Morality Without Mindreading

SUSANA MONSÓ

Abstract: Can animals behave morally if they can't mindread? Does morality require mindreading capacities? Moral psychologists believe that mindreading is contingently involved in moral judgements. Moral philosophers argue that moral behaviour necessarily requires the possession of mindreading capacities. In this paper, I argue that while the former may be right, the latter are mistaken. Using the example of empathy, I show that animals with no mindreading capacities could behave on the basis of emotions that possess an identifiable moral content. Therefore, at least one type of moral motivation does not require mindreading. This means that, *a priori*, non-mindreading animals can be moral.

1. Introduction

There appears to be something intuitively compelling about the idea that being a moral creature requires the ability to mindread, understood as the capacity to conceptualise and attribute mental states to others. The claim that one cannot be moral unless one possesses the ability to grasp, firstly, that others have interests and, secondly, the impact of one's actions on other beings' mental lives, seems to be obviously correct. Many moral philosophers and moral psychologists have been seduced by the intuitive plausibility of this idea, and assume its veracity. In this paper, I intend to go against this common trend and argue that, despite appearances, morality does not require mindreading, so that a nonhuman creature may be moral, in the sense that she may be capable of behaving morally, even if she lacks mindreading capacities.

There are two main reasons why I want to make this argument. The first one is the intention of adding some conceptual clarity to debates on animal morality. What exactly do we mean when we ask whether animals are moral? It is common to link morality to prosociality. However, not all behaviour that is prosocial is necessarily moral. For example, Nowbahari *et al.* (2009) found that ants will reliably free non-anesthetised conspecifics who are trapped in a snare. The ants' behaviour is surely prosocial, since it benefits others in the colony, but we cannot assert that it is moral until we determine what motivates the ants to behave this way, for it is in the

Acknowledgements: This research has been funded by the Spanish Ministry of Economy and Competitiveness, under grants BES-2012-052504 and FFI2014-57258-P. The author would also like to thank Judith Benz-Schwarzburg, Samuel Camenzind, José A. Gascón, Javier González de Prado, Marco A. Joven, Mark Rowlands and two anonymous referees for their helpful comments on previous drafts of this paper.

Address for correspondence: Unit of Ethics and Human-Animal Studies, Messerli Research Institute, University of Veterinary Medicine Vienna, Veterinärplatz 1, 1210 Vienna, Austria. **Email:** susanamonso@gmail.com

Mind & Language, Vol. 32, No. 3 June 2017, pp. 338–357.

underlying motivations that morality can be found. Vasconcelos *et al.* (2012) have distinguished between behaviour that is *functionally* intended to benefit another—i.e. behaviour that has been favoured by natural selection because it contributes to this end—and behaviour that is *psychologically* intended to benefit another. They assert that the latter only applies when the behaviour in question is 'goal-directed towards acting on the actor's *internal representation of the receiver's wellbeing*' (Vasconcelos *et al.*, 2012, p. 911, my emphasis), which would appear to link morality to mindreading capacities. While this may be sufficient for moral behaviour, I shall argue that we can conceive of more minimal forms of morality that will not require mindreading capacities while, at the same time, ensuring that not all prosocial behaviour qualifies as moral.

The second motivation behind this paper stems from the current situation in experimental studies on animal cognition. On the one hand, there is a growing amount of evidence that suggests the presence of something akin to moral behaviour in quite a large number of species, including monkeys (Masserman *et al.*, 1964; Lakshminarayanan and Santos, 2008), great apes (O'Connell, 1995; Clay and de Waal, 2013; Palagi and Norscia, 2013), elephants (Douglas-Hamilton *et al.*, 2006; Bates *et al.*, 2008; Plotnik and de Waal, 2014), cetaceans (Porter, 1977; Park *et al.*, 2012), canids (Cools *et al.*, 2008; Custance and Mayer, 2012; Palagi and Cordoni, 2009), rodents (Church, 1959; Rutte and Taborsky, 2007; Bartal *et al.*, 2011; Sato *et al.*, 2015) and corvids (Seed *et al.*, 2007; Fraser and Bugnyar, 2010). On the other hand, however, debates on mindreading capacities in animals are dominated by a rather pessimistic attitude on behalf of researchers, as there is only a limited amount of evidence, which is largely restricted to great apes and corvids, and is viewed as deeply problematic by several prominent researchers (for a review on the current state of this debate, see Lurz, 2011). Arguing for the logical independence of morality and mindreading presents itself as a crucial endeavour, if we want to ensure that the discouraging results in the animal mindreading debate don't trump the possibility of concluding that (some) animals are, nevertheless, moral.

This paper is divided into three parts. In the first part, I discuss the relationship between mindreading and moral judgement, as it is viewed from the eyes of moral psychologists. While I shall concede that it seems entirely plausible that moral judgement requires mindreading capacities, I will also show that this does not affect the case I want to make in this paper, as my concern lies specifically in moral behaviour. In the second part, I will argue against Cova (2013) and Nichols (2001), two philosophers who have, for different reasons, defended the claim that moral behaviour requires mindreading. Making use of the conceptual framework developed by Rowlands (2011, 2012), I will show that neither of them has succeeded in establishing a necessary connection between morality and mindreading. To make my case, I shall refer to the hypothetical example of an emotion that I shall term *minimal moral empathy.* In the final part of this paper, I shall present a more systematic definition of this emotion and some further considerations to defend the idea that, while this emotion would not require mindreading, it would nevertheless be moral in character.

2. Mindreading and Moral Judgement

Moral judgement, understood as the process whereby we subject situations, behaviours, motivations, etc. to scrutiny in order to determine their moral status, is commonly thought to engage our mindreading capacities. This is a claim that is backed up by a good deal of empirical research. Several studies from developmental psychology suggest that a higher development of children's mindreading capacities is positively correlated with more sophisticated performance in moral reasoning and moral judgement tasks (Baird and Astington, 2004; Cushman *et al.*, 2013; Dunn *et al.*, 2000; Lane *et al.*, 2010; Piazzarollo Loureiro and de Hollanda Souza, 2013). Furthermore, moral psychologists have found evidence in adult humans of a strong reliance of moral judgements upon the attribution of mental states, especially intentions (Cushman, 2008; Darley *et al.*, 1978; Woolfolk *et al.*, 2006). And lastly, studies in social neuroscience have also found that brain areas that are associated with mindreading are often activated during moral judgement and moral reasoning tasks (Bzdok *et al.*, 2012; Harenski *et al.*, 2009; Young *et al.*, 2007; Young and Saxe, 2008).

