Animal Moral Psychologies

Susana Monsó and Kristin Andrews


1. Animal morality?

In August 2018, newspapers around the world reported on the case of an orca nicknamed Tahlequah whose newborn calf died 30 minutes after birth. Tahlequah was witnessed carrying the dead body for 17 days and over 1,000 miles, in what the media called a ‘tour of grief’ (Cuthbert & Main, 2018). In May 2014, CCTV cameras in California captured the moment in which a four-year-old child was attacked and pulled off his bike by a neighborhood dog. At that moment, the boy’s family cat appeared and chased the dog away, thus saving the child. The family later reported to have adopted the cat five years ago, after she followed them home from the park, and that she had formed a strong bond with their son (Hooton, 2014). In December 2014, a monkey was reported to have saved the life of another monkey who had been electrocuted after walking on the electric wires at an Indian train station. The monkey was filmed apparently trying to revive her unconscious friend by rubbing, hitting, biting, and dipping her in water until, after some minutes, the stunned monkey at last showed signs of life (“Monkey saves dying friend,” 2014).

These sorts of observations have led academics and the public alike to ask whether morality is shared between humans and other animals. Some philosophers explicitly argue that morality is unique to humans, because moral agency requires capacities that are only
demonstrated in our species, such as self-awareness, reflective scrutiny, the capacity to construct and act according to rules, normative self-government, or moral concepts (e.g. Kagan, 2000; Kitcher, 2006; Korsgaard, 2006, 2018; Dixon, 2005). Other philosophers argue that some animals can participate in morality because they possess these capacities in a rudimentary form (Sapontzis, 1987; Pluhar, 1995; DeGrazia, 1996), or because they are moral subjects whose emotional reactions reliably track objective moral facts (Rowlands, 2012).

Scientists have also joined the discussion, and their views are just as varied as the philosophers’ (e.g. Ayala, 2010; Hauser, 2006; Bekoff & Pierce, 2009; Tomasello, 2016; de Waal, 2006, 2009, 2013). Some research programs examine whether animals countenance specific human norms, such as fairness. While one research group found that animals do so (e.g. Brosnan and de Waal, 2003), another found that they do not (e.g. Jensen et al., 2007). Other research programs investigate the cognitive and affective capacities thought to be necessary for morality, from Tomasello’s (2016) claim that animal lack the distinctive sorts of joint intentionality and cooperation required for moral agency to de Waal’s (2013) insistence that morality is continuous between humans and other animals in the domains of empathy and reciprocity.

There are two sets of concerns that can be raised by these debates. They sometimes suffer from there being no agreed upon theory of morality and no clear account of whether there is a demarcation between moral and social behavior; that is, they lack a proper philosophical foundation. They also sometimes suffer from there being disagreement about the psychological capacities evident in animals. And at their worst, some views suffer from both. For example, Ayala (2010, p. 9015) argues that animals lack morality because they lack what he identifies as the three necessary conditions for being a moral agent that “exist as a consequence of the
eminent intellectual capacity of human beings”: the ability to anticipate the consequences of one’s own actions, the ability to make value judgments, and the ability to choose between alternative courses of action. The notion that humans alone have these three capacities is dubious (as will be demonstrated), and we suspect that many ethicists would be unmoved by the relationship between these capacities and a moral sense, as they may be only trivially necessary, and nowhere near jointly sufficient. Ayala’s three abilities are at least as relevant to deciding whether to buy a house or which university to attend as they are to saving a drowning child or not stealing from the tip jar.

Of these two sets of concerns—the nature of the moral and the scope of psychological capacities—we aim to take on only the second. In this chapter we will defend the claim that animals have three sets of capacities that, on some views, are taken as necessary and foundational for moral judgment and action. These are capacities of care, capacities of autonomy, and normative capacities. Care, we argue, is widely found among social animals. Autonomy and normativity are more recent topics of empirical investigation, so while there is less evidence of these capacities at this point in our developing scientific knowledge, the current data is strongly suggestive.

We recognize that some of these capacities are not themselves uncontroversially defined or operationalized and that the science is far from complete on any of these issues. Despite these problems, we think it important to track the current evidence for these capacities in nonhuman animals. The better handle we have on the scope and extension of the psychological capacities that may be implicated in moral practice, the better we will be able to understand the nature of moral practice. This is important for our self-understanding as a species, as well as for determining whether and to what extent animals can be said to be moral agents. In addition, it is
also crucial for determining what it means for an animal to lead a good life, and therefore has implications for the moral status of animals. We hope that this investigation will also help us determine what we owe to other animals (see Monsó et al., 2018).

2. Capacities of Care

Care is widely, though not universally, identified as a significant aspect of morality. It is one of the foundations in Moral Foundations Theory (Haidt and Graham, 2011), and it is key to the ethics of care (Gilligan, 2016; Noddings, 2003; Tronto, 1994), as well as to sentimentalist moral theories (Prinz, 2009; Nichols, 2004; Gibbard, 1998). The role of care in the theories reflects our intuitive sense that care is a crucial aspect of morality. One of the most profound moral criticisms we can offer is “You don’t care about other people.” Nothing can make up for not caring, not a good sense of humor, a hard-working nature, or artistic genius. Care can be defined in terms of identifying and meeting the needs of others with whom we are in relationships, as well as being affected by the plight of others. Caregiving, in this sense, is seen widely among social animals who require lengthy parental supervision (Hrdy, 2011). A particularly moral sense of care would require, in addition to caregiving behavior, moral emotions or other forms of moral motivation that underlie the caring behavior. Appropriate caring emotions may include empathy, sympathy, compassion, grief, or love (see, e.g. Nussbaum, 2001; Slote, 2007).

Scientists have long had interest in looking for the capacities of care in other species. The recent focus of care in animals has been bolstered by the work of Frans de Waal and his students, who have been investigating capacities of care in chimpanzees and other mammals. The first form of care de Waal focused on was consolation behavior in chimpanzees—comforting

---

1 Thanks to John Doris for making this point obvious to us.
behavior directed at an individual in distress, often in the aftermath of a conflict or a fight. De Waal observed this behavior in chimpanzees and reported on it in his first book (de Waal, 2007), and subsequent work has found consolation behaviors in a number of species (see Table 1). Another recent trend is to examine mourning behavior in animals. Mourning is a term that is used to refer to expressions of distress over the death of an individual, manifested in, for example, attempts to elicit a response from a corpse or insistent returning to the carcass or carrying the body. Scientists have observed apparent examples of mourning behaviors in a wide range of species (see Table 1). A third topic of interest has been on helping behaviors, which refers to cases in which an animal intentionally benefits another individual. Scientists have argued for this behavior in a number of species as well (see Table 1). Such behaviors support the claim that we see empathy in other animals (e.g. Andrews and Gruen 2014; Gruen 2015; Monsó 2015).

In order to discuss some of the interpretational debates surrounding the evidence of care capacities in animals, we will focus on the tradition of research on care in rats (and some other rodents), which encompasses studies on consolation and helping (to the authors’ best knowledge, there is as yet no evidence of mourning behavior in rodents). Given that rats are only distantly related to the ‘usual suspects’ (apes, corvids, cetaceans, and elephants), we think that focusing on the evidence for care capacities in the rat will help to bolster the evidence for care capacities in these other taxa, which are primarily studied using behavioral methods. In the rat literature, scientists have not just adopted behavioral approaches, but also neurobiological and neurochemical methods.

Before delving into the debate, we would like to briefly note that the methodologies used can be heavily criticized from an ethical perspective. The studies that use neurobiological and
neurochemical methods are often very invasive and can leave the rats permanently impaired in their social capacities. Even the purely behavioral studies have often involved deliberately placing the rats in very stressful situations. It is, of course, no coincidence that animals who are deemed less cognitively sophisticated are also subjected to much more invasive and stressful methodologies. As Hernandez-Lallement et al. (2018) candidly state in their defense of a rodent model of callousness, “[r]odents offer a cheap, convenient and ethically less controversial alternative to non-human primate [sic] in the study of social cognition” (p. 124). While we will bracket this issue in what follows, it is our hope that reflecting on what these studies teach us about the morality of these animals will also help us reconsider the morality of some of these experiments (see also Monsó et al., 2018).

