Monkeys, Men, and Moral Responsibility:

A Neo-Aristotelian Case for a Qualitative Distinction

Paul Carron, Baylor University

I.   Introduction

On February 17, 2009, a large chimp named Travis attacked a friend of his owner, blinding her, severing her nose, ears, hands, and severely lacerating her face. The police subsequently shot and killed Travis. The Primatologist Frans de Waal wrote about the incident on several occasions, bemoaning the state of the laws that allow humans to own chimps as pets.[1] But one thing was curiously absent from de Waal's comments: he never *blamed* Travis for his actions. Surely if Travis were a mentally competent human male there would be no doubt about who to blame – we would blame Travis. But no one – including one of the foremost advocates of continuity between primate behavior and human morality – blames a chimp for a brutal murder.

Initially this doesn't surprise us. After all, we don't hold very young children morally responsible for their actions in the same way that we hold adults accountable. Furthermore, many of us accept that non-human primates do not have the cognitive and psychological faculties necessary for responsibility ascriptions. Humans can deliberate and weigh reasons about whether or not to harm another person or animal, and can form an intention based on that deliberation and act on that intention; primates and other non-human animals are not privy to the same complex mental processes. But de Waal has spent his career arguing for *continuity* between primate behavior and human morality. He argues that the building blocks of human morality are evident in primates, that

---

[1] de Waal, "Another Chimp Bites the Dust," *Huffington Post*, February 17, 2009. http://www.huffingtonpost.com/frans-de-waal/another-chimp-bites-the-d_b_167768.html; "Chimp Attack: Missed Opportunities," *Huffington Post*, December 7, 2009. http://www.huffingtonpost.com/frans-de-waal/chimp-attack-missed-oppor_b_383078.html; http://www.emory.edu/LIVING_LINKS/sciam_travis.html.

humans and primates are social to their core, and that the root of morality for both humans and

primates is empathy. So if de Waal believes in this "total gradualism," why not blame Travis?[2]

De Waal is part of a philosophical school known as moral sentimentalism. According to de

Waal, moral sentimentalism "firmly anchors morality in the natural inclinations and desires of our

species" and emphasizes the role of the emotions in human morality.[3] Sentimentalist stresses

"empathetic caring as the touchstone of virtuous agency."[4] For a sentimentalist, moral judgments

stem from a person's reactive attitudes—primarily empathy.[5] De Waal thinks that he demonstrates

that some primates have the capacity for empathy—they can read other agent's minds and react

appropriately. Based in part on de Waal's research, the psychologist Jonathan Haidt concludes that

"[g]iven so many close parallels between the social lives of humans and chimpanzees, the burden of

proof must fall on those who want to argue for discontinuity—that is, that human morality arose ex

nihilo when we developed the ability to speak and reason."[6] Although I have no interest in

defending the position that morality arose *ex nihilo*, I argue that de Waal's "total gradualism" is

problematic: while his research provides evidence for the *descent* half of Darwin's project, he does

not properly account for the *modification* half.[7] The human modification most relevant to moral

responsibility is the ability to emotionally regulate. A creature lacking the capacity for emotion

---

[2] This is Christine Korsgaard's term. "Morality and the Distinctiveness of Human Action," in *Primates and Philosophers: How Morality Evolved,* Waal, F. B. M. de, Stephen Macedo, Josiah Ober, and Robert Wright. Princeton, N.J.: Princeton University Press (2006), p. 104.

[3] *Primates and Philosophers,* p. 18

[4] Damian Cox, "Agent-based Theories of Right Action," *Ethical Theory and Moral Practice* 9 (2006), p. 506.

[5] Hume's classic formulation: ""The final sentence . . . which pronounces characters or actions amiable or odious, praiseworthy or blameable . . . depends on some internal sense or feeling which nature has made universal in the whole species." *Enquiry Concerning Human Understanding and Concerning the Principles of Morals.* 3rd Edition. Edited by L. A. Selby-Bigge and P. H. Nidditch. Oxford: Clarendon Press, (1777) 1975. Also consider Edward Westermarck: "To name an act good or bad, ultimately implies that it is apt to give rise to an emotion of approval or disapproval in him who pronounces the judgment..." *The Origin and Development of the Moral Ideas.* London: Macmillan and Co. 1906.