With this sort of research in mind, several authors have postulated that mindreading capacities play a crucial role in (human) moral judgement. Guglielmo *et al.* (2009), for instance, have defended the argument that in order to judge a certain behaviour as being intentional, and thus blameworthy, people require, amongst other things, 'evidence for the agent's *desire* for an outcome, *beliefs* about the action in question leading to the outcome, the *intention* to perform the action, *awareness* of the act while performing it' (Guglielmo *et al.*, 2009, p. 451, their emphasis), the evaluation of which engages our mindreading capacities. Gray *et al.* (2012) have asserted that the essence of morality is 'the combination of harmful intent and painful experience' (Gray *et al.*, 2012, p. 106), and so, at the heart of most moral judgements, lies the perception of a dyad of mental states: harmful intent on behalf of the agent, and painful experience on behalf of the patient. Young and Waytz (2013) assert that the main role of our mindreading capacities can be found in moral cognition, as they constitute a fundamental mechanism when it comes to understanding the actions of moral agents, predicting those actions of others that will affect us, as well as determining who is our ally and who our enemy (Young and Waytz, 2013, p. 93).

What these different authors are claiming is that *people's moral judgements involve mindreading*. This is something I'm not interested in disputing. In fact, I think it's a plausible and empirically supported claim. My interest is in arguing against a slightly different idea, namely, that *mindreading capacities are a necessary requirement for moral behaviour*. When these authors claim that 'at the heart of morality lies folk psychology' (Guglielmo *et al.*, 2009, p. 449), that 'mind perception is the essence of morality' (Gray *et al.*, 2012, p. 101), or that 'mind attribution is for morality' (Young and Waytz, 2013, p. 93), they are not defending the thesis I want to challenge. There are two important differences between the claim made by these authors and the one I want to debunk. The first difference is that, despite their use of the term 'morality', these authors' common claim is one that refers to moral judgement, and not

to moral behaviour. What they are asserting is that, at the heart of our capacity to engage in moral judgement, lies our capacity to attribute mental states to others. They generally do not address the question of whether behaving morally (that is, on the basis of motivations that are moral in character) requires the presence or involvement of mindreading capacities—which is precisely where my interest lies.

At this point, one could well ask whether moral behaviour can ever be separated from moral judgement, and indeed, the intuition that moral behaviour involves moral judgement is one of the reasons why philosophers defend the idea that morality requires mindreading. I shall argue against this intuition more carefully in the following sections, but for now, it suffices to introduce the conceptual framework offered by Rowlands (2011, 2012), and distinguish between moral subjects and moral agents. Rowlands has argued that the key to advancing the animal morality debate is to acknowledge that we can separate moral motivation from moral responsibility. Accordingly, the category of moral *subject* refers to individuals who sometimes behave on the basis of *moral motivations*, and the label of moral *agent* applies only to those moral subjects who can be held *morally responsible* for their behaviour. All moral agents are moral subjects, but not necessarily vice versa. The notion of moral subjecthood allows us to make sense of the idea that moral behaviour may sometimes obtain in the absence of moral responsibility. This enables us, for instance, to describe the behaviour of a child who is a bully or a mentally ill person who commits a crime as *morally bad*, even though different considerations may prevent us from holding them responsible or blaming them for their behaviour. Among the considerations that would determine to what extent they are morally responsible, would lie the question of whether these two individuals can fully understand and properly engage in moral judgements. If they can't understand that what they have done is morally bad, it doesn't make much sense to hold them morally responsible for their behaviour.[1] On the other hand, it seems much less controversial to say that what they did was morally bad. In the following section, I will argue that while moral agency almost certainly requires moral judgement, this is not the case for moral subjecthood.

The second big difference between the claim Guglielmo *et al.* (2009), Gray *et al.* (2012) and Young and Waytz (2013) defend and the one I want to dispute is that these authors are stating a *contingent* fact about (human) morality, and not describing a *necessary* one. Their arguments are made after empirical examinations of people's moral judgements in different circumstances. What is being argued is not that mindreading *must* be involved in order for one to perform a moral judgement, but rather, that mindreading is *actually* involved whenever people judge the morality of others' behaviour. The idea that (normal adult) humans make use of their mindreading capacities whenever they engage in moral judgements is not one that I intend to question. Even if these authors were defending the claim that humans engage in the attribution of mental states to others whenever they *behave* morally (as seems

[1] Cova (2013) has positioned himself against this claim. I shall address his ideas in the following section.

to be suggested by Young and Waytz, 2013, pp. 97–100), I would not be interested in disputing that claim, either. This is because if it is true that all human moral behaviour involves mindreading *de facto*, it does not follow that moral behaviour *necessarily* involves mindreading. My interest lies precisely in arguing that mindreading does not have to be involved in order for a behaviour or a motivation to count as moral and, moreover, that an individual does not have to possess mindreading capacities to be a moral subject. These scholars do not take a stand on this issue.

I want to argue that morality does not require mindreading, and this should be understood as the claim that there may be creatures that are moral *subjects* even though they lack mindreading capacities. As stated in the introduction, I will understand that behaviour is moral when it has been triggered by motivations of the appropriate sort. The assertion that moral subjecthood does not require mindreading capacities amounts to the claim that certain individuals who lack these capacities may, nevertheless, behave on the basis of motivations that are moral in character. To state the obvious, I do not intend to argue that non-mindreaders may behave on the basis of *any* moral motivation whatsoever. The absence of mindreading capacities may preclude them from being the subjects of certain kinds of moral motivations, such as considerations of the following form: 'I want to φ but if I φ, Jones will suffer; making people suffer is morally wrong, so I should refrain from φing.' Accordingly, my claim is a rather restricted one—namely, that there is at least one kind of moral motivation that does not require the possession of mindreading capacities. This is what I shall term *minimal moral emotions*.

3. Mindreading and Moral Behaviour

In the previous section, we saw that empirical evidence from disciplines such as developmental psychology and social neuroscience has led some moral psychologists to assert that mindreading capacities are (contingently) involved whenever humans engage in moral judgements. As stated above, I do not intend to discuss this claim. Instead, I will focus, in this section, on the arguments put forward by moral philosophers to account for a necessary connection between mindreading and moral behaviour. In contrast to moral psychologists' analysis of moral judgement, where the capacity to attribute *beliefs* and *intentions* to others is regarded as occupying a central role, moral philosophers, as we shall see, consider the attribution of *affective states* to others as being crucial for moral behaviour.[2] My aim is to show that moral behaviour can obtain even in the absence of an ability to engage in phenomenal mindreading.

I will focus, in this section, on two philosophers who have defended the idea that moral behaviour *necessarily* requires mindreading capacities: Cova (2013) and

[2] My thanks to an anonymous referee for encouraging me to make this clarification.

Nichols (2001). They both seem untroubled by the idea that animals may be moral, and offer an account of the minimal cognitive requirements for morality that they interpret as allowing animals into the realm of moral creatures (Cova, 2013, pp. 128–9; Nichols, 2001, p. 450). While I sympathise with the de-intellectualised accounts of morality that they put forward, I worry that their inclusion of mindreading amongst these cognitive requirements may serve to undermine the possibility of animal morality. I believe, however, that neither of them succeeds in establishing a necessary connection between mindreading and moral behaviour. In this section, I will attempt to refute their arguments, and defend the idea that minimal forms of morality that do not require the capacities for moral judgement or mindreading are perfectly conceivable, and indeed, entirely plausible.