The rat care research can be traced back to the 1950s, when scientists found that rats refused to press a lever to obtain food when doing so shocked a rat in an adjacent cage (Church, 1959). (Similar results were later obtained with pigeons [Watanabe & Ono, 1986] and rhesus macaques [Wechkin et al., 1964] — see Table 1). In his book on the evolution of morality, Marc Hauser argued that this behavior is likely due to an egoistic motivation, rather than a caring one (Hauser 2006, pp. 353ff.). The animals, he argued, may have been simply experiencing the other’s pain as an aversive stimulus, or acting out of a fear of retribution. Although the same debunking arguments can and have been made for humans, which are the uncontroversially ‘moral animals’ (see Batson 2011), subsequent research has attempted to address this worry by trying to determine what motivations underlie this behavior, with some scientists arguing that rats in these situations are behaving on the basis of the caring emotion of empathy. That is, rats are said to recognize and share the emotion of other rats, and that this is what motivates them to help the other in need.
This recent spate of “rat empathy” studies started in 2011, when rats were found to release a conspecific who was trapped in a restrainer. If there was no rat or only a fake rat in the restrainer, the free rats generally didn’t open the restrainer, and when they did it was at a much higher latency (Bartal et al. 2011). The authors claimed that the helping rat acted altruistically, since when a second restrainer full of chocolate chips was placed into the compartment, the helping rat would tend to open both restrainers and share the food with her cagemate. However, it is still possible that the helping rat was acting out of an egoistic desire for social contact. This desire might have been stronger than the desire to have the chocolate all for herself, but still egoistic in nature. In order to rule out this social contact hypothesis, the test conditions were modified so that the trapped rat was released into a different compartment, thus preventing the helping rat from interacting with her upon release. The authors found that the rats would continue freeing their cagemate. Bartal et al. concluded that they were “empathically sensitive to another rat’s distress” and “acted intentionally to liberate a trapped conspecific” (Ibid., p. 1430). The media had a field day with this study, with The Washington Post calling the rat a “new model of empathy” (Brown, 2011).

The studies continued, finding that rats who had an unpleasant experience were more likely to help a cagemate rat in that same unpleasant experience. Sato et al. (2015) placed a rat into a compartment that was half-way full of water, which would cause her to display distress. A second rat—the experimental subject, who had been previously housed with the target rat for two weeks—was then placed into an adjacent compartment that had an elevated floor and a door to the first compartment. The walls between both compartments were transparent, and the subject had the possibility of opening the door, thus allowing her conspecific to exit the water. The experimenters found that 90% of the subjects displayed door-opening behavior, at a latency that
decreased across trial sessions. When the roles were switched, the rats who had previously been in the water opened the door at an even shorter latency. Furthermore, when the conspecific was not in water, and thus not displaying distress behavior, only one experimental subject displayed door-opening behavior. The authors interpreted these results as providing evidence that the subjects were motivated by a desire to rescue the rat from the water, as well as evidence of empathy that was facilitated by what animal cognition researchers called “experience projection,” improved interaction given experience with the situation some other is in, since rats’ own prior experience in the water compartment increased the speed at which they freed their distressed cagemate.

The claim that these studies demonstrate care in rats, in that rats have the right kinds of emotions driving their behavior, has been controversial. Other explanations for these behaviors have been suggested. Silberberg et al. (2014), for instance, ran a follow-up experiment to Bartal et al.’s 2011 to further test the social contact hypothesis. In order to do so, they subjected pairs of rats to a test consisting of three consecutive conditions (i.e., performed as serial trials with the same rats). In the first condition, opening the restrainer would release the trapped rat into a separate compartment, where social contact would not be possible. In the second condition, opening the restrainer would release the trapped rat into the same compartment as the free rat’s. The authors found that the free rat would open the restrainer in these two conditions, but in the asocial condition the latency would increase across trials, whereas in the social condition it would decrease. After the social condition came a third condition, in which the trapped rat was once again released into a separate compartment, thus returning to the setting in the asocial condition. The researchers found that the door-opening latency and response frequencies across trials in this last condition remained as short as they were in the second, social condition. In this
last condition, the trapped rat was also often observed returning to the restraining tube after having been released into the distal chamber.

Silberberg et al. argued that these results are incompatible with an explanation in terms of empathy, because if that were the sole motivation of the free rats, then response frequencies and latencies should have been very similar across trials. Instead, they argued that what was likely operating in the free rats was a desire for social contact, which is why response frequencies and latencies increased across trials in the first asocial condition and decreased across trials in the social condition. The authors explained the results of the third condition, where the setting returned to the asocial condition but the free rats performance was indistinguishable from the social condition, by arguing that rats are usually neophobic, which means that the trapped rat would have been afraid of the apparatus in the first condition, but that by the third condition this fear would have been extinguished. This is presumably why the trapped rats were observed returning to the restrainer after release, and also why the response frequencies of the free rat remained high, since they learned that by staying close to the restrainer they could spend more time with the other rat. This alternative explanation posits a motivation in the free rats that is still likely emotional in nature, but the emotions involved are presumably not the right sort of emotions for it to be an explanation in terms of care, since their focus is the free rat’s own interests (i.e. she wants the trapped rat to be released into her own compartment because she herself wants social contact).

However, the behavior of the free rats in this experiment is especially difficult to interpret, due to the fact that the restrainer was operated by a sensor, so that being in close proximity to it was enough for the trapped rat to be released. This means that we cannot know how much of the restrainer-opening behavior was actually intentional. In contrast, the restrainer used in the
original Bartal et al. 2011 study required the rats to purposefully interact with it in order for it to open, which means that accidental, non-intentional openings were highly unlikely. In addition, it should be noted that in Silberberg et al.’s experiment the free rats tended to open the restrainer in the different conditions, even if latency and frequency varied across trials, which suggests an interest in releasing the trapped rat.

While debates continue about the interpretation of the data, there has been little discussion about whether the social contact hypothesis actually undermines claims about rat empathy. Humans have mixed motives for our actions; if we free a prisoner in part because we want social contact, it doesn’t follow that the act was amoral, because we might also think that the person should be freed. There can always be a combination of motivations operating in the rats, too, so they could be acting out of a desire for social contact plus empathy for the other’s condition. Even if the motivation for social contact is stronger than the empathic motivation, this does not exclude the possibility that the rats were indeed caring for the other. Evidence that rats do care for the other is provided by Schwartz et al. (2017), who gave rats a choice between liberating a conspecific who was in a container full of water or one which was in a regular Plexiglas enclosure. They found that the rats consistently preferred to liberate the soaked rat, which the authors explained in terms of a desire for social contact plus an interest in the wet container. However, if rats are indeed neophobic and strongly driven towards social contact, one would expect them to prefer to liberate the dry conspecific, or if their only motivation was a desire for social contact they should have been just as likely to liberate either of the conspecifics across the different trials. The fact that they strongly preferred to free the wet rat suggests that they could be motivated by both a desire for social contact and empathy, or some other form of care.
Another objection is that the results of these experiments are evidence of prosociality, not care (Vasconcelos et al., 2012). The idea is that the evidence only shows that rats are predisposed to help conspecifics in need, and that all that this allows us to postulate in them is a biological mechanism whose function is to benefit others in the social group in order to ultimately promote the fitness of the individual. In a similar vein, Alex Kacelnik argues that the Bartal et al. 2011 study does not require an explanation in terms of empathy, since “the reproductive benefits of this kind of behavior are relatively well understood as, in nature, they are helping individuals to which they are likely to be genetically related or whose survival is otherwise beneficial to the actor… To prove empathy any experiment must show an individual understands another’s feelings and is driven by the psychological goal of improving another’s wellbeing. Our view is that, so far, there is no proof of this outside of humans” (Kacelnik, 2012). The requirement that a proof of empathy must come with evidence that “an individual understands another’s feelings” sets the bar quite high. It is likely that all that empathy requires is some form of behavior-reading, and not a theory of mind (see Monsó, 2015; 2017). Still, the worry remains that all that these experiments show is that rats are prone to prosociality, that is, that they engage in behaviors that benefit others, but not that they are operating on the basis of a caring motivation, such as empathy. In order to make this point, the Bartal et al. studies are often contrasted with those performed by Nowbahari et al., who showed that ants will also reliably free nest-mates who are trapped in a snare (Nowbahari et al. 2009). For many, it seems unlikely that ants have moral capacities, so the fact that they can still engage in helping behaviors gives rise to the question, what evidence is there that these rat behaviors are caused by the right mechanism to count as care?
To answer this question, scientists have turned to methods involving brain lesions and drug interventions in order to block emotional responses in the rats. If rats with impaired emotions show impaired helping behavior, we have evidence that the mechanism involved in helping is at least in part an emotional one.

