

## Two Kinds of Moral Competence: Moral Agent, Moral Judge

COVA, Florian

### Abstract

In this paper, I argue that some of the disagreements about the continuity or discontinuity of human moral life with that of animals can be assuaged by drawing a distinction between two senses in which someone can be a 'moral being': being a moral agent (i.e. being morally responsible for one's action) and being a moral judge (i.e. being able to form moral judgments). More precisely, I argue that it is not necessary to be a moral judge to be a moral agent, because moral actions (actions we are morally responsible for) don't need to stem from moral judgments. Consequently, I argue that, even if moral judgment is highly likely to be a human specificity, moral agency is something that we might share with other animals, given that the only requisite to be a moral agent is to be able to be motivated by the fact that other entities do have interests.

### Reference

COVA, Florian. Two Kinds of Moral Competence: Moral Agent, Moral Judge. In: Musschenga, B. & van Harskamp, A. *What Makes Us Moral? On the capacities and conditions for being moral*. Dordrecht : Springer Netherlands, 2013. p. 117-130

DOI : 10.1007/978-94-007-6343-2\_7

Available at:

<http://archive-ouverte.unige.ch/unige:109903>

Disclaimer: layout of this document may differ from the published version.



UNIVERSITÉ  
DE GENÈVE

# Metadata of the chapter that will be visualized online

Series Title		
Chapter Title	Two Kinds of Moral Competence; Moral Agent, Moral Judge	
Chapter SubTitle		
Copyright Year	2013	
Copyright Holder	Springer Science+Business Media Dordrecht	
Corresponding Author	Family Name	<b>Cova</b>
	Particle	
	Given Name	<b>Florian</b>
	Suffix	
	Division	Swiss Center for Affective Sciences
	Organization	University of Geneva
	Address	Geneva, Switzerland
	Email	florian.cova@gmail.com
Abstract	In this paper, I argue that some of the disagreements about the continuity or discontinuity of human moral life with that of animals can be assuaged by drawing a distinction between two senses in which someone can be a ‘moral being’: being a moral agent (i.e. being morally responsible for one’s action) and being a moral judge (i.e. being able to form moral judgments). More precisely, I argue that it is not necessary to be a moral judge to be a moral agent, because moral actions (actions we are morally responsible for) don’t need to stem from moral judgments. Consequently, I argue that, even if moral judgment is highly likely to be a human specificity, moral agency is something that we might share with other animals, given that the only requisite to be a moral agent is to be able to be motivated by the fact that other entities do have interests.	

# Chapter 7

## Two Kinds of Moral Competence; Moral Agent, Moral Judge

Florian Cova

### 7.1 What Makes Us Moral? And the Continuism/Discontinuism Debate

There are many ways of understanding the Big Question ‘what makes us moral?’ One way to understand it is to understand it as bearing on the psychological capacities that allow us to be moral beings. Thus, it can be paraphrased in the following way:

(BQ) *What psychological capacities allowed us to become moral beings?*

Note that the question is not what capacities are *necessary* for a being to count as a moral being. For all we know, there might be different capacities that are separately sufficient, but not necessary, to make a being a moral being. What we are interested in here is what capacities *actually* made us (and continue to make us) moral beings.

This question is undoubtedly fascinating, and this is why it is currently investigated by philosophers, psychologists, neuroscientists, evolutionary biologists, economists, anthropologists and others I surely forgot. It would be preposterous to claim that I actually have the answer to this question, so my goal won’t be to answer it. My goal is rather to point out that a preliminary question must be answered if we are to find the answer to our Big Question. To discover what made us moral, we must first understand what it is ‘to be moral’ and what it means to be a ‘moral being’.

Indeed, answers to (BQ) come into two very different forms: *Continuists* believe that the psychological capacities required for being a moral being are not unique to humans and that at least rudimentary traces of these capacities can be found in other species (most likely apes). *Discontinuists*, on the contrary, believe that these psychological capacities are unique to humans and that human beings are the only moral beings we know of.

F. Cova (✉)

Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland  
e-mail: [florian.cova@gmail.com](mailto:florian.cova@gmail.com)

A paradigmatic example of this debate is offered in Frans de Waal's book *Primates and philosophers: How morality evolved*. In this book, De Waal takes a continuist stance in describing the school of thought he defends:

[This school] views morality as a direct outgrowth of the social instincts that we share with other animals. In the latter view, morality is neither unique to us nor a conscious decision taken at a specific point in time: it is the product of social evolution. (De Waal 1997, p. 6)

De Waal champions this theory by pointing to the different psychological capacities he claims we share with apes (perspective taking, emotions like gratitude) and to reports of 'moral behaviours' (help, consolation). This line of argument directly clashes with the conception of certain of his respondents. For example, Christine Korsgaard (De Waal 1997, p. 104) claims that 'morality represents a break with our animal past'. And indeed, in accordance with this conception, some psychologists have argued that morality is the product of new psychological capacities that are not to be found in non-human animals, some 'sense of fairness' (Baumard 2010) or 'universal moral grammar' (Mikhail 2011).