3.1 Cova (2013): Mindreading is for Caring

Cova (2013) distinguishes two positions in the animal morality debate: (1) *continuism*, or the idea that some of the capacities that make humans moral are present, if somewhat rudimentarily, in some nonhuman species, and (2) *discontinuism*, or the thesis that only humans possess these capacities. In a similar vein as Rowlands, Cova argues that in order to make progress in debates on animal morality, we need to introduce modifications into the way we use our moral terms. However, in contrast to Rowlands' (2011, 2012) proposal of separating moral motivation from moral responsibility, Cova (2013) argues that the disagreement amongst continuists and discontinuists can be at least partially solved by separating moral judgement from moral agency. He understands moral agency as the quality of being 'morally responsible of (some of) [one's] action' and moral judgement as the ability to 'judge whether something is right or wrong' (Cova, 2013, p. 118). The key to solving the debate, according to Cova, is realising that one can be a continuist about moral agency and a discontinuist about moral judgement, and thus, that we can establish that 'we share the psychological bases for moral agency with other animals' while acknowledging that 'we are the only known species able to form moral judgments' (Cova, 2013, p. 118).

In order to defend his position, Cova attempts to construct an account of moral agency without moral judgement. To this end, he gives a series of examples of people performing good or bad deeds without stopping to reflect on their goodness or badness, and argues that it is counterintuitive to consider that the agents were not morally responsible for these actions just because they didn't execute their moral judgement capacities while performing them. It seems, however, that he is not properly addressing the issue at hand, because in all the cases he considers, the agents are healthy adult humans who *do* have moral judgement capacities. As a result, they can understand (even if they haven't stopped to reflect upon it in these particular instances) that what they're doing is right or wrong, and that they should be praised or blamed for it. Cova doesn't succeed in showing that this isn't part of the reason

why their actions are worthy of praise or blame. For example, Cova considers the following case:

> Let's say that Jack had a bad day, was irritated, and smashed the window [sc. of the car parked in front of him] without taking the time to assess whether it was right or wrong. Let's also say that, though he realized afterwards that it was the wrong thing to do, he did not regret this action at great length. Should we say that Jack is not responsible and does not deserve blame for what he did? That he hasn't the duty to pay for repairs? That seems very counter-intuitive (Cova, 2013, p. 123).

An obvious response would be to say that the fact that Jack has moral judgement capacities, but has failed to make proper use of them, is an important part of what warrants our holding him morally responsible for his action. The fact that he has these capacities does, indeed, make it counterintuitive to say that he is not responsible. On the other hand, if Jack were a severely disabled individual who lacked all capacity to understand that what he did was wrong, then holding him responsible or making him pay for repairs would also be very counterintuitive. Cova does not put forward any argument to prevent us from interpreting his examples this way.

Cova's characterisation of a 'mere' moral agent is also a rather awkward one. He states that a moral agent that is incapable of engaging in moral judgements would be 'responsible for his action' but that this would not 'necessarily entail that he can be punished,' because 'one has to be both a moral agent and a moral judge to be an appropriate target of punishment' (Cova, 2013, p. 129). Cova defends this last claim by appealing to his own intuition ('it seems to me that we want the people we punish to understand why they are punished'), and also to an experimental study by Gollwitzler and Denzler (2009) that suggests that people 'consider revenge satisfactory only if the offender understands (and acknowledges) that revenge was taken against him because and in virtue of a prior unfair behaviour' (Cova, 2013, p. 129). However, unless we speak of degrees of moral responsibility (which Cova does not), it is very difficult to see how someone can fulfill the cognitive requirements to be held morally responsible for an offence and for this not to entail that she can be punished. It seems that any consideration that makes reference to the individual's personal characteristics (and not to circumstantial facts) and serves to undermine the legitimacy of punishment will also entail that the person in question is not a full-blown moral agent. The whole point of holding someone *morally* responsible is to open the possibility of praising or blaming her, and rewarding or punishing her if the circumstances warrant it. Cova's appeal to the intuition that those we punish need to understand why they are punished only works against his own interest, because it points, precisely, to the fact that moral responsibility and moral judgement cannot be separated.

We have seen that Cova's argumentative strategy to defend the logical independence of moral agency and moral judgement is unsatisfactory and, also, that his characterisation of a 'mere' moral agent is rather counterintuitive. However, the most

important point that needs to be made here is a more fundamental one—namely, that Cova *does not need* to separate moral agency from moral judgement. He doesn't need to do so, because he is merely trying to make sense of behaviours that can be *morally motivated*, that occur 'for good reasons' (Cova, 2013, p. 127), in the absence of moral judgement capabilities. Appealing to the category of moral subjecthood is a much easier way of achieving this. An individual who can behave on the basis of morally good reasons (or morally bad ones) is a moral subject. If that subject *also* possesses the ability to engage in moral judgements, then she is *also* a moral agent, and can be held responsible (and, thus, praised or blamed) for her behaviour. This allows us to make sense of the morality of the examples Cova uses even if we suppose that the subjects involved lack all capacity to engage in moral judgements. At the same time, we can also avoid the awkward consequences that come from separating moral agency from moral judgement.

I have a further concern with Cova's paper, which will bring us to the issue of mindreading. It seems to me that he doesn't fully succeed in separating moral behaviour from moral judgement. To see why this is so, let us consider one of his main aims in this paper, which is to give a minimal account of what it means to act for morally good reasons. Cova reaches the conclusion that 'one acts for the good reasons when one actually cares about the person one is trying to help' (Cova, 2013, p. 127). In order to be able to say that an individual *cares* for another, two conditions must be met, according to Cova: (1) she must be 'able to understand that this person has *interests*,' and (2) she must 'give an intrinsic (i.e. non-instrumental) positive value to the fact that this person's interests are preserved and augmented' (Cova, 2013, p. 127, his emphasis). These two conditions, in his view, mean that being a moral creature only requires '[acting] according to what we attach importance to and a bit of theory of mind [i.e. mindreading capacities]' (Cova, 2013, p. 128).

Cova, therefore, considers it necessary for mindreading capacities to be present in order to be able to say that an individual *cares* for others, and thus that she helps others for morally good reasons. My reason for thinking that Cova does not succeed in letting go of moral judgement capacities resides in the other condition he puts forward, namely, the idea that the moral individual must also be capable of *attaching importance* or *giving an intrinsic positive value* to the preservation of others' interests. Attaching importance or giving a positive value to X seems to mean, precisely, (consciously or unconsciously) judging that X is a good thing. How is this different from moral judgement? How can a person attach importance to something without the ability to engage in moral judgements? Cova does not offer a solution to this question, which means it poses a potential threat to his whole argument. I believe, however, that there is a way around this problem; a way that will allow us to effectively separate moral behaviour from moral judgement, and additionally, from mindreading capacities.