Recent studies show that physical and chemical interventions do impact rat helping behavior. Rats with lesioned amygdalae behave differently than intact rats in care-related tasks. For example, in the prosocial choice task, rats can choose to secure food for themselves and a partner or secure food only for themselves. Intact rats (Hernandez-Lallement et al., 2015; Márquez et al., 2015), as well as capuchin monkeys (Lakshminarayanan and Santos, 2008), common marmosets (Burkart et al., 2007), cotton top tamarins (Cronin et al., 2010), and chimpanzees (Horner et al., 2011, but see Silk et al., 2005 for a negative report), have been found to prefer provisioning a partner as well as the self. However, when a rat’s amygdala is damaged, her responses change significantly, and she fails to reliably choose the mutual reward option (Hernandez-Lallement et al., 2016). Amongst other functions, the amygdala is involved in emotional processing and social cognition. Non-functioning amygdalae have been associated with impairments in affiliative behaviors and sensitivity to social cues, and with psychopathic traits in humans (Hernandez-Lallement et al., 2018). Accordingly, the authors interpreted their results as showing that damaging their amygdala induced in the rats “[a] deficit in attaching affective salience to social cues,” resulting in “a general insensitivity to the affective value of social information” (Hernandez-Lallement et al. 2016, p. 8). In other words, while the damaged rats could possibly still process the social signals emitted by the partner, they could no longer emotionally experience them as rewarding or aversive, which was presumably needed to motivate them to choose the prosocial option. This suggests that the rats’ care behavior is driven
by emotions and that, by lesioning the seat of the emotions, rats were no longer able to care for their cagemates.

Chemical interventionals also impact rat care behavior. When scientists administer anxiolytics to rats, their helping behavior is reduced (Bartal et al., 2016). The authors of this study used a benzodiazepine anxiolytic that “acts in the brain to reduce anxiety, which in turn reduces ... sympathetic activation” (Ibid., p. 2). While unimpaired rats and rats injected with a saline solution would reliably free a trapped conspecific, anxiolytic-treated rats did not free the trapped rats, even though they would still open a restrainer door to gain access to chocolate. This pattern suggests that their reluctance to free their conspecific was not due to the anxiolytic having general sedative effects. Given that the anxiolytic impairs emotional capacities, we have further evidence that emotion motivates the helping behavior. This was indeed the interpretation of the authors, who concluded that the anxiolytic “interfered specifically with social affective processing that appears necessary to motivate a free rat to help a trapped rat” and that the results of the study “support the idea that affective resonance between helper and victim rats is responsible for motivating pro-social actions” (Ibid., p. 10).

It isn’t just rats that show an impairment in care behavior after drug interventions. Another set of experiments demonstrate the role of emotion in care behavior among prairie-voles, rodents who form life-long pair bonds. Pairs of bonded prairie voles were separated, and one of them (the demonstrator) was either left in isolation, or subjected to a form of Pavlovian fear conditioning, by presenting her with tones followed by light electric shocks delivered to her feet (Burkett et al., 2016). The other vole (the observer) was then returned to the cage, and her spontaneous behavior was observed. It was found that observers licked and groomed the demonstrator for a longer period of time and following a shorter latency when the latter had
undergone the shocks, than they did during the control condition. This licking and grooming was found to have an alleviation effect on the demonstrator’s distress, as she displayed less anxiety-related behaviors than when she was left alone after the shocks. This response on behalf of observers was thought to be triggered by oxytocin, for those subjects who were injected with an oxytocin antagonist did not show a consolation response. Given that, in humans, oxytocin is involved in empathy and socioemotional engagement, the authors interpret these results as suggesting “conserved biological mechanisms for consolation behavior between prairie vole and human” (p. 378).

One might object that the emotion that is turned off in these studies is not a care emotion, but rather a negative emotion caused by the conspecific. According to this interpretation, a vocalizing trapped conspecific is annoying, and the rats act to release trapped conspecifics to eliminate the annoying stimulus, not to offer care toward them. Rats who are trained that they can turn off sound recordings by pressing a lever will do so to stop a recording of a rat cries, for example (Lavery & Foley, 1963). Perhaps releasing the trapped conspecific serves the same purpose. The rats may be emotionally motivated (which is why the surgery and the drug interventions impair their behavior), but the emotion that is driving the behavior is self-centred. It is the wrong kind of emotion for their reaction to be labeled as an instance of care. Instead, the rats would more appropriately be said to be helping the other in order to escape from a situation that they themselves find aversive (Silberberg et al., 2014, p. 610; Schwartz et al., 2017, p. 299).

This objection raises the spectre of a simplistic type of psychological egoist objection to moral action according to which all action is motivated by desire, and all desires are, by definition, selfish. A straightforward response to this objection is that while all my desires may be selfish in the sense that they are mine, there is no evidence that all desires are self-directed
and that none of my desires are other-directed (see Tiberus, 2014 for a discussion of this objection and response). And as we already pointed out, mixed motives are always possible. In addition, as Monsó (2015; 2017) has argued, following a suggestion by Rowlands (2012, p. 11), rat aversion to the crying of conspecifics does not preclude an explanation in terms of care anymore than humans who find crying babies annoying cannot be said to be acting morally when helping an injured infant. In fact, finding something unpleasant may be in itself a moral reaction, if what is functioning in the individual in question is a moral emotion that has this unpleasantness built into its phenomenal character. That is, for the rats in these studies, caring for an individual in need might precisely imply experiencing her distress cries as an unpleasant stimulus.

Further insight into the motivations of the helping rats was provided by a very recent study that was modeled after Batson’s “aversive-arousal” studies on humans. In one of Batson’s studies, participants watch a video of a person who appears to be suffering from electric shocks, and are told that they could take her place if they choose to. Half of the participants have been primed to be empathetic, and the other half have not. Half of each of those groups can easily escape the situation, and the other half cannot. Humans in the high empathy conditions choose to help, regardless of how easy or difficult it is to escape the situation (Batson et al., 1981). In a rat version of the aversive-arousal experiment, Carvalheiro et al. (forthcoming) modified the original Bartal et al. 2011 test to give the free rat the option to escape to a dark compartment instead of opening the restrainer. The free rats were first given the chance to familiarize themselves with the arena, which consisted of a compartment with glass walls and a door to an adjacent dark compartment. Then, in the test condition, the restrainer with a trapped rat was placed into the lit compartment and the door to the dark compartment was either open or closed.
Each pair of rats was tested for twelve consecutive days in the same condition (i.e., with either an escape option or not). In theory, this test would have the capacity to disentangle a caring response from a purely selfish one. If the free rats were to open the restrainer in both conditions, then they would be clearly motivated to free the trapped rat. If they were to only open the restrainer when there was no possibility to escape, then we would know that they were only motivated by an egoistic desire to calm their own distress.² The researchers found that rats in the no-escape condition opened the door significantly more often and at a shorter latency than rats in the escape condition. While this initially seems to support the egoistic hypothesis, a closer look at the data reveals that things are not as straightforward as this.