[6] Haidt, "The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review* 108 (2001), p. 826.

[7] As Philip Kitcher highlights in his comments on de Waal's lectures, Darwin's phrase "descent with modification" has two necessary aspects, descent and modification. Kitcher rightly notes that while de Waal deals adequately with the descent aspect, he largely ignores the modification aspect, which is just as important and interesting. "Ethics and Evolution: How to Get Here from There," in *Primates and Philosophers: How Morality Evolved,* Waal, F. B. M. de, Stephen Macedo, Josiah Ober, and Robert Wright. Princeton, N.J.: Princeton University Press (2006), p.124.

regulation cannot be held morally responsible for its actions. Thus emotion regulation marks a boundary between primate behavior and human morality.

Unlike other critics, I will engage de Waal from within the sentimentalist tradition itself broadly construed and the Aristotelian virtue tradition more specifically.[8] On an Aristotelian understanding of the emotions, the emotional capacities of non-human animals are both continuous and discontinuous with the emotional capacities of human animals. Emotions stem from the perceptive faculty: both humans and emotions have a perceptive faculty, and both humans and animals can discriminate or discern[9] whether or not what they perceive should cause fear (for instance). But only human animals can freely and intentionally cultivate their character so that the perceptual object (i.e. the cause of fear) *appears* differently.[10] And when the end appears differently—i.e. the perception changes—the emotion also changes. Finally, Aristotle's understanding of perception, rational speech, and decision account for how human animals can make lasting changes in how the end appears, resulting in different character traits. Aristotle is keenly aware that many animals are capable of emotions and many other forms of intelligence, but non-human animals do not have sufficient control over those feelings and those feelings lack the necessary rational content to warrant moral praise or blame.

I focus on one point: de Waal's understanding of empathy and the place of cognitive mental processes in primate empathy. I then examine Aristotle's understanding of pity and anger and the place of cognitive mental processes in these emotions. This de Waal/Aristotle conversation is

---

[8] Here I am in basic agreement with Damian Cox who argues – contrary to Michael Slote – that both virtue theory and sentimentalism are *agent-based theories*. As Cox puts it, both virtue (agent-prior) and sentimentalism (agent-based) ethical theories "derive the moral significance of actions from virtuous agency and do this without appealing to non-*aretaic* moral notions." More fundamental to my purposes, on both accounts an agent cannot *reason* morally or *perform* moral actions without having virtuous feelings or emotions.

[9] Greek κρίνειν.

[10] "Someone might say that everyone aims at the apparent good, but does not control its appearance; but the end appears to each person in a way that corresponds to his character. For if each person is somehow responsible for his own state of character, he will also be himself somehow responsible for its [viz. the end's] appearance [phantasias]" (1114b1–3; cf. 1114b17). From Sherman, *The Fabric of Character*, p. 32.

valuable because de Waal claims that his sentimentalist understanding of morality is part of a tradition dating back to Aristotle's emphasis on the role of emotions in morality (which Aristotle certainly does emphasize), yet critiques of de Waal usually come from other perspectives.[11] Furthermore, Aristotle does attribute intentionality and cognitive responsiveness to the appetitive states of some animals, but he will only ascribe moral responsibility to human actions.[12] Therefore Aristotle can help us understand why we intuitively do not hold our primate cousins morally accountable despite our shared traits.

II.   de Waal on the Moral Capacities of Non-Human Primates

De Waal has long argued that there is more continuity than discontinuity between primate and human behavior, specifically in regards to behavior that is relevant to morality. However, he will admit differences. For instance, when discussing Smith's "impartial spectator" he states that in the area of disinterestedness human emotions "seem to go radically further than other primates' [emotions]."[13] But he stops short of admitting a difference in *kind* between human morality and primate behavior. Rather, the seeds of morality present in primate behavior are the same seeds that germinate and blossom into the essential characteristics of human morality. This is most evident in de Waal's support of Darwin's claim that "[a]ny animal whatever, endowed with well-marked social instincts, the parental and filial affections being here included, would inevitably acquire a moral sense or conscience, as soon as its intellectual powers had become as well developed, or nearly as well developed, as in man."[14] He argues that primates have the social instincts as well as the basic

---

[11] Most notably deontological and utilitarian.