My proposal is the following. We can construct a minimal account of care that will not require the ability to (consciously or unconsciously) judge that preserving someone's interests is a good thing, by focusing on the emotions that an individual might undergo when deciding to help someone. So, let's imagine an individual

called Higgins who feels sad whenever he sees others in distress. Let's imagine that this sadness is what triggers Higgins' urge to help or comfort them, and that he feels happy once they're no longer in distress. Even if Higgins were incapable of engaging in moral judgements, Rowlands' (2012) framework allows us to categorise Higgins' behaviour as moral because it is motivated by an emotion that *tracks* the moral proposition: 'This creature's distress is bad.' An emotion can be said to track a moral proposition if there is a truth-preserving relation between the emotion and the proposition in question, so that the truth of the proposition is guaranteed whenever the emotion is not misguided (see section 4 below; Rowlands, 2012, Chapters 2, 9).

To see what this means, imagine that Higgins is sad because he sees that Jane is crying, and that his sadness is *intentionally* directed at Jane's crying—that is, Higgins is sad *that* Jane is crying. His being sad that Jane is crying means he is *experiencing* Jane's crying as something unpleasant, as something bad. Built into his sadness at Jane's crying is an urge to comfort her. This experiential form that Higgins' emotion takes is what allows us to speak of it tracking the proposition 'This creature's distress is bad' (see also section 4). Now, suppose that Jane's crying were due to an entirely mundane reason, such as the fact that she's watching a sad film. Or suppose that Jane had committed a serious offence and her crying was the result of her well-deserved punishment. In these cases, Jane's crying would (at least arguably) *not* be a bad thing. Higgins' sadness would then be *misguided*, because he would be experiencing as bad something that is not, in fact, a bad thing. This is what it means to say that the truth of the proposition 'This creature's distress is bad' is guaranteed by the non-misguided status of Higgins' emotion.

The idea of emotions that track moral propositions allows us to separate moral behaviour from moral judgement entirely. We do not need Higgins to be capable of entertaining a proposition such as 'This creature's distress is bad.' Higgins can lack all capacity to engage in moral judgements *and still be a moral subject*, because the morality of his behaviour comes by virtue of the fact that it is triggered by an emotion that tracks a moral proposition. And here comes the crucial bit—since we can let go of all capacity to entertain the proposition 'This creature's distress is bad,' we can also let go of all capacity to entertain the proposition 'This creature is distressed.' That is, we do not need Higgins to be capable of attributing any mental states to others, because what is important is that he reliably undergoes sadness when witnessing someone's distress. It is enough for his sadness to be triggered by, and intentionally directed at, the *superficial behavioural cues* that accompany Jane's distress—we do not need him to understand anything about the underlying mental states.

For Higgins to be a minimal moral subject, we need his behaviour to be motivated, at times, by a (minimal) moral emotion. This (minimal) moral emotion can be of different sorts. In the example I have been using, Higgins is motivated by an emotion we can call *minimal moral empathy*[3] (see section 4 for a definition of this

[3] I use this cacophonous phrase, instead of merely the term 'empathy,' to distinguish the concept the former refers to from the many uses of the term 'empathy' that can be found in the literature,

notion). For this specific kind of moral emotion to obtain, Higgins must possess a reliable mechanism that ensures that, upon detection of the superficial markers of distress in others, he undergoes emotional contagion, where this is understood as a process of affective resonance that results in an affective state that is isomorphic to the target's. Thus, witnessing the target's distress behaviour must trigger distress in Higgins himself.[4] Crucially, however, the emotional contagion that Higgins undergoes must be of a special sort, insofar as his resulting distress has to be intentionally directed at the state-of-affairs that is the other individual in distress. We need this intentional form of distress to be what moves Higgins to help or comfort the other. If these conditions are fulfilled, Higgins will be a moral subject, regardless of whether or not he is a moral judge, a mindreader, or indeed, a human being.

3.2 Nichols (2001): Mindreading is for Motivation

In this paper, Nichols gives a minimal account of the cognitive mechanisms that underlie what he considers one of the 'basic moral capacities: the capacity for altruistic motivation' (Nichols, 2001, p. 425). He focuses on the core cases of human altruistic motivation—namely, 'cases of helping or comforting others in distress' that emerge in early childhood and are pervasive among adults (Nichols, 2001, p. 428)—and argues that they can be best accounted for by postulating the existence of what he calls a 'Concern Mechanism' (Nichols, 2001, p. 426). This is an affective system that produces the motivation to engage in helping or comforting behaviour when we encounter someone in distress. One of Nichols' aims in this paper is to determine what kind of mindreading mechanisms must be in place for the 'Concern Mechanism' to be activated. He argues against two different options: (1) the possibility that all altruistic motivation may be triggered by perspective-taking processes, which he finds too intellectually demanding and empirically implausible (Nichols, 2001, pp. 440–3), and (2) the possibility that no mindreading capacities at all are required for altruistic motivation to occur. It is in his arguments against option (2) where we find a defence of the idea that mindreading capacities are necessary for moral behaviour. I shall now address his arguments, in an attempt to show that he does not succeed in proving that there is a necessary connection between mindreading and moral motivation.

Minimal moral empathy (MME) would be an example of what Nichols considers the most 'radical' view of the relationship between mindreading and altruistic motivation—namely, the view that no mindreading at all is required for

---

many of which refer to abilities that are either not moral (e.g. de Vignemont and Jacob, 2012; Gallagher, 2012; Goldman, 2006; Zahavi, 2010), or far from minimal (e.g. Dixon, 2008; Jamison, 2014). See Monsó (2015) for a comparison of *minimal moral empathy* and other forms of empathy.

[4] I'm using the term 'distress' in a broad sense to encompass all affective states that are forms of suffering, such as pain, fear, anxiety, sorrow, etc. Thus, while the emotion delivered by Higgins' process of emotional contagion need not be exactly the same as Jane's (e.g. he can be *anxious* as a result of her *sorrow*), they both have to be forms of distress, broadly construed. I am grateful to an anonymous referee for encouraging me to make this clarification.

altruistically-motivated behaviour to occur (Nichols, 2001, p. 426). The core cases of altruistic motivation, according to Nichols, cannot be adequately accounted for from this 'radical' view because without mindreading one will not be motivated to help another when escaping is easier. In particular, Nichols considers that we need at least a 'minimal mindreading capacity to attribute negative affective or hedonic states to others' in order for altruistic motivation to take place (Nichols, 2001, p. 346). Continuing with our previous example, Nichols would say that the lack of mindreading capacities means that Higgins, due to his emotional contagion, experiences Jane's distress behaviour as 'bad music' that he'd like to turn off, and escaping will be just as good a solution as comforting her (Nichols, 2001, p. 429). Possessing a capacity to represent negative affective states, on the other hand, means that 'escape is not an adequate alternative' because 'the motivation comes from an enduring *internal* cause' (Nichols, 2001, p. 435, his emphasis). If Higgins possessed mindreading capacities, his motivation to help Jane would not have been triggered by mere superficial cues, but by a representation of her distress, which would mean that 'merely escaping the perceptual cues of pain won't eliminate the consequences of the enduring representation that another is in pain' (Nichols, 2001, p. 436).