To begin with, the door-opening frequency increased and the latency decreased across trials in both conditions and by the twelfth day every free rat was releasing her conspecific, so there was some motivation to free the other rat in both conditions. In fact, the authors describe this as “surprising,” since they expected decreased helping in the escape condition. In addition, each free rat was only tested in one of the conditions, and in neither case did they have previous experience opening the restrainer. In the no-escape condition, there was nothing else to do in the arena, so the free rats had more incentive to explore and find out sooner how to open the restrainer. In the escape condition, the rats had the option of escaping what would surely be perceived as a threatening environment into one that they would feel as safer (rats have a strong

² Note that this is only in theory. In practice, moral creatures may sometimes choose to escape instead of helping, especially when this can be done at little or no cost. For instance, some humans who are otherwise caring may nevertheless look the other way when a beggar approaches them on the street.
preference for dark places). As the days went by, the rats would presumably learn that they were in no real danger and could then attend to their trapped cagemate.

We know from studies performed on humans that too much personal distress is detrimental to helping behavior (Batson et al., 1983). While a certain level of distress appears necessary for the rats to help (Bartal et al., 2016), too much may be overwhelming and paralyzing, just as in humans. A close analysis of the rats’ behavior supports this interpretation. Studies have shown that rats are prone to emotional contagion, i.e. the spontaneous ‘catching’ of another’s emotion (Atsak et al., 2011; Kiyokawa et al., 2019). The distress displayed by the trapped rat in the first trials, coupled with the brightly lit area, could have evoked too much distress in the free rat. In later trials, when both the free rat and the trapped rat were displaying less behaviors associated with stress, the free rats were more likely to open the restrainer. So, the effect of stress on helping behavior seems to follow an inverted U-shape, with very high and very low levels associated with less helping.

Further studies could continue probing into the helping rats’ motivation. Carvalheiro et al., for instance, suggest making both compartments dark, to disentangle the stress evoked by the trapped rat from that derived from the light. However, we believe that there is sufficient evidence to support the claim that rats are capable of acting on caring motivations. The fact that their motivations aren’t always straightforward and may be mixed with selfish inclinations does not undermine the hypothesis that rats can be caring. Rats, like humans, may be moral without being saints.

3. Capacities of Autonomy
Autonomy is another capacity that is often cited as necessary for an individual to participate in moral practice. While the autonomy requirement is perhaps most closely associated with Kantian moral theories, it is also an important aspect of liberal approaches to morality, as well as to some feminist approaches (Christman, 2018). However, what autonomy amounts to varies widely in these literatures, and it shares an uncomfortable connection with the free will debate. At its most basic, we can describe autonomy as the ability to act flexibly. This description requires that an autonomous individual be cognitively flexible—able to respond differently toward the same set of stimuli, such that their behavior isn’t determined by observable external factors. Flexible action is widely seen across animal taxa, and the recognition of animal flexibility helped to motivate the cognitive revolution in animal behavior studies. While a minimal requirement for autonomy, it is worth recognizing that we find evidence of flexible action in birds, mammals, reptiles, fish, and cephalopods.

We can also describe autonomy as the ability to act authentically, in that the action is coming from the agent’s self, or is otherwise the agent’s own action. In this sense, an autonomous agent is not a wanton who is driven by occurrent first-order desires (Frankfurt, 1971), but rather someone who can adjudicate between competing desires. This aspect of autonomy requires more than Frankfurtian identification. It requires the capacity of self-control, or inhibition of an initial response. There is evidence for self-control and inhibition in animals from rats and pigeons to great apes, which we will review below.

Finally, we can describe autonomy in the Kantian sense of acting on reasons that have been subject to self-scrutiny. This aspect of autonomy requires metacognitive access to one’s own mental states. However, it has been argued that even this is not enough to possess full normative self-government, or the highest level of autonomy (e.g. Korsgaard, 2006). A fully autonomous
individual can not only access her own motivations, but also reflect on the very reasons that drive her behavior, and ask herself whether these are reasons that are worth pursuing; whether they correspond to the sort of being she wants to be. Normative self-government, thus understood, is very intellectually demanding, and it is unlikely that any animal can engage in it (indeed, we question how often humans actually engage in such reflections, see e.g. Doris, 2015). Nevertheless, it seems plausible that an animal who has metacognitive access to her own mental states will have some level of control over her behavior, in the sense that metacognition in animals may precisely have the function of enabling them to choose which course of action better accommodates their current desires (though see Rowlands, 2012 for arguments against the link between metacognition and control).

The capacities of autonomy have been studied primarily in the ‘usual suspects’, though that is changing. While the discussion that follows will focus on primates, there is growing evidence of self-control in other taxa. We shall divide the available evidence into two groups that correspond to the capacities of autonomy that have been studied the most in these and other animals: self-control and metacognition.

3.1 Self-control

The orangutan “Princess” lived in a human-orangutan community at Camp Leakey since early infancy. One afternoon, Princess was socializing with human visitors on their bunkhouse porch, and after a while she left, climbing up the bunkhouse wall and to the roof peak. But, rather than continuing on her way, Princess turned around and climbed back to the porch and stayed there all day. She was still on the porch at 8:00 pm, apparently asleep, when the humans living there locked the door, stepped over her, and left for dinner. Two hours later, the humans
returned to find that Princess was inside the bunkhouse, their suitcases ransacked and food gone. Princess likely entered via torn screening at the top end of the wall at the roof peak, where she had been early that morning.

Psychologist Anne Russon, who described this incident, says, “Princess probably noticed the break-in spot in the afternoon, when she went to the roof peak, but knew she had little hope of entering with humans present because they regularly foiled such attempts. She began behaving atypically right after pausing at the break-in spot. She feigned indifference to this spot and even moved away from it. She became noticeably more friendly to humans at the bunkhouse, probably to create a false image of her reasons for staying; this likely increased her chances of monitoring the scene without detection. She waited until she knew humans would be inattentive to break in, the nightly dinner hour” (Russon, personal communication). This example of delayed gratification and deception shows Princess’s capacity for self-control in an unusual environment for the typical orangutan, but it reflects a species typical capacity. Great apes in the wild have been observed to find food, then wait until no one is around to get it, even moving away from the food and postponing their feast by several hours (Byrne & Whiten, 1988).

This observation of self-control in a nonhuman animal occurred in the context of deception. Deception is a natural place to look for the ability for self-control, as it requires acting against some normal responses. In a classic study, Richard Byrne and Andrew Whiten collected via a survey of researchers working with primates cases of tactical deception, defined as “acts from the normal repertoire of the agent, deployed such that another individual is likely to misinterpret what the acts signify, to the advantage of the agent” (Byrne & Whiten 1988, p. 661). They found cases of deception in all primate species examined. Primates have been observed to restrain from food calls, sex calls, and natural facial expressions, in order to hide their behavior,
goals, or emotions from another. More recently, experiments and observations on deception have reported flexible tactical deception in species including dogs (Heberlein et al., 2017), corvids (Clayton et al., 2007) and cephalopods (Brown, 2012). In this last case, cuttlefish, who can change their shape and color, were observed tactically using this ability to deceive competitors during courtship. On one side of their body, males would display male courtship patterns to an attentive female, while on the other side they would simultaneously display female patterns to a rival male. This strategy of disguising as a female is used to prevent competitors from interrupting the courtship, but the cuttlefish only use it when there is a single male observing them, which reduces the risk of their deception being discovered.

Experimental self-control tasks with nonhuman animals take a variety of forms, and have been performed on a number of different species. One of the most common task types is the delay of gratification type tasks, inspired by the classic marshmallow test. In the original task, young children are offered a choice between a small reward now and a larger reward later. Some children have more difficulty than others at waiting to gain a second treat, and those children who can delay gratification display a number of strategies to distract themselves, such as turning around so they couldn’t see the marshmallow. Waiting for the two treats is interpreted as evidence of self-control (Mischel, 1958; Mischel & Ebbesen, 1970).