[12] Nussbaum, Martha Craven, *The Therapy of Desire: Theory and Practice in Hellenistic Ethics*. Princeton, N.J.: Princeton University Press (1994), p. 81. On the one hand this is technically true: only humans are capable of *proairesis* and thus full moral and legal sanctions. However, I would argue that on a properly reconstructed account of *hekousios* from *NE* III.1-5, some non-human animals are apt targets of reactive attitudes such as praise and blame. I argue that Aristotle leaves room for asserting that many non-human animals are morally appraisable agents, but only humans are morally responsible agents.

[13] *Primates and Philosophers,* p. 20.

[14] Darwin, Charles, *The Descent of Man, and Selection in Relation to Sex*, Princeton, N.J.: Princeton University Press (1981), 71-72. Quoted in *Primates and Philosophers* p. 14.

"intellectual powers" necessary for human morality. The most important social instinct is empathy; the most important intellectual power is the ability to recognize other agents' mental states, what de Waal calls "theory of mind" (ToM). I will briefly deal with each of these two related claims.

The foundation of de Waal's argument for the continuity of primate behavior and human moral behavior is his claim that human beings are "social to the core," a core that is largely shared with primates such as chimpanzees and bonobos.[15] This shared social core includes the traits of cooperation, reciprocity, and empathy. De Waal says that empathy "covers a wide-range of emotional linkage patterns, from the very simple and automatic to the highly sophisticated"[16] and that both the very simple and the highly sophisticated are observed in primates and are an essential part of human morality. He calls the most sophisticated form "attribution," or fully adopting another's perspective, referred to as ToM. In order to adopt that perspective, the agent must not only have the ability to look for reasons for the other's emotions, but be able to understand the other agent's mental states. De Waal gives a more succinct definition of sympathy: "an affective response that consists of feelings of sorrow or concern for a distressed or needy other (rather than the same emotion as the other person)."[17] A good example of primate sympathy is the chimp Kuni who tried to help a bird fly by climbing to the top of the highest tree in the enclosure, wrapping her feet around the tree branch in order to leave her hands free, and then spreading out the bird's wings and launching the bird into flight. Kuni obviously understood the difference between a chimp's needs and the needs of a bird, and responded appropriately. De Waal cites a number of similar examples that he takes as sufficient proof that primates can adopt another's viewpoint. Chimps respond to the pain they see in another chimp; a chimp helped another chimp who had fallen into a moat; chimps will protect comrades who are being attacked. All this requires a basic understanding

---

[15] *Primates and Philosophers,* p. 5.
[16] Ibid., 41.
[17] Ibid., 26.

of the other's situation and emotional cues, and the ability to purposefully respond. However, on Aristotle's virtue account much more is necessary to ascribe moral responsibility to Kuni and her comrades.

De Waal emphasizes emotions such as sympathy in primate behavior, and believes these characteristics are fundamental aspects of human morality. He argues that primates are able to recognize the needs of others and respond appropriately to their emotional states. He concludes that there is fundamental continuity between primate and human behavior, and that the building blocks of morality are evident in primate behavior. Analyzing the human versions of sympathy and anger makes it clear that these emotions require not only the ability to understand and respond to another agent's intentions, but intentionality, self-conscious awareness of one's own mental states, and perhaps most fundamentally, the ability to self-consciously control and give dispositional continuity to one's emotional responses. Although there are no exact equivalents of sympathy in Aristotle, Aristotle does analyze the closely related pity in the *Rhetoric*, and he explicitly discusses anger—another reciprocal emotion—in the *Nicomachean Ethics*.