The most obvious response here would be to argue that MME can also take the form of an 'enduring internal cause.' We are not supposing that Higgins' emotional contagion results in self-directed personal distress, or in a form of distress with no intentional object, but rather, that as a result of his emotional contagion, Higgins is distressed *that* Jane is displaying distress behaviour. Jane's distress behaviour is the intentional object of Higgins' distress—he experiences it *emotionally* as distressing. If his emotional contagion gives rise to an emotion that is *intentionally directed* at Jane's distress behaviour, then the cause of Higgins' distress can persist in his mind even after he escapes the situation. Nichols considers this possibility but dismisses it:

> An emotional contagion theorist might continue to deny any role for mindreading and maintain that altruistic motivation comes from an enduring representation of the behavioral, acoustic, or physiognomic cues that cause emotional contagion. … The problem is that … if one knows that the cues leading to emotional contagion are merely superficial, this typically does not prevent one from experiencing emotional contagion, but it does undermine altruistic motivation (Nichols, 2001, p. 435).

This response is unsatisfactory because it presupposes that the being in question possesses mindreading capacities and can '[know] that the cues leading to emotional contagion are merely superficial'. If an individual lacked mindreading capacities altogether, then she wouldn't be capable of distinguishing 'merely superficial' cues from those that were markers of underlying mental states, and so there is no reason to suppose that the superficial cues couldn't provide the adequate motivation. All that is required is for the individual's past learning experiences or hardwired behavioural dispositions to move him to engage in affiliative behaviour as a result of his (intentional) emotional contagion.

In contrast to an account of altruistic motivation based on emotional contagion, Nichols proposes that the attribution of distress to others may be precisely what triggers the affective process that results in a motivation to behave altruistically. This is what Nichols terms the 'Concern Mechanism,' which he proposes as underlying much of our altruistic behaviour, working as follows:

> The distress attribution might produce a kind of second order empathic distress in the subject. For example, representing the sorrow of the target might lead one to feel sorrow. This would provide a kind of empathic motivation for helping. And the motivation would be effective even when escape is easy. For the cause of the emotion is still the representation of the other's mental state and as a result, one is motivated not simply to escape the situation since that would not rid one of the representation (Nichols, 2001, pp. 445–6).

We have seen that there is no reason to suppose that emotional contagion cannot trigger an enduring representation. At the same time, there is no reason to suppose that 'escaping the situation' cannot rid one of the representation of another's distress. We have all experienced being overwhelmed by a particularly tragic story in the news and changing channel to avoid being put off our dinner. There need not be images of people in distress to trigger an emotional contagion in order for this to occur. It could simply be the case that a journalist is reporting a story that makes us very aware of the suffering of the people involved in it. By changing channel, we become distracted and thus stop thinking about their distress—that is to say, our internal representation of it is 'turned off.' Of course, the cases that Nichols is thinking about are those in which we are faced with the dilemma of helping the person in distress or leaving the scenario—in this respect, my example is not a perfect analogy. Nevertheless, it serves to illustrate the fact that having an internal representation of someone's distress does not mean that it must endure once we escape the situation—we can, after all, make the effort to become distracted and think about something else. Having mindreading capacities does not ensure that one will help when escaping is easier. The 'Concern Mechanism' that Nichols proposes would, therefore, not necessarily be infallible. The same applies to MME, about which we can never say that it will *necessarily* result in comforting or helping behaviour—there could always be other intervening factors that motivate the individual towards a different behavioural outcome. For both Nichols' 'Concern Mechanism' and MME to result in helping or comforting behaviour, a number of further conditions must probably obtain: the individual must have the right beliefs, her relation to the distressed person must be of a certain sort, there must not be any other pressing issue to attend to, and so on.

It is important to emphasise that Nichols is directing his arguments at the thesis that *all* altruistic motivation is caused by emotional contagion, and this is not what I'm interested in defending. It certainly seems plausible to me that something akin to Nichols' 'Concern Mechanism' may be at the basis of much human (perhaps even nonhuman) altruistic behaviour. My point is that MME may *also* be a kind of altruistic motivation and that the arguments put forward by Nichols do not succeed in proving otherwise. Of course, whether MME ever does take place is an empirical

matter on which I take no stand. My interest is in arguing that, if an individual (at times) behaved on the basis of moral emotions such as MME, she would be a moral subject, regardless of whether or not she possessed mindreading capacities.

4. What Makes *Minimal Moral Empathy* Moral?

In the previous section, I have shown that the arguments put forward by Cova (2013) and Nichols (2001) to defend a necessary connection between mindreading and moral behaviour are not convincing. In the process, I have presented the notion of MME as an example of a hypothetical caring or helping motivation that would be entirely independent of mindreading and moral judgement. An obvious objection emerges at this point: sure, the reader may say, MME does not require the possession of mindreading capacities or the ability to engage in moral judgements, and it may indeed be enough to trigger helping or comforting behaviour, but what reason do we have to believe that this is a *moral* emotion? What warrants the attribution of moral subjecthood to those who behave on the basis of MME? This is a very important objection, since failing to address it correctly would mean that I had not succeeded in my defence of the independence of morality and mindreading. I have thus chosen to devote this last section to disentangling what makes MME moral. Let us begin by defining this notion more carefully:[5]

Creature C possesses minimal moral empathy (MME) if: (1) C has an ability to detect distress behaviour in others, and (2) due to the action of a reliable mechanism, the detection of distress behaviour in others results in a process of emotional contagion that (3) generates a form of distress that has the other's distress behaviour as its intentional object, and built into which is (4) an urge to engage in other-directed affiliative behaviour.