Parallel studies have been performed with a number of different species, finding similar results. Pigeons were trained that they could either gain access to a hopper of palatable but subpar food now, or that they could wait to gain access to a preferred food later. Like human children, some pigeons did delay gratification (Grosch & Neuringer, 1981). Like children, the pigeons were better at the task when food items were not visible, and when a distractor object was made available. When pigeons’ attention was drawn to the food items, pigeons did worse on
the task, just as children did when they were instructed to think about the treat during the waiting period. In another study, rats were trained that they could access a hopper that was accumulating food pellets, and the subjects demonstrated the ability to delay gratification so as to acquire a greater amount of food (Killeen et al., 1981).

More recently, primate researchers began to examine delay of gratification in great apes and monkeys. Language-trained and naïve chimpanzees were able to restrain from pushing a doorbell to get a small treat, and wait until the end of a trial in order to receive a larger reward (Beran et al., 1999). Using an accumulation task similar to the one in the rat study, four chimpanzees and an orangutan demonstrated the ability to wait until all 20 chocolate pieces were placed into their bowl, delaying response for 60 to 180 seconds (Beran, 2002). Like the children and the pigeons, apes are also able to wait longer when there is a distractor object in their enclosure (Beran et al., 2014; Evans & Beran, 2007). Rhesus macaques (Evans, 2007) and capuchin monkeys (Bramlett et al., 2012) also show the ability to delay gratification for a larger reward.

Other studies examined self-control or self-regulatory capacities in a range of mammalian and avian species using two tasks as the measures for self-control: the A-not-B task and the cylinder task (MacLean et al., 2014). The A-not-B task examines subjects’ ability to switch to a new behavior, when there is reason to choose the old behavior as well as reason to choose the new behavior. This test can be understood as a choice between two competing desires. The cylinder task invites subjects to access a reward that is visible through the transparent material of the cylinder, and the question is whether or not the subjects can inhibit their response to move directly toward the reward, and instead move away from the reward to successfully access it through the open end of the cylinder. MacLean and colleagues examined 36 species on these two
tasks, and found evidence that brain mass tracks performance, in that great apes had better self-control than did monkeys and carnivores, who in turn were better than lemurs, and below them were rats and birds. However, the idea that self-control relates to absolute brain mass is disputed by Kabadayi and colleagues, who tested additional avian species, and found that ravens, New Caledonian crows, and jackdaws’ performance on the cylinder task was indistinguishable from that of the great apes (Kabadayi et al., 2016).

In the animal literature, debates continue about the kinds of self-control that exist, and the different methods of studying self-control. In addition, claims that animals have self-control based on their performance in a single kind of test may be hasty, as a general capacity for self-control should be displayed in a variety of contexts. Likewise, failure to find self-control in a single domain isn’t evidence that the individual (or species) has no capacity for self-control; it may be a domain-specific failure. One may be bad at resisting desserts, but good at resisting moral violations. Likewise, the findings that many species have difficulty with reversal learning (e.g., selecting the set of objects that one doesn’t want, or selecting a smaller array to gain a larger one) shouldn’t lead to the conclusion that such animals have a generalizable failure of self-control. That being said, given what we do currently see in nonhuman animals, there is evidence

3 As in many studies of associations between brain size and behavior, there are outliers. The Asian elephant, for example, had the lowest score on the A-not-B task. Whether this speaks against the brain-size hypothesis or raises methodological questions is not directly relevant to the point here, namely that at least some other species demonstrate self-control.

4 For a recent review of the self-control literature in humans and other animals, see Michael Beran’s book Self-Control in Animals and People (2018).
that other species have at least something akin to autonomy as the ability to act authentically, in that the action is coming from the agent’s self, given the ability to restrain from acting on initial impulses. Others animals, like humans, are not mere wantons. Acting on one’s first impulse is not going to be an effective strategy for a species living in a complex physical and social environment, so it should not be too surprising that other animals also have the ability to modulate their impulses—to some extent.

3.2 Metacognition

In the animal literature, metacognition is typically understood along the lines of the ability to monitor and control one’s own cognitive processes (Basile et al., 2015). This is in contrast to the theoretically-laden approach of some philosophers who understand metacognition as requiring metarepresentation (Carruthers, 2009; Bermúdez, 2003), but in keeping with other philosophical accounts, such as Joëlle Proust’s nonpropositional affordance-sensing account (Proust, 2013). This review will remain silent on the debate about the representational nature of metacognition and related questions about the vehicle needed for metacognitive representation (e.g. language, concepts, or analog representations) and instead will sketch the evidence that animals are capable of accessing, evaluating, and controlling their cognitive processes.

The capacity for metacognition is implicated in a number of different kinds of actions. Knowing what they don’t know allows an agent to realize when they need to seek additional information, to critically evaluate their own decisions and past actions, and to evaluate their current motivational state. It has a social role, permitting a kind of perspective-taking so as to consider how others might see one, and the metacognitive capacity has been referenced for its
role in the creation of cumulative culture, learning, and teaching (e.g. Heyes, 2018; Sterelny, 2012; Tomasello, 2016).

We can discuss two kinds of tasks that have been given to animals to examine their ability for metacognition: seeking information before acting and opting out from difficult tasks. The seeking information before acting studies are designed to test whether a subject knows what they need to know in order to achieve a goal. A common paradigm used to test this ability in great apes offers subjects the opportunity to check their answer before acting. If the apes know that they know the answer to a puzzle, and checking is costly, we should predict that the apes won’t check. But if we vary the cost of checking, making it easy, we should predict that apes would check more often. (This checking behavior is likened to human behavior before international flights—we might often pat our pocket or look into our bag to make sure that the passport is there, when checking is easy and the cost of not having it is high.) One way this has been tested in orangutans is to give them a simple disjunctive problem by hiding a treat under one of two upside-down cups. Then the subject is shown that one cup is empty. If the cups are on a transparent barrier and the barrier is easy to duck under, subjects can check which cup is hiding the treat, and they do. If the barrier is low and harder to duck under, subjects are less likely to check, and make the choice based on an inference. Finally, if another cup is added and subjects don’t have enough information to solve the task, they check regardless of the cost (Marsh and Macdonald 2012). Versions of this task have demonstrated the checking behavior in great ape subjects (gorillas, chimpanzees, bonobos, and orangutans) (Call, 2010) as well as monkeys (Serman & Smith, 2011; Marsh, 2014).

Another type of metacognitive task is the uncertainty-monitoring task, which gives subjects the opportunity to reject difficult perceptual or memory tasks. In this sort of task,
subjects can choose not to take a test, and gain a small reward, or choose to take the task and risk gaining nothing if they fail it, but with the opportunity for a better reward if they pass it. Macaques (Hampton, 2001; Basile et al., 2015; Smith et al., 1997), pigeons (Sole et al., 2003), dolphins (Smith et al., 1995), and orangutans (Suda-King, 2008) have been found to pass this sort of test. Here, as with the self-control experiments, we stress the importance of not relying on any one test to provide thorough evidence of a cognitive capacity. For example, in the pigeon study, subjects were asked to discriminate between sparse and dense visual arrays, and the authors note that the pigeons could have solved that task without metacognition, by associating the hard to classify stimuli with the uncertainty response. In contrast, in the Basile et al. study with macaques, seven different experiments were run using the uncertainty-monitoring methodology, each of which was designed to test for a different alternative hypothesis, thus providing stronger evidence of metacognition in this species. Given the current science, there is evidence that at least some other animals demonstrate something akin to autonomy understood as acting on reasons that have been subject to self-scrutiny.

