails (for example that suffering, anxiety, and fear are bad, and that pleasure, joy, and love for kin and playmates are good) B) it is unethical to use any animal's life or negatively affect their wellbeing for any amount of beneficial gain for human beings.

Theorists and scientific researchers endorsing 1) typically cite the dramatic advances in the medical sciences that most humans rely upon for their (lives and/or) well-being and accuse those who endorse 2) of ignoring the fact that humans are also animals. In particular, they point out that it seems right to regard humans as animals with far more valuable concerns and greater rights than other species of animals. On the other hand, those who endorse 2) typically accuse their opponents of being inconsistent when they

hold that human life and welfare is valuable in a way that does not permit experiments on humans for instrumentally valuable purposes, while at the same time holding that other species of animals' lives ought to be used for such experiments. Often those who endorse 2) also point out that there is no principled and ethically relevant distinction that can be drawn between the value, concerns, and rights of human beings as a species and the value, concerns, and rights of other animal species. The seeming incompatibility of 1) and 2) has led to a temporary impasse that the account of emotions put forward here is, I believe, capable of redressing.

If any progress can be made on this debate, it hinges on developing a neutral, justified, and accurate way to comparatively assess the ethical importance of animal and human life, emotions, and welfare. 1) and 2) are in fact largely compatible because there are principled reasons to think that the suffering of some animals and some humans in experiments to improve the well-being of all animals and all humans is justified under certain circumstances. To endorse this claim is to reject 2B) as a universal claim (which is dubious at best), but not 2A). This endorsement also preserves the intuition behind both 1A) and 1B) that there are some situations in which a great benefit justifies risky experiments potentially or actually involving suffering, impairment, or even death. If this is correct, then what is needed is a way to determine under what circumstances the suffering of an organism of one species is justified by the (potentially or actually) improved or saved lives of another (and/or the same) species. I believe that the impasse caused by these dichotomous attitudes can be bridged by developing a way of comparing the well-being of organisms, and the value of their unencumbered and flourishing life, across species.

The application of my theory of emotions and Rowlands' tracking proposal to this debate might make such a comparison of well-being tractable, in so far as animals and humans share both the core relational themes that are represented in basic emotions and empathically modulated emotions that have a common phenomenological and value-laden structure. Not only can we now in principle compare social emotions (and their implicit values) across species; this dissertation also provides the beginning of model for such comparisons. Systematizing the details of such a model is a much larger project than can be, or should be, undertaken here, but it is nevertheless a promising avenue of theorizing in applied medical ethics that is suggested by the positive account of emotions in this dissertation.

The thesis I have defended is that some animals, including all mammals, are legitimately ascribed moral emotions. Our duties to animals depend on the types of emotions, values, and purposes that they can have. If their suffering is as (or more) profound than ours when confined, isolated, constricted, and led to the slaughter without ever having lived the type of life their negative emotional states heedlessly propel them towards, are our crimes to them are indescribably worse than the mere propagation of pain on lower life forms for the brief higher pleasure we attain by treating animals as resources for food and commodities. Treating something morally means more than merely not directly causing them pain or a painful death. It means understanding and respecting their purposes, their values, and the emotional states and cognitive abilities that determine and make possible these purposes and values, and acting with such an understanding in mind. Such an understanding is impossible without an account of what it is to have an emotion, of what emotions are and what kinds of cognitive abilities they

imply or require, and how emotions give rise to values and purposes. This dissertation provides such an account, and thereby helps us better understand and respect our *emotional and moral* animal brethren.

References

Allen, C. (1999). "Animal Concepts Revisited: The Use of Self-monitoring as an Empirical Approach". *Erkenntnis.* 51, 1: 33-40.

Allen, C. and Bekoff, M. (1997). *Species of Mind: The Philosophy and Biology of Cognitive Ethology*. MIT Press.