To determine whether or not MME-based behaviour would be moral, we have to first establish what makes a certain behaviour deserve this label. In the introduction, I already hinted at a sufficient (if not necessary) condition for determining that a behaviour is moral—namely, whether it has been triggered by a moral motivation. Now, this of course puts the *explanandum* in the *explanans*, but it is not an entirely vacuous definition. It means that we cannot determine whether a behaviour is moral solely by looking at its outer characteristics but, rather, that we must also look at its underlying motivations. The key question, then, is What makes a motivation moral? Since there is no universally accepted definition of morality, it would be very difficult, and surely beyond the scope of this paper, to give an explanation that would satisfy everyone. There are, however, a number of considerations that can be made about MME, which taken together build quite a strong case for its moral character. To see this, let's return to the Higgins example, now explicitly thinking

[5] Earlier characterisations of MME can be found in Monsó (2015) and Rowlands and Monsó (2017).

of Higgins as a nonhuman individual, such as a dog. Here are the reasons why he should be considered a moral subject:

1. Even though Higgins cannot conceptualise and attribute to others the mental state that underlies displays of distress behaviour, his emotional contagion occurs in the presence of others in distress. This is important insofar as distress is a *morally relevant* feature of situations. This is something about which most (if not all) theories in normative ethics will agree. From the point of view of normative ethics, whether a situation involves distress or whether an action produces distress in others or not is something that must generally be taken into account when deciding how to act or how to evaluate the actions of others. Higgins possesses an *emotional sensitivity* to this morally relevant feature, since he not only feels sad whenever he sees others in distress, but also emotionally experiences their distress behaviour as distressing. Higgins may also emotionally experience the noise of the vacuum cleaner as distressing, but this sensitivity is not a moral one because vacuum cleaning noises are not morally relevant features of situations, whereas distress behaviour is a reliable marker of the bad-making feature of situations that is distress as a mental state. And we are supposing that Higgins' sensitivity is not accidental or contingent but, rather, grounded in what Rowlands (2012, Chapters 5, 9) would call a 'moral module,' that is, a psychological mechanism, such as a perception-action mechanism (Preston and de Waal, 2002), which entails that Higgins' emotional sensitivity is *reliable*.

2. Moral emotions are usually thought to involve moral judgements, so for MME to be a moral emotion, some theorists (Dixon, 2008, pp. 129–40) would require Higgins to be capable of entertaining a proposition such as 'This creature's distress is bad.' It seems safe to assume that Higgins, being a dog, cannot entertain such a proposition, for he lacks the requisite concepts. However, we can consider that his emotion *tracks* this proposition. To understand what this means, it is important to point out that moral propositions are not merely a record of how the world is—the proposition 'This creature's distress is bad,' if true, means that this creature's distress is indeed a bad thing—as they also have a motivational component, such that if I believe in the truth of 'X's distress is bad', then, *ceteris paribus*, upon seeing that X is in distress, I will be moved to help or comfort X. We can establish that Higgins' distress at Jane's distress (behaviour) tracks the proposition 'Jane's distress is bad' because of the phenomenal character that Higgins' emotion takes. On the one hand, Higgins reliably feels distressed whenever he sees Jane displaying distress behaviour, and his distress is not merely triggered by, but also directed at, her distress behaviour. This means he experiences Jane's distress as something negative, unpleasant. The *descriptive* aspect of the moral proposition 'This creature's distress is bad' is thus present in Higgins' aversion to Jane's distress (behaviour). On the other hand, Higgins' emotion also has the adequate *motivational* component to it, insofar as, built into his distress, is an urge to engage in affiliative behaviour directed at the target. Even though Higgins cannot entertain the moral proposition that MME tracks, it is implicit in the phenomenal character of Higgins' emotion, in the sense that Higgins reliably experiences Jane's distress (behaviour) as something negative, and as something he wishes to eliminate.

3. The fact that MME tracks the proposition 'This creature's distress is bad' means that there is a truth-preserving relation between Higgins' emotion and this proposition, such that, if the emotion is not misguided, the proposition must be true (see Rowlands, 2012, Chapters 2, 9). This allows us to say that Higgins' emotion would be moral even if it were misguided. It doesn't necessarily have to be true that Jane's distress in these particular circumstances is a bad thing. She could be crying simply because she is reading a sad book. In this case, Higgins' MME would be misguided, but we could still say that it tracks this moral proposition because it has the adequate phenomenal character (see the previous paragraph) and because, were it not to be misguided, the truth of this proposition would be guaranteed.

4. Since MME takes the form of an emotion that has Jane's distress (behaviour) as its intentional object, it becomes a *reason* for Higgins' comforting behaviour, rather than a mere cause. Recall the ants in the experiment we saw in the introduction (Nowbahari *et al.*, 2009). Their helping behaviour was presumably triggered by a 'chemical call for help' from their entrapped conspecific (Nowbahari *et al.*, 2009, p. 3) and probably had no cognitive or affective component, which, if true, would mean that such a 'call for help' was the *cause* of their behaviour, but not the *reason* behind it. Regardless of the mechanisms underlying the ants' behaviour, in the case of Higgins we are supposing that his behaviour is triggered by an emotion with intentional content, which means it has been done *for a reason*. The fact that this emotion has intentional content turns MME into what Higgins *should* feel, given the circumstances. Higgins not only helps Jane; he does so, we are supposing, on the basis of a motivation that has an adequate phenomenal character. This allows us, as external evaluators, to say that he helps her *for the right reasons.*

5. MME resembles the phenomenology of what humans often experience when moved by another's distress to help or comfort them. Explicitly thinking 'This person's distress is bad' is not usually a step along the way—we may even be considered callous if we need to stop and make this sort of consideration before helping someone in distress. The fact that we can engage in these sorts of considerations *post hoc* ('It was good of me to help her; after all, her distress was a bad thing') is relevant when it comes to granting us moral agency, but our behaviour before engaging in these considerations was already moral in character—because it was triggered by an emotion of the appropriate sort. Since we are only considering granting moral subjecthood to Higgins, and not moral agency, the fact that his motivation to comfort Jane is phenomenologically similar to what ours would be like is surely relevant.

6. MME, being a fairly reliable disposition to react emotionally to others' distress, can be considered a character trait of Higgins, and, since it incorporates an urge to engage in affiliative behaviour, we can say that it systematically produces good consequences, for it will tend towards alleviating others' distress. This is enough, on some de-intellectualised accounts of virtues (Driver, 2006), to consider that Higgins is virtuous. The fact that MME systematically produces good consequences also allows us to evaluate Higgins' MME-based behaviour, from an objective consequentialist perspective, as morally good. While we cannot praise Higgins for his

behaviour, we can say that it is 'a good thing that the world contains a subject like this, an individual who acts in this way' (Rowlands, 2012, p. 254).