4. Capacities of normativity

Creatures who care for one another and who have agential control over their own behaviors have mental capacities that can play important roles in an agent’s moral psychology. As we saw above, some of these capacities appear to be common across animal species. For creatures who have capacities of care and autonomy, we can ask a further question—do they have capacities of normativity? That is, do they demonstrate norm-guided behavior? There has been growing

5 A good anthology that covers the evidence and debate about metacognition in animals is *Foundations of Metacognition* (Beran et al., 2012).
interest among scientists in the possible existence of social norms in nonhuman animal societies, and in particular in the great apes. Insofar as moral norms are a subset of social norms, uncovering social norms in animal societies can illuminate the question raised at the beginning of the chapter as to whether nonhuman animals participate in moral practice of some sort. Two lines of research have supported normativity in nonhuman animals. For one, great apes and cetaceans, and likely other species, have been found to share with humans practices that have been identified as moral foundations (Shweder & Haidt, 1993; Haidt, Graham & Joseph, 2009). In addition, there is evidence that other animals have the psychological capacities we take to be involved in a kind of norm Andrews (in prep) calls an animal social norm. We will briefly look at both these lines of argument.

4.1 Moral foundations

One line of investigation into the nature of moral psychology takes the form of looking for shared practices that are though to form the foundation of all moral thought. A well-known example of this project is seen in Jonathan Haidt and colleagues’ Moral Foundations Theory, which identifies between five and six dimensions that serve as the psychological foundations of human morality across cultures: harm/care, fairness/reciprocity, in-group/loyalty, authority/respect, purity/sanctity, liberty/oppression (Haidt, Graham & Joseph, 2009; Haidt & Graham, 2011; Iyer et al., 2012).

Haidt and colleagues appear to be open to the possibility that other animals share at least some of these foundations with humans; they write that there is “some evidence of continuity with the social psychology of other primates” (Haidt et al., 2009, p. 111). In a review of the existing literature, Vincent et al. (2019) provide evidence that great apes and cetaceans (whale and dolphin) demonstrate many of the same moral foundations. Vincent and colleagues integrate
Moral Foundations Theory with Krebs and Janicki’s (2002) five categories of moral norms (obedience norms, reciprocity norms, care-based and altruistic norms, social responsibility norms, and norms of solidarity) to identify behaviors that would fall under each category of moral norms. *Obedience norms* are identified that include social hierarchies, punishment, and teaching. *Reciprocity norms* are identified that include fairness and cheating, cooperation, mutualism, proportionality, and preference for individuals. *Care norms* include consolation, targeted helping/hurting, grief, and emotion recognition. *Social responsibility norms* include loyalty/betrayal, aversion or protesting, distribution of labor based on skill, and indirect reciprocity or cooperation for the benefit of the group. Finally, *solidarity norms* include sanctity/degradation, liberty/oppression, group identity or culture, and self-sacrifice.

The authors find plausible cases of each of these norm types in chimpanzees, and most of them across various cetacean species. Given the current state of the science, no one cetacean species was seen as having all or many of these capacities, but the distribution of behaviors in the family suggests that they may well exist across cetacean species. While chimpanzees have been studied since the 1960s, wild research on cetaceans is much more recent, and additional research on these species will be required.

The observation that the human moral foundations may be shared with other species suggests that there may be a deep structure to moral psychology that is widely conserved across species. Insofar as moral practice and cognition evolved to help us solve our social living problems, it should not be too surprising that the core practice types underlying a variety of solutions to social living are similar in this way.

### 4.2 Social norms
Social norms are defined differently by various theorists, but one thing these definitions have in common is an appeal to some motivation or authority that drives a behavioral regularity. Norms tell one what is a duty, what is permissible, and what is obligatory. Sripada and Stich define a norm as “a rule or principle that specifies actions which are required, permissible or forbidden independently of any legal or social institution” (Sripada & Stich, 2007, p. 281). Heath claims norms are social rules that “classify actions as permissible or impermissible” (Heath, 2008, p. 66). Bicchieri defines a social norm as a rule of behavior that individuals choose to follow because they believe two things: (a) that others in their community follow the rule and (b) that others also believe that community members ought to follow the rule (2006, 2017).

Even a cursory look at the range of discussion on norms can help us identify two widely shared elements of the accounts: norms are rules that are represented by the norm holders, and norm violators face sanctions on behalf of community members. Both these elements are sometimes presented as if they require a sophisticated cognitive capacity and practice. For example, Bicchieri’s account makes having social norms dependent on the capacity of community members to formulate a belief about the beliefs of others in their community; this makes mindreading a cognitive capacity necessary for having a social norm. Likewise, the requirements for punishment include sophisticated conceptual capacities when sanctioning norm violators is construed as requiring third party punishment—when a bystander responds to a violation with the goal of retribution or rehabilitation, and the violator understands the response to be punishment for the violation.

The project of looking for norms in nonhuman animals shouldn’t start by looking for the intellectualist pinnacle of the practice in human cultures, just as the project of looking for communication in animals shouldn’t start by looking for evidence that animals produce poetry.
And, just as all human communicative behavior is not poetry, not all human normative behavior elicits our metacognitive capacities. For example, consider human practices that are norm-like even when they fail to elicit beliefs about others’ beliefs: standing distance, greeting norms (hugging vs. kissing vs. no touch), or hygiene norms (how to blow a nose or use the toilet). There is nothing intrinsically functional about many such norms—they are not needed for the biological flourishing of the species, unlike norms about how to process food to make it safe for eating. Often, individuals don’t even know that their behavior is part of a cultural norm until they travel to another culture, meet a foreign guest, or otherwise see a violation. Nonetheless, humans are naturally motivated to follow these norms, and we are upset by violations.

To explain these sorts of human normative practices and to offer a schema that will be of more use when examining social norms in animals, Andrews has developed a less cognitively demanding account of a norm type that she calls *animal social norms*, modeled on Bicchieri’s account (Andrews, in prep). An animal social norm has the following three properties: (a) there is a pattern of behavior demonstrated by community members; (b) individuals are motivated to conform to the pattern of behavior; (c) individuals expect that community members will also conform, and that they will sanction those who do not conform.

Animal social norms require that individuals countenance rules of behavior, even if they do not recognize the rules or could not state them, and it has them *intrinsically motivated* to countenance the rule, given that the rule is practiced by in-group members. That is, the desire to follow the group’s norms is a part of the individual’s basic psychology, and it does not have to be learned. Despite widely cited claims to the contrary (e.g. Tomasello, 1999, 2016; Moore, 2013; Heyes, 2018), great apes have demonstrated the kind of selective social learning that we see in human norm learning, including prestige bias—imitating high ranking individuals—in
wild (Kendal et al., 2015) and captive communities (Horner et al., 2006); copying cultural natives after moving to a new community, either by a natural process of immigration (Luncz & Boesch, 2014; Luncz et al., 2012, 2015), or by a human-enforced process (Russon & Galdikas, 1993); and even overimitating in-group members (Myowa-Yamakoshi & Matsuzawa, 2000) (see Andrews, 2017, in prep for discussions). The cognitive capacities of imitation, and an early arising motivation to imitate in-group members, helps to explain why individuals to follow the norms they do; it’s how we do things around here.

Animal social norms also require that individuals sanction rule violations, which includes any attitude or act of disapproval. While third-party punishment is one sort of sanction, retaliation, withholding of cooperation, emotional discomfort around the violator, assuming a cost to watch a norm violator get punished (e.g. pushing open a heavy door to get a better view, spending time to watch a tyrant get executed on television), and shunning are examples of other sorts. For humans and nonhuman animals alike, third-party punishment is often not available given power dynamics in our societies. Just as less powerful humans protest the actions of the more powerful by avoiding violators, sharing information about violations to close friends, or often, by experiencing privately held emotions, nonhuman sanctions shouldn’t be expected to take the form of third-party punishment, or even to be easily apparent without careful attention. Punishment can be meted out by the powerful, but we certainly don’t want to limit our analysis of what counts as a sanction such that sanctions can only be exhibited by those in power.

Andrews argues that there are four cognitive capacities required for individuals to have animal social norms: identification of agents, sensitivity to in-group/out-group differences, social learning of group traditions, and the awareness of appropriateness. Evidence for this set of
capacities, which Andrews refers to as capacity for *naïve normativity*, is found in human children as well as nonhuman animals such as chimpanzees (Andrews, in prep).