III.   Aristotle on Moral Virtue

According to Aristotle, a moral virtue regulates emotions and feelings,[18] and the moral virtues along with their constitutive emotions are guided by a person's rational capacity.[19] In other words, human affect is essential for morality, and human affective capacities have essential cognitive elements. On an Aristotelian account, a number of elements are required in order to deem a *pathos*/emotion[20] morally praiseworthy or blameworthy. There are requirements that reference the

---

[18] "Now virtue is about feelings and actions, in which excess and deficiency are in error and incur blame, whereas the intermediate condition is correct and wins praise, which are both proper to virtue." *NE* 1106b 25 (Irwin Translation).

[19] "For while the plantlike [part] shares in reason not at all, the [part] with appetites and in general desires shares in reason in a way, insofar as it both listens to reason and obeys it." *NE 1102b;* "virtue…is defined by reference to reason…" *NE* 1107a 2 (Irwin).

[20] Though it is not quite right to equate the ancient Greek *pathos* with the modern English *emotion*, Aristotle's use of pathos generally resembles the contemporary usage of emotion in moral psychology. Although *pathos* typically

agent's freedom or control with regard to the action, and there are requirements that focus on the cognitive content of human emotions. For Aristotle, free—and thus morally responsible—mental acts (including emotions) require that the agent have within herself "the 'origin' (*arche*) of the action."[21] In other words, to act freely and responsibly, the action must be determined by or stem from the agent. Furthermore, morally responsible action requires that the agent possess a kind of dual power: "…when acting is up to us, so is not acting."[22] Morally responsible action requires that the agent have it within her power to perform the action or not.[23] This is especially true in regards to virtues and vices: "Hence virtue is also up to us, and so also, in the same way, is vice."[24] Second, there are requirements that reference the cognitive aspects of emotions. Many commentators agree that Aristotle's understanding is that emotions are intentional mental states that are about or directed at an object.[25] Furthermore, emotions are *discriminating capacities*: emotions discern particular features of the agent's environment in a way that purely rational discrimination cannot.[26] As Sherman puts the point, "We thus come to have relevant points of view for discrimination as a result of having certain emotional dispositions. We notice through feeling what might otherwise go unheeded by a cool and detached intellect."[27] Due to their cognitive nature, emotions are not simply events that the agent *suffers*, but mental events that are at least to some extent under the agent's control.

---

refers to "suffering, feeling, emotion," or literally "what befalls one," Aristotle's *pathos* is regulated by moral virtue, can be guided by a rational principle, is subject to choice and responsibility ascriptions, etc. *Pathos* isn't simply something that happens to a person according to Aristotle, but rather is an activity of the soul that the agent is responsible for.

[21] Sorabji, Richard, *Necessity, Cause, and Blame: Perspectives on Aristotle's Theory*, (Ithaca, N.Y.: Cornell University Press, 1980), 234. Quoted in Kane, Robert, *The Significance of Free Will*, New York: Oxford University Press, 1998), 33.

[22] Aristotle, *Nicomachean Ethics*, 1113b6. Quoted in Watson, "Free Action and Free Agency," p. 1 (fn1).

[23] Kane, Robert, *The Significance of Free Will*, (New York: Oxford University Press, 1998), p. 33.

[24] Aristotle, *Nicomachean Ethics*, 1113b4. Irwin's translation.

[25] For instance, see Nussbaum, *The Therapy of Desire*, p. 80.

[26] *De Motu Animalium* describes cognitive capacities this way: "We see that the things which move the animal are thought and *phantasia* and decision and wish and appetite. And all these can be reduced to intellect (*nous*) and desire (*orexis*). For both *phantasia* and perception (*aisthesis*) hold the same place as intellect, for they are all cognitive (*kritika*) (MA 700b17-21)." Translation from Moss, Jessica. (2012). *Aristotle on the Apparent Good: Perception, Phantasia, Thought, and Desire*. Oxford Scholarship Online, pp. 9-10.

[27] Sherman, *The Fabric of Character,* p. 45.