These six conditions warrant our attributing moral subjecthood to Higgins. Before concluding, let us address one final objection. Imagine that instead of sweet old Higgins, Jane had a cat called Lunchbox, who was completely indifferent to her distress behaviour. Suppose that Jane decided to try to make a moral subject out of Lunchbox by systematically stimulating the punishment centre in his brain while simultaneously presenting him with individuals displaying distress behaviour. Lunchbox would soon acquire an aversion to distress behaviour. Suppose that Jane then managed to teach Lunchbox that he could stop the punishment signal by engaging in affiliative behaviour towards the target. Lunchbox's aversion to distress behaviour would eventually be accompanied by an urge to engage in affiliative behaviour. While Lunchbox's emotion clearly would not count as MME, for it would not have been triggered by a process of emotional contagion, could it nevertheless be said to track the proposition 'This creature's distress is bad'? Would Jane have succeeded in turning Lunchbox into a moral subject?[6]

There are two important differences between Higgins and Lunchbox that justify attributing moral subjecthood to the former but not the latter. First, Higgins' emotional sensitivity to distress is grounded in the operations of a hardwired psychological mechanism, or 'moral module,' while Lunchbox's is not. Higgins is thus *naturally* inclined to experience distress when in the presence of distress behaviour in others, while Lunchbox has merely been subjected to a form of conditioning. Second, while both Higgins' and Lunchbox's distress is caused by the presence of distress behaviour in the immediate environment, only Higgins' distress can be said to be not merely triggered by, but also directed at, the target's distress behaviour. In the case of Lunchbox, we know that the distress behaviour is not what he is distressed *about* because Jane could easily substitute it for any other stimulus and generate that same aversion. The target's distress behaviour is the *cause* but not the *content* of Lunchbox's distress. For these reasons, only Higgins can be said to behave on the basis of a *moral* emotion. Insofar as the motivation behind Lunchbox's behaviour is not moral, he fails to qualify as a moral subject.

5. Conclusion

We have seen that there are at least two ways in which we can interpret the question of whether morality requires mindreading capacities. We can, firstly, interpret it as the question of whether moral judgement requires mindreading capacities. This is how moral psychologists understand it, and, as we've seen, they tend to consider that mindreading is contingently involved whenever humans engage in moral

[6] I am grateful to an anonymous referee for raising this objection.

judgements. Instead of questioning this claim, I have focused on the second possible way of interpreting the question—namely, as the question of whether mindreading capacities are among the necessary cognitive requirements for moral behaviour to obtain, where the latter is understood as behaviour that has been triggered by a moral motivation. Contrary to what Cova (2013) argues, I have shown that moral responsibility cannot be separated from moral judgement, but that we can, however, construct a minimal account of moral emotions (a type of moral motivation) that is independent of both moral judgement and moral responsibility. Using the framework introduced by Rowlands (2011; 2012), and focusing on the specific case of empathy, I have presented the notion of MME; an emotion that possesses an identifiable moral content, and does not require a capacity to mindread or engage in moral judgements. I have also argued that Nichols (2001) has not provided us with any convincing reason to suppose that MME couldn't provide the adequate motivation for an individual to engage in other-directed helping or comforting behaviour. In the final section, I introduced a systematic definition of MME, and gave a list of reasons why we should consider that whoever behaves on the basis of MME is, indeed, a moral subject. If my arguments have been correct, the case for animal morality should be considered to be independent from the case for animal mindreading.

Department of Logic, History, and Philosophy of Science
Universidad Nacional de Educación a Distancia (UNED),
Madrid

Unit of Ethics and Human-Animal Studies
Messerli Research Institute,
Vienna

References

Baird, J. A. and Astington, J. W. 2004: The role of mental state understanding in the development of moral cognition and moral action. *New Directions for Child and Adolescent Development*, 2004 (103), 37–49. http://doi.org/10.1002/cd.96

Bartal, I. B.-A., Decety, J. and Mason, P. 2011: Empathy and pro-social behavior in rats. *Science*, 334 (6061), 1427–30. http://doi.org/10.1126/science.1210789

Bates, L. A., Byrne, R., Lee, P. C., Njiraini, N., Poole, J. H., Sayialel, K. and Moss, C. J. 2008: Do elephants show empathy? *Journal of Consciousness Studies*, 15(10–11), 204–25.

Bzdok, D., Schilbach, L., Vogeley, K., Schneider, K., Laird, A. R., Langner, R. and Eickhoff, S. B. 2012: Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. *Brain Structure and Function*, 217(4), 783–96. http://doi.org/10.1007/s00429-012-0380-y

Church, R. M. 1959: Emotional reactions of rats to the pain of others. *Journal of Comparative and Physiological Psychology*, 52(2), 132–34. http://doi.org/doi: DOI: 10.1037/h0043531

Clay, Z. and de Waal, F. B. M. 2013: Bonobos respond to distress in others: Consolation across the age spectrum. *PLoS ONE*, 8(1), e55206. http://doi.org/10.1371/journal.pone.0055206

Cools, A. K. A., Van Hout, A. J.-M. and Nelissen, M. H. J. 2008: Canine reconciliation and third-party-initiated postconflict affiliation: do peacemaking social mechanisms in dogs rival those of higher primates? *Ethology*, 114(1), 53–63. http://doi.org/10.1111/j.1439-0310.2007.01443.x

Cova, F. 2013: Two kinds of moral competence: moral agent, moral judge. In B. Musschenga and A. van Harskamp (eds.), *What Makes Us Moral? On the Capacities and Conditions for Being Moral* (pp. 117–30). Springer Netherlands. Retrieved from http://link.springer.com/chapter/10.1007/978-94-007-6343-2_7

Cushman, F. 2008: Crime and punishment: distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, 108(2), 353–80. http://doi.org/10.1016/j.cognition.2008.03.006

Cushman, F., Sheketoff, R., Wharton, S. and Carey, S. 2013: The development of intent-based moral judgment. *Cognition*, 127, 6–21.

Custance, D. and Mayer, J. 2012: Empathic-like responding by domestic dogs (Canis familiaris) to distress in humans: an exploratory study. *Animal Cognition*, 15(5), 851–9.

Darley, J. M., Klosson, E. C. and Zanna, M. P. 1978: Intentions and their contexts in the moral judgments of children and adults. *Child Development*, 49(1), 66–74. http://doi.org/10.2307/1128594

de Vignemont, F. and Jacob, P. 2012: What is it like to feel another's pain? *Philosophy of Science*, 79(2), 295–316. http://doi.org/10.1086/664742

Dixon, B. A. 2008: *Animals, Emotion and Morality: Marking the Boundary*. New York: Prometheus Books.

Douglas-Hamilton, I., Bhalla, S., Wittemyer, G. and Vollrath, F. 2006: Behavioural reactions of elephants towards a dying and deceased matriarch. *Applied Animal Behaviour Science*, 100(1–2), 87–102. http://doi.org/10.1016/j.applanim.2006.04.014

Driver, J. 2006: *Uneasy Virtue* (First paperback version). New York: Cambridge University Press.

Dunn, J., Cutting, A. L. and Demetriou, H. 2000: Moral sensibility, understanding others, and children's friendship interactions in the preschool period. *British Journal of Developmental Psychology*, 18(2), 159–77. http://doi.org/10.1348/026151000165625

Fraser, O. N. and Bugnyar, T. 2010: Do ravens show consolation? Responses to distressed others. *PLoS ONE*, 5(5), e10605. http://doi.org/10.1371/journal.pone.0010605

Gallagher, S. 2012: Empathy, simulation, and narrative. *Science in Context*, 25(03), 355–81. http://doi.org/10.1017/S0269889712000117

Goldman, A. I. 2006: *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading* (1st ed.). New York: Oxford University Press.