These two theses seem to me equally plausible, though they seem to cancel one another: when someone describes the delightful stories of seemingly moral behaviour in the animal realm, we are moved, and find the conditions set by the first theory too high. But when we listen to advocates of the human specificity, we can't help notice that there's really something special about human beings, i.e. the complexities of their interrogations about what is right or wrong. After all, we are the only known species to do moral philosophy.

Here, I want to loosen this tension by suggesting that the disagreement between continuists and discontinuists can be (partly) resolved because it (partly) stems from the conflation between two ways of counting as a moral being. First, one can be a moral being because one is morally responsible of (some of) his action: thus, one is a moral being in the sense of being a *moral agent*. Then, one can be a moral being in the sense that one is able to judge whether something is right or wrong: in this sense, one is a moral being in the sense of being a *moral judge*. Indeed, once this distinction made one can argue that both sides are right (and wrong) respective to one sense of counting as a moral being. More precisely, I will argue that it is possible to be a continuist about moral agency while being a discontinuist about moral judgment, and argue that we share the psychological bases for moral agency with other animals while we are the only known species able to form moral judgments.

Consequently, the main point I propose to defend here will be that these two ways of being moral can be dissociated and that one can be a moral agent without being a moral judge. I won't try to argue for a *double* dissociation and to prove that there are also cases of moral judges that are not moral agents. Maybe such cases can be found in children (that we do not always judge responsible for their action but that are able to form at least rudimentary moral evaluations)<sup>1</sup> or in patients suffering from

<sup>1</sup>Famous psychological studies led by Turiel suggest that, by the age of 4, children are able to understand moral concepts and use them to form moral judgments. For example, they are

particular neuropsychological deficits.<sup>2</sup> Nevertheless, such cases being about non-  
matured or pathological individuals that are not representative of the full potential  
of their species, they won't be relevant to the present discussion. I will then stick  
to the more plausible claim that a species can be endowed with the psychological  
requirements for moral agency while being unable to form moral judgment.

## 7.2 The Epistemic Argument Against the Moral Agency/Moral Judgment Dissociation

I will take as a starting point that it is not *prima facie* implausible that one can  
be a moral agent without being a moral judge. As a proof, we can see that one of  
the foundational myths of occidental culture, in which many people have believed,  
relies on such a dissociation. Let's go back to the famous episode of *Genesis*, when  
Adam and Eve eat the forbidden fruit. What happens then? What is said is that, by  
eating the fruit, their 'eyes opened' and they came to 'know good and evil' (*Genesis*  
3:5). Now, remember that Adam and Eve are finally punished for having eaten the  
fruit: so, by eating the fruit, they did something wrong. Nevertheless, at the time  
they weren't able to recognize right from wrong yet. So, what is suggested by this  
story is that a being can be morally responsible for an action he performed at a time  
he wasn't able to recognize right from wrong.

Once granted that it is not obvious that moral judgment is required for moral  
agency and that one cannot be a moral agent without being a moral judge, I will  
argue that the burden of the proof lies with those who consider moral judgment as a  
requirement for moral agency. The thesis that one cannot be a moral agent without  
being a moral judge amounts to the thesis that the two concepts of 'moral agency'  
and 'moral ignorance' are incompatible.<sup>3</sup> But, since each concept can be understood  
separately, and since two logically independent concepts are *prima facie* compatible,  
it seems to be the position by default that moral agency and moral ignorance are  
compatible. Thus, if there is no argument in favour of the thesis that moral agency  
and moral ignorance are incompatible, we are justified in thinking that they are  
compatible.

---

able to distinguish moral rules (that are universally valid and independent from authority) from  
conventional rules (that are only locally valid and dependent on authority) or prudential rules (see  
Turiel 2002 for a review).

<sup>2</sup>For example, patients with lesions in the prefrontal cortex are still able to make moral judgments,  
but are much less prone to act according to these judgments, due to emotional and motivational  
deficits (Damasio 1995). Patients suffering from aboulia, who have lost all motivation, are also  
plausible cases of people able to form moral judgments but lacking moral motivations (Marit and  