Besides the freedom required for moral emotions, and the cognitive nature of the emotions, Aristotle is also clear that praise or blame for an emotion requires that the emotion be in accord with reason and stem from an internalized characteristic—a disposition to action and emotion—freely cultivated over a long period of habitual actions.[28] Again, emotions for which an agent can be praised or blamed must not merely be phenomena that happens to the agent, but must be closely connected to the agent's intellect and will.

IV.   Aristotle on Pity and Anger

Aristotle defines pity as "[p]ain at an appearing evil, destructive or painful, belonging to one who does not deserve to have it happen—the sort of evil that one might expect oneself to suffer, or some member of one's family."[29] Nussbaum outlines three relevant aspects of pity in Aristotle's definition. First, the recipient of the agent's pity must be undeserving of the misfortune. Second, the agent must be aware of his or her own vulnerability to a similar misfortune. Finally, the potential misfortune must be significant.[30] Note that Aristotle uses pity in a very similar way that empathy and more specifically sympathy are used by some contemporary neuroscientists: "empathy occurs when an observer perceives or imagines someone else's (i.e., the target's) affect and this triggers a response such that the observer partially feels what the target is feeling."[31] When that affect is one of distress, the emotional state of the agent (or "observer") is sympathy. In this way sympathy is a species of empathy.

Aristotle's definition of pity reveals the rich cognitive structure of human emotions. Emotions are not mindless responses or purely affective states. Rather, *human emotions are complex intentional mental states* directed at specific objects. An agent cannot have pity or sympathy for another

---

[28] Moral virtue is "a state (ἕξις– Latin *habitus*, related to "hold" or "have") [that] results from [the repetition of] similar activities" (II.1.7), and it produces actions that are in "accord with the correct reason" (II.1.1).

[29] *Rhetoric,* 1385b13-15.

[30] Nussbaum, p. 87.

[31] Singer, Tania, and Claus Lamm, "The Social Neuroscience of Empathy," *Annals of the New York Academy of Sciences* 1156:1 (2009), p. 82.

person without seeing that person, understanding their situation, and being concerned about their state. An agent must also be aware of the self's own vulnerability; this means understanding the difference between subject and object. De Waal claims that primates are capable of both of these cognitive abilities, and they are at least to some extent. More importantly, however, is the fact that the human agent is aware of those feelings of pity, can reflect on the object of and reasons for the pity, and can self-consciously choose (at least at times) to be motivated by or to act on that pity. In her response to de Waal, the philosopher Christine Korsgaard argues that nonhuman animals operate at a basic level of intentional action: "the animal is aware of his purposes, and thinks about how to pursue them. But he does not *choose to pursue* those purposes. The animal's purposes are given to him by his affective states: his emotions and his instinctual or learned desires."[32] Furthermore, when the animal appears to choose a course of action, the animal's choices are actually "made for him by the strength of his affective states."[33] Korsgaard goes on to argue that a rational animal is conscious of motivational states such as fear and desire, and is aware that she fears or desires a certain object. This awareness allows the agent to ask "should I be moved in this way?", and to no longer be moved by the strongest motivational or affective state at a given moment.[34]

Once an agent becomes aware of her motivational and affective states, and gets some mental distance on those states, this allows the agent to regulate or develop those very states in light of her self-conscious awareness of those states. Therefore, not only are human emotions (and other motivational states) robustly cognitive in a way that nonhuman animal states are not, but humans

---

[32] Korsgaard is of course a Kantian, but John McDowell—whose views are much more heavily influenced by Aristotle—makes a similar claim: "sentience is in the service of a mode of life that is structured exclusively by immediate biological imperatives" and that "merely animal life is shaped by goals whose control of the animal's behavior at a given moment is an immediate outcome of biological forces." McDowell, John Henry. 1994. *Mind and World*. Cambridge, Mass.: Harvard University Press, p. 115.

[33] Korsgaard, "Morality and the Distinctiveness of Human Action," p. 110. Please note that this is Korsgaard's gendered language.