Gollwitzer, M. and Denzler, M. 2009: What makes revenge sweet: seeing the offender suffer or delivering a message? *Journal of Experimental Social Psychology*, 45, 840–44.

Gray, K., Young, L. and Waytz, A. 2012: Mind perception is the essence of morality. *Psychological Inquiry*, 23(2), 101–24. http://doi.org/10.1080/1047840X.2012.651387

Guglielmo, S., Monroe, A. E. and Malle, B. F. 2009: At the heart of morality lies folk psychology. *Inquiry*, 52(5), 449–66. http://doi.org/10.1080/00201740903302600

Harenski, C. L., Antonenko, O., Shane, M. and Kiehl, K. A. 2009: A functional imaging investigation of moral deliberation and moral intuition. *NeuroImage*, 49(3), 2707–16. http://doi.org/10.1016/j.neuroimage.2009.10.062

Jamison, L. 2014: *The Empathy Exams: Essays*. Minneapolis, MN: Graywolf Press.

Lakshminarayanan, V. R. and Santos, L. R. 2008: Capuchin monkeys are sensitive to others' welfare. *Current Biology*, 18(21), R999–R1000. http://doi.org/10.1016/j.cub.2008.08.057

Lane, J. D., Wellman, H. M., Olson, S. L., LaBounty, J. and Kerr, D. C. R. 2010: Theory of mind and emotion understanding predict moral development in early childhood. *The British Journal of Developmental Psychology*, 28(Pt 4), 871–89.

Lurz, R. W. 2011: *Mindreading Animals: The Debate over what Animals Know about Other Minds*. Cambridge, Mass.: MIT Press.

Masserman, J., Wechkin, S. and Terris, W. 1964: "Altruistic" behaviour in rhesus monkeys. *American Journal of Psychiatry*, 121(6), 584–5.

Monsó, S. 2015: Empathy and morality in behaviour readers. *Biology and Philosophy*, 30(5), 671–90. http://doi.org/10.1007/s10539-015-9495-x

Nichols, S. 2001: Mindreading and the cognitive architecture underlying altruistic motivation. *Mind and Language*, 16(4), 425–55. http://doi.org/10.1111/1468-0017.00178

Nowbahari, E., Scohier, A., Durand, J.-L. and Hollis, K. L. 2009: Ants, Cataglyphis cursor, use precisely directed rescue behavior to free entrapped relatives. *PLoS ONE*, 4(8), e6573. http://doi.org/10.1371/journal.pone.0006573

O'Connell, S. M. 1995: Empathy in chimpanzees: evidence for theory of mind? *Primates*, 36(3), 397–410. http://doi.org/10.1007/BF02382862

Palagi, E. and Cordoni, G. 2009: Postconflict third-party affiliation in Canis lupus: do wolves share similarities with the great apes? *Animal Behaviour*, 78(4), 979–86. http://doi.org/10.1016/j.anbehav.2009.07.017

Palagi, E. and Norscia, I. 2013: Bonobos protect and console friends and kin. *PLoS ONE*, 8(11), e79290. http://doi.org/10.1371/journal.pone.0079290

Park, K. J., Sohn, H., An, Y. R., Moon, D. Y., Choi, S. G. and An, D. H. 2012: An unusual case of care-giving behavior in wild long-beaked common dolphins

(Delphinus capensis) in the East Sea. *Marine Mammal Science*, 29(4), E508–E514. http://doi.org/10.1111/mms.12012

Piazzarollo Loureiro, C. and de Hollanda Souza, D. 2013: The relationship between theory of mind and moral development in preschool children. *Paidéia (Ribeirão Preto)*, 23(54), 93–101. http://doi.org/10.1590/1982-43272354201311

Plotnik, J. M. and de Waal, F. B. M. 2014: Asian elephants (Elephas maximus) reassure others in distress. *PeerJ*, 2(e278). http://doi.org/10.7717/peerj.278

Porter, J. 1977: Pseudorca stranding. *Oceans*, 10, 8–15.

Preston, S. D. and de Waal, F. B. M. 2002: Empathy: its ultimate and proximate bases. *Behavioral and Brain Sciences*, 25(1), 1–20; discussion 20–71.

Rowlands, M. 2011: Animals that act for moral reasons. In T. Beauchamp and R. G. Frey (eds.), *Oxford Handbook of Animal Ethics*. New York: Oxford University Press.

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

Rowlands, M. and Monsó, S. 2017: Animals as reflexive thinkers: the aponoian paradigm. In L. Kalof (ed.), *The Oxford Handbook of Animal Studies* (pp. 319–342). New York: Oxford University Press. http://doi.org/10.1093/oxfordhb/9780199927142.013.15

Rutte, C. and Taborsky, M. 2007: Generalized reciprocity in rats. *PLoS Biology*, 5(7), 1421–5. http://doi.org/10.1371/journal.pbio.0050196

Sato, N., Tan, L., Tate, K. and Okada, M. 2015: Rats demonstrate helping behavior toward a soaked conspecific. *Animal Cognition*, 18(5), 1039–47. http://doi.org/10.1007/s10071-015-0872-2

Seed, A. M., Clayton, N. S. and Emery, N. J. 2007: Postconflict third-party affiliation in rooks, Corvus frugilegus. *Current Biology: CB*, 17(2), 152–8. http://doi.org/10.1016/j.cub.2006.11.025

Vasconcelos, M., Hollis, K., Nowbahari, E. and Kacelnik, A. 2012: Pro-sociality without empathy. *Biology Letters*, 8(6), 910–12. http://doi.org/10.1098/rsbl.2012.0554

Woolfolk, R. L., Doris, J. M. and Darley, J. M. 2006: Identification, situational constraint, and social cognition: studies in the attribution of moral responsibility. *Cognition*, 100(2), 283–301.

Young, L., Cushman, F., Hauser, M., & Saxe, R. 2007: The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences*, 104(20), 8235–8240. http://doi.org/10.1073/pnas.0701408104

Young, L. and Saxe, R. 2008: The neural basis of belief encoding and integration in moral judgment. *NeuroImage*. http://doi.org/doi:10.1016/j. neuroimage.2008.01.057

Young, L. and Waytz, A. 2013: Mind attribution is for morality. In S. Baron-Cohen, M. Lombardo and H. Tager-Flusberg (eds.), *Understanding Other Minds: Perspectives from Developmental Social Neuroscience* (pp. 93–103). New York: Oxford University Press.

Zahavi, D. 2010: Empathy, embodiment and interpersonal understanding: from Lipps to Schutz. *Inquiry*, 53(3), 285–306. http://doi.org/10.1080/00201741003784663