### 4.2.1 Candidate animal social norms

The animal social norm concept allows for the systematic investigation of *potential* animal social norms across, and within, species. We briefly list a number of candidate animal social norms for primates that are worthy of further investigation using this framework. The following list includes practices that do fulfill animal social norms criteria (a), (b), and that may fulfill criteria (c):

- **Infanticide avoidance**—chimpanzee females protest infanticide (Rudolf von Rohr et al., 2011, 2015; see Nishie & Nakamura, 2018 for a description of the first observed case of a wild chimpanzee killing and eating an infant chimpanzee).

- **Treatment of infants**—chimpanzee infants enjoy permissive parenting for the first year of life, and are not punished by community members for any behavior. “They can do nothing wrong, such as using the back of a dominant male as a trampoline, stealing food out of the hands of others, or hitting an older juvenile as hard as they can” (de Waal, 2014, p. 189).

- **Helping**—chimpanzees help conspecifics even when there is no direct benefit to self (Yamamoto et al., 2009). Male and dominant chimpanzees aid females and youth in road crossing (Hockings et al., 2006). Chimpanzees destroy hunting snares that can injure group members (Ohashi & Matsuzawa, 2011). Gorillas have also been observed dismantling snares,
and in one observation juveniles worked together to destroy two snares just days after a snare had captured an infant member of their group. This report appeared in National Geographic:

[T]racker John Ndayambaje spotted a trap very close to the Kuryama gorilla clan. He moved in to deactivate the snare, but a silverback named Vubu grunted, cautioning Ndayambaje to stay away, Vecellio said. Suddenly two juveniles—Rwema, a male; and Dukore, a female; both about four years old—ran toward the trap. As Ndayambaje and a few tourists watched, Rwema jumped on the bent tree branch and broke it, while Dukore freed the noose. The pair then spied another snare nearby—one the tracker himself had missed—and raced for it. Joined by a third gorilla, a teenager named Tetero, Rwema and Dukore destroyed that trap as well. (Than, 2012)

- Food—chimpanzees share food with friends but not with non-friends (Engelmann & Herrmann, 2016). Chimpanzees as well as other species have calls indicating the presence of food. Withholding such calls so as to monopolize the food resource has been observed in rhesus monkeys (Hauser, 1992), capuchin monkeys (Di Bitetti, 2005), and in chimpanzees (Hauser & Wrangham, 1987). Violators of food call practices may be sanctioned by group members.

- Copulation rules—primates have strict rules about who copulates with whom. Juvenile chimpanzee males who venture too close to an estrus female risk being attacked by adult males (de Waal, 2014); macaques will have sex more often when bystanders are not around, especially the alpha males (Overduin de Vries et al., 2013); geladas engaging in extra-pair copulations are less likely to vocalize and more likely to copulate when the other pair-member is some distance away (le Roux, 2013).
• Immigrant conformity—immigrant chimpanzees have been observed to modify their tool use to conform to the practices of their new community, even though the adopted practice is less functional than their original practice (Luncz et al., 2012; Luncz & Boesch, 2014); vervet monkeys modify their food choices to conform to their new community, leaving untouched the food source they grew up with and that is not subject to competition (van de Waal, 2013).

• Arbitrary conventions—a female chimpanzee started wearing a straw-like blade of grass in her ear, and other chimpanzees began to do the same (van Leeuwen et al., 2014); a male capuchin monkey introduced hand-sniffing (mutual inserting of fingers in one another’s nostrils or eye sockets) and tail-biting games, which spread through the community (Perry et al., 2003); chimpanzees prefer to open a puzzle box in the way demonstrated by higher-ranking group members (Horner et al., 2006).

• Inequity avoidance—preference for fairness or resistance to inequalities, such as gaining the same reward for the same work. Chimpanzees and monkeys refuse to participate in tasks upon witnessing another receive a higher-valued reward (Brosnan et al., 2005, 2010; de Waal & Brosnan, 2003 – see also Table 1). Chimpanzees in an ultimatum game make more equitable divisions after partner protests (Proctor et al., 2013).

• Cooperation—working together to achieve a joint goal, such as cooperative hunting in chimpanzees (Boesch, 1994).
• Consolation—chimpanzees engage in higher levels of affiliation with a social partner after a conflict. They have been observed to console those who lose fights, reconcile after fights and facilitate reconciliation between fighting parties (de Waal & van Roosmalen, 1979; Kutsukake & Castles, 2004; de Waal, 2009).

• In-group preference—chimpanzees patrol boundaries between neighboring communities, sometimes invading and killing adult males and infants and kidnapping adult females (Watts & Mitani, 2001; Watts et al., 2006).

Determining whether these behaviors qualify as animal social norms will require evidence that the animals are motivated to conform to the behavior because their in-group members perform them, and that individuals who do not conform face sanctions of some variety. In many of the above cases, the behaviors are observed in wild, free-ranging populations, and the evidence can only come from trained field observers who work at the sites in which those behaviors have been observed.

Candidate animal social norms have also been identified in captive populations. Captive animals and their human caregivers may form a community in which animal social norms might be created. For example, group-housed captive animals may come to expect from their human caregivers that they provide the animals an equitable distribution of food. A violation of that expectation occurs when a human caregiver gives one individual a lesser-valued food reward than another individual, when both perform the same task. In response to the violation of an expectation, monkeys and chimpanzees behave in ways that can be taken as sanctioning (as do other species — see Table 1). Monkeys who are the victims of inequity express negative
emotions, and stop engaging with the human who perpetrated the inequity (Brosnan & de Waal, 2003). Chimpanzees who are the victims of inequity express negative emotions, and stop working with the human as well. In addition, some chimpanzees who observe their companions being victimized also express negative emotions and stop working with the human (Brosnan et al., 2005). These experiments suggest that captive primates may have an animal social norm that can be stated as “Humans should provide an equitable distribution of food.” To determine whether this is indeed an animal social norm, scientists could search for corroborative evidence of this norm by violating it in different contexts, and by looking for long term impacts on the social relationships between the human violator and the primate community members.

Experiments and observations in the wild and in the lab also offer support that great apes are sensitive to norm violations. For one, bonobos are able to discriminate between an expected aggressive encounter (i.e. one that is compatible with a response to a norm violation) and an unexpected aggressive encounter, offering different types of vocalizations in each context (Clay et al., 2016). This behavior is interpreted as evidence that bonobos recognize violations of social expectations, and that the vocalization serves to elicit social support in the face of a violation.

Another kind of preliminary evidence that nonhuman animals sanction violations comes from experiments that offer subjects the opportunity to “pay” to watch violators get punished. In one study that compared human children and chimpanzees’ responses to antisocial behavior, researchers found that individuals of both groups made an effort to watch the antisocial individual get punished for the antisocial behavior. Chimpanzees and children observed a scene in which a human acted prosocially (giving food to another person) or antisocially (teasing another and not giving them food). Chimpanzees later saw the antisocial individual get approached by a punisher who expressed rage and hit him. The antisocial individual and the
punisher then moved to another part of the room that was only visible to the chimpanzee subject if they opened a heavy door. Watching the antisocial actor get beaten up was worth the effort for the chimpanzees, who moved the heavy door more frequently when the antisocial individual was punished then when a prosocial individual was attacked (Mendes et al., 2018).