[34] As MacIntyre puts the point, as responsible persons we can "evaluate or reasons as good or bad reasons and by so doing change our reasons for acting and in consequence our actions." MacIntyre, Alasdair C. 1999. *Dependent Rational Animals: Why Human Beings Need the Virtues*. Chicago, Ill.: Open Court, p. 72.

can intentionally transform those states. This potential—whether or not it is actualized by the agent—forms a key part of a neo-Aristotelian model of moral responsibility. Consider Aristotle's discussion of anger in the *Nicomachean Ethics*. Anger—like other Aristotelian virtues—is a mean between excessive and deficient motivational states. A certain amount of anger is proper in certain situations, but it is easy to have an excess of feeling. Aristotle says "the person who is angry at the right things and toward the right people, and also in the right way, at the right time, and for the right length of time, is praised."[35] Conversely, someone who sufficiently deviates from any of these means is blamed. Utilizing Aristotle's explanation of pity, we can deduce that proper anger will be at a person who has committed an intentional offense or harm. For anger to be appropriate, the agent must perceive the situation in a certain way, and believe that the offense was intentional and sufficiently significant. Human emotions are evaluative perceptions—perceptions of significance[36]—and the agent plays an active role in this evaluative state.[37]

De Waal would likely counter that primates also seem to perceive situations in a certain way, seem to be aware of other agent's mental states and probably even have beliefs about the situation since they sometimes respond to those perceived mental states in ways that we deem appropriate.[38] However, on Aristotle's account no one is naturally or effortlessly virtuous.[39] For Aristotle, virtue

---

[35] *NE* 1125b 32. Irwin's translation.

[36] This is Rick's Furtak's phrase. He is using it in reference to the Stoics and Kierkegaard, but I think that it is also appropriate for Aristotle. Furtak, Rick Anthony. 2005. *Wisdom in Love: Kierkegaard and the Ancient Quest for Emotional Integrity*. Notre Dame, Ind.: University of Notre Dame Press, p. 4.

[37] As Nussbaum notes, "...what is stressed is the fact that it is the way that things are seen by the agent, not the fact of the matter, that is instrumental in getting emotions going. Intentionality...is the issue." Nussbaum, *The Therapy of Desire,* p. 85.

[38] More recently de Waal gives a more complicated account of animal morality, what he calls "natural normativity." He claims—among other things—that primates engage in fair distribution, impulse control, and conflict resolution. Although I think this recent research does show that non-human animals can and do sometimes conform their behavior to norms, even de Waal admits that this behavior falls short of human morality: "One could argue that their (primates) behavior is normative in that it seeks certain outcomes, but that animals manage to do so without normative judgment. They may evaluate social behavior as successful or unsuccessful in furthering their goals, but not in terms of right or wrong." De Waal, Frans, 2014. "Natural Normativity: The 'Is' and 'Ought' of Animal Behavior." *Behaviour* 151 (2–3): 200.

[39] For an action or emotion to be virtuous - that is the appropriate recipient of moral praise - the action must be shaped or guided by *phronesis* or practical wisdom. See for instance *NE* 1144b 25-35.

requires the very kind of *self-conscious awareness* that Korsgaard described, because immediate emotional responses (such as too much anger at an insignificant or unintentional infraction) are not morally praiseworthy unless they are guided by practical wisdom. Therefore, the agent's emotional response must be either the result of intentional effort at the moment of affect, or the result of a long process of deliberation and habituation that has molded the agent's dispositional emotional responses. Having the right response to a personal offense means having the right amount of anger—an emotion—and this emotion is up to the agent because it stems from a deeply engrained character trait that results from intentional action and repetition over a long period of time. For these reasons emotions are morally praiseworthy or blameworthy.

V.   Conclusion

The emotions of a non-human primate are clearly intentional in that they purposely respond to objects, and that intentionality is more complex because primates seem to at least some times be aware of the mental states of others and purposely respond to those states. But part of the reason we don't hold primates morally accountable for their mental states is that *the emotional states of primates are given to them*; primates are not able to self-consciously regulate or develop those emotional states, which means that while primate emotional states are certainly more cognitive than the emotional states of many other animals, those states still lack the cognitive robustness of human emotions that is necessary for responsibility ascriptions. It is for this reason that de Waal was justified it putting the blame on Travis's owner while letting Travis off the hook.