Allen, C. and Hauser, Marc D. (1991). "Concept Attribution in Nonhuman Animals: Theoretical and Methodological Problems in Ascribing Complex Mental Processes". *Philosophy of Science.* 58.

Aristotle. (1998). *The Nicomachean Ethics*. Trans. David Ross. New York: Oxford University Press.

Aristotle. (1968). *De Anima* Books II and III (with certain passages from Book I), translated by D. W. Hamlyn, Clarendon Press.

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

Bekoff, M. (2007). *The Emotional Lives of Animals.* CA: New World Library.

Bekoff, M ed. (2000). *The Smile of a Dolphin: Remarkable Accounts of Animal Emotions*, New York: Discovery Books.

Carruthers, Peter (1996). *Language, Thought and Consciousness* (Cambridge: Cambridge University Press).

Carruthers, Peter (2004). "Why the Question of Animal Consciousness Might Not Matter Very Much". *Philosophical Psychology* 17: 83-102.

Cheney, D & Seyfarth, S. (1990). *How Monkeys see the World*, University of Chicago Press.

Chater, N & Heyes, C (1994). "Animal Concepts: Content and Discontent". *Mind and Language.* 9: 209-46.

Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason and the Human Brain.* New York: Putnam.

Darwin, C. (1872/1965). *The Expression of the Emotions in Man and Animals.* Chicago: University of Chicago Press.

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

Davidson, D (1982). "Rational Animals". *Dialectica.* 36: 317-28.

DeGrazia, D (1996). *Taking Animals Seriously*, Cambridge University Press.

Descartes, Rene (1649/1988). *The Passions of the Soul.* In J. Cottingham, R. Stoothoff, & D. Murdoch (Trans. And Eds.), *Selected Philosophical Writings of Rene Descartes*. Cambridge, UK: Cambridge University Press.

DeSousa, Ronald. (1987). *The Rationality of Emotions.* MIT Press.

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

Dixon, B. A. (2008). *Animals, Emotions, and Morality: Marking the Boundary.* Amherst, NY: Prometheus Books.

Ekman P. & Friesen, W. V. (1971). "A New Pan-cultural Facial Expression of Emotion. *Motivation and Emotion.* 10: 159-168.

Ekman, P. (1977). "Biological and Cultural Contributions to Body and Facial Movement. In J. Blacking (Eds.) *Anthropology of the Body* (pgs. 34-84). London: Academic Press.

Ekman, P. (1992). "An Argument for Basic Emotions." *Cognition and Emotion.* 6: 169-200.

Gallup, GG Jr. (1970). "Chimpanzees: Self-Recognition". *Science.* 167: 86–87.

Griffiths, P. E. (1997). *What Emotions Really Are.* Chicago: University of Chicago Press.

Korsgaard, Christine M. (2010). "Reflections on the Evolution of Morality." *The Amherst Lecture in Philosophy.* 5: 1-29.

Korsgaard, Christine M. (2006). "Morality and the Distinctiveness of Human Action." in Frans de Waal, Primates and Philosophers. Princeton: Princeton University Press.

Korsgaard, Christine M. (2005). "Fellow Creatures: Kantian Ethics and Our Duties to Animals." *The Tanner Lectures on Human Values.* 25: 79–110.

Kriegel, Uriah. (2009). "Self-representationalism and phenomenology." *Philosophical Studies.* 143: 357-381.

Lazarus, R. S. (1991). *Emotion and Adaption.* New York: Oxford University Press.

LeDoux, Joseph. (1996). *The Emotional Brain The Mysterious Underpinnings of Emotional* Life. Simon & Schuster Paperbacks. New York.

Lyons, William. (1980) *Emotion*. Cambridge University Press.

Millikan, Ruth (1984) *Language, Thought and Other Biological Categories.* Cambridge, MA: MIT Press.

Nagel, Thomas. (1974/2001). "What Is It Like To Be A Bat". *Philosophical Review. 83.* in Analytic Philosophy. Ed. Martinich, A. P. and Sosa, David. Blackwell Publishing.