Another candidate animal social norm in a human/nonhuman captive context comes from biologist Diana Reiss, whose research focuses on dolphin cognition. Reiss described teaching a newly captive dolphin named Circe how to eat dead fish and to perform basic husbandry behaviors. Reiss quickly learned that Circe didn’t like the spiny tails of the dead fish she was being fed, so she started trimming the fish tails as she prepared food before a training session. One thing Reiss taught Circe was the meaning of a “time-out”—a negative reinforcer given to a dolphin when they don’t perform as expected. In a time-out the trainer steps away from the station, standing in an upright position without engaging with the subject for 30 seconds to a minute. Circe quickly learned what Reiss was teaching her, too. As the two individuals were learning about one another, they were also creating a community and forming expectations about one another. Reiss describes what happened during this period of getting to know one another:

One day during a feeding I accidentally gave her an untrimmed tail. She immediately looked up at me, waved her head from side to side with wide-open eyes, and spat out the fish. Then she quickly left station, swam to the other side of the pool, and positioned herself vertically in the water. She stayed there against the opposite wall and just looked at me from across the pool. This vertical position was an unusual posture for her to maintain … I could hardly believe it. I felt that Circe was giving me a time-out! (Reiss, 2011, p. 75)

In order to test her interpretation of Circe’s behavior, Reiss ran an experiment, purposefully inserting untrimmed fish tails into regular feedings, and she found that Circe always gave a time out when fed the untrimmed tails (Reiss, 1983). One might object that in both the
case of the inequity aversion, and in the case of the untrimmed tails, individuals were responding primarily to harm caused to self, so that it isn’t a case of a sanction because of a violated norm, but it is an expression of aversion toward an unwanted event. We agree that these experiments are not sufficient evidence for a norm in these communities, and that more evidence would be needed. Nonetheless, we find the current state of the science to support the hypothesis that some other animals live in communities guided by social norms.

5. Conclusion

In the introduction, we pointed out that questions related to animal morality can be divided into two broad concerns: the nature of the moral and the scope of the psychological capacities. We have attempted to shine some light on this second question and show how the evidence supports the idea that capacities of care, autonomy, and normativity extend beyond the boundaries of the human species. However, we acknowledge that the research on animal moral psychologies is still in its early stages and much more work needs to be done to warrant confident claims in this domain. A fundamental first step that must be taken is for scientists to take the question of animal moral psychology seriously, and realize that the error of mistakenly denying that moral capacities in animals exist is just as bad as the error of mistakenly attributing moral capacities to them (Andrews & Huss, 2014). There is no reason to fear one mistake over the other, since both would constitute a failure to describe the world with scientific accuracy. If, as we believe we’ve shown, there is enough prima facie evidence that points to the existence of moral capacities beyond Homo sapiens, we have good reason to incorporate this question into our research agenda.
The debate on the moral psychologies of animals can also benefit from research into the kind of relation that holds between the three capacity clusters that we have identified. Are these capacities related in the way they develop from an ontogenetic or phylogenetic perspective, or are they completely independent from each other? At first glance, it seems that capacities of autonomy are necessary for capacities of care and norms to count as such. If we return to the example of the ants freeing their entrapped conspecifics (Nowbahari et al., 2009), one of the reasons why this behavior is not generally viewed as an instance of care is because it seems to be automatically triggered by certain chemicals, with no cognitive or affective correlates. True instances of care may require, not just the adequate motivations to be in place, but also perhaps a higher level of intentionality than is seen in the ants, and thus some degree of autonomy. The same potentially applies to normative behavior. In order to distinguish it from the sort of behavioral regularities that we see in some insect species, such as ants, we might need there to be a certain level of autonomy: the animals choosing to conform to a behavioral regularity rather than being simply hardwired to do so. In addition, there might be some connection between care capacities and normativity, insofar as the sense of belonging to a group that enables normativity to emerge may be fostered by the presence of care capacities in the members of the community.

We don’t want to finish this chapter without mentioning the big elephant in the room: what warrants us labeling these capacities and behaviors as ‘moral’? We have opted for sidestepping this question because it is notoriously difficult to answer, and have instead chosen to focus on capacities that, under many popular views, are thought to be important or even crucial to morality. The question of the nature of the moral itself is so hard to answer that some authors, such as Fitzpatrick (2017) and Nado et al. (2009), have defended that it should be relinquished altogether. While it goes far beyond the scope of this paper, we nevertheless think that this
question is worthwhile asking, and would like to also call for more research in this area, particularly research that attempts to de-intellectualize morality and disentangle it from anthropocentric biases. Although definitive answers will be hard (if not impossible) to arrive at, we believe that research into the nature of the moral can illuminate much about our own species and the human-animal divide. So long as we are willing to confidently consider ourselves as moral, we think it is warranted to ask: can animals be moral too?

Acknowledgements

Research for this chapter was partially funded by the Austrian Science Fund (Project Number P 31466-G32) (Monsó) and by the Social Sciences and Humanities Research Council of Canada (435-2016-1051) (Andrews). Thanks to Dan Kelly, John Doris, and Brian Huss for helpful comments.

References


Rosenfeld, M. (1983). Two female northwest Atlantic harbor seals (P. vitulina concolor) carry dead pups with them for over two weeks - some unusual behavior in the field and its implication for a further understanding of maternal investment. Presented at the 5th Biennial Conference on Biology of Marine Mammals, Boston.


<table>
<thead>
<tr>
<th>Animal Group</th>
<th>Consolation behavior</th>
<th>Mourning behavior</th>
<th>Helping behavior</th>
<th>Inequity aversion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Great apes</strong></td>
<td>De Waal &amp; van Roosmalen, 1979; Kutsukake &amp; Castles, 2004; Palagi et al., 2004; Cordoni et al., 2006; Fraser et al., 2008; Clay &amp; De Waal, 2013; Palagi &amp; Norscia, 2013</td>
<td>Hosaka et al., 2000; Warren &amp; Williamson, 2004; Biro et al., 2010; Anderson et al., 2010; van Leeuwen et al., 2016</td>
<td>Warneken &amp; Tomasello, 2006; Warneken et al., 2007; Yamamoto et al. 2009; Hornet et al., 2011; Matsumoto et al., 2016</td>
<td>Brosnan et al., 2005; Brosnan et al., 2010</td>
</tr>
<tr>
<td><strong>Elephants</strong></td>
<td>Plotnik &amp; De Waal, 2014</td>
<td>Payne, 2003; Douglas-Hamilton et al., 2006</td>
<td>Bates et al., 2008</td>
<td>--</td>
</tr>
<tr>
<td><strong>Monkeys</strong></td>
<td>Call et al., 2002; Schino &amp; Marini, 2012; McFarland &amp; Majolo 2012; Palagi et al., 2014</td>
<td>Sugiyama et al., 2009; Fashing et al., 2011; Campbell et al., 2016; Yang et al., 2016</td>
<td>Wechkin et al., 1964; Masserman et al., 1964; Turner et al. 2005; Burkart et al., 2007; Lakshminarayanan &amp; Santos, 2008; Cronin et al., 2010</td>
<td>Brosnan &amp; de Waal, 2003; van Wolkenten et al. 2007; Massen et al., 2012</td>
</tr>
<tr>
<td><strong>Rodents</strong></td>
<td>Knapska et al., 2010; Burkett et al., 2016; Walker et al., 2003; Atsak et al., 2011</td>
<td>--</td>
<td>Church, 1959; Rice and Gainer, 1962; Greene, 1969; Evans and Braud, 1969; Bartal et al., 2011; Sato et al., 2015; Hernandez-Lallement et al., 2015; Márquez et al., 2015</td>
<td>Oberliessen et al., 2016</td>
</tr>
<tr>
<td><strong>Canids</strong></td>
<td>Cools et al., 2008; Palagi &amp; Cordoni, 2009; Baan et al., 2014; Cavalli et al., 2016; Quervel-Chaumette et al., 2016</td>
<td>Appleby et al., 2013</td>
<td>Bräuer et al., 2013; Piotti &amp; Kaminski, 2016</td>
<td>Range et al., 2009; Brucks et al., 2016</td>
</tr>
<tr>
<td><strong>Cetaceans</strong></td>
<td>Tamaki et al., 2006; Yamamoto et al., 2015</td>
<td>Fertl &amp; Schiro, 1994; Reggente et al., 2016</td>
<td>Park et al., 2012</td>
<td>--</td>
</tr>
</tbody>
</table>

Table 1. Some of the animal evidence of (proto-)moral behavior