Newen, A & Bartels, A (2007). "Animal Minds and the Possession of Concepts". *Mind and Language*. 20: 283-308.

Nussbaum M. C. (2001). *Upheavals of Thought: The Intelligence of the Emotions.* Cambridge, UK: Cambridge University Press.

Panskepp, J. (1982). "Toward a General Psychobiological Theory of Emotions. *Behavior and Brain Sciences.* 5: 407-467.

Panskepp, J. (2000). "Emotions as Natural Kinds Within the Mammalian Brain". In M. Lewis & J. Haviland-Jones (Eds.), *Handbook of Emotions* (2nd. Ed., pp. 137-156). New York: Guilford Press.

Panksepp, J. (2004). "Affective Consciousness: Core Emotional Feelings in Animals and Humans." *Consciousness and Cognition.* 14: 30-80.

Panksepp, J. (2004). *Affective Neuroscience: The Foundations of Human and Animal Emotions.* Oxford University Press.

Plato. (1987). *The Republic.* Trans. Lee, Desmond. New York: Penguin Group.

Plato. (1999). *Symposium.* Trans. Benjamin Jowett. NJ: Prentice Hall/Library of Liberal Arts.

Preston, Stephanie D. and de Waal, Frans B. M. (2002). "Empathy: Its Ultimate and Proximate Bases" *Behavioral and Brain Sciences.* 25: 1–72.

Prinz, Jesse. J. (2004). *Gut Reactions*. Oxford University Press.

Prinz, Jesse. J. (2007). *The Emotional Construction of Morals.* Oxford University Press.

Prinz, Jesse. J. (2008) "Embodied Emotions". in *Mind and Cognition*. 3rd Edition. Ed. Lycan, William G. and Prinz, Jesse J. Blackwell Publishing.

Rolls, Edmund T. (1999) *The Brain and Emotion.* Oxford University Press.

Rosenthal, David (1986) 'Two Concepts of Consciousness'. *Philosophical Studies.* 49: 329-59

Rosenthal, David (1993) "Thinking That One Thinks". in G. Humphreys and M. Davies (Eds) *Consciousness.* Oxford: Basil Blackwell.

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

Rowlands, M. (1998/2009). *Animal Rights: Moral Theory and Practice*, Macmillan

Rowlands, M. (2007). *Body Language,* The MIT Press

Rowlands (2001). "Consciousness and Higher-Order Thoughts". *Mind and Language.* 16, 3: 290-310.

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

Ryle. G. (1949). *The Concept of Mind*. Chicago: University of Chicago Press.

Sartre, Jean-Paul. (1939/1971). *Sketch for a Theory of the Emotions,* Mariet, Methuen (Trans.) Routledge.

Schachter, S. and Singer, J. (1962). "Cognitive, Social and Physiological Determinants of Emotional State". *Psychological Review.* 69.

Skinner, B. F. (1974). *About Behaviorism,* Cape.

Slobodchikoff, C. N. (2002). "Cognition and Communication in Priarie Dogs". in *The Cognitive Animal* Bekoff, M, Allen, C and Burghardt, G (Eds.) MIT Press.

Slote, Michael. (2007). *The Ethics of Care and Empathy*. New York: Routledge.

Slote, Michael. (2010). *Moral Sentimentalism*. New York: Oxford University Press.

Solomon, Robert. (2008) "Emotions and Choice." in *Mind and Cognition*. 3rd Edition. Ed. Lycan, William G. and Prinz, Jesse J. Blackwell Publishing.

Speaks, Jeff. (2005) "Is There a Problem about Nonconceptual Content?" *The Philosophical Review,* Vol. 114, 3.

Stich, S (1978) "Do Animals Have Beliefs?" *Australasian Journal of Philosophy*. 57:15-28.

Watson, J. B. (1925) *Behaviorism,* Norton.

Williams, Bernard. (1981). *Moral Luck*. New York: Cambridge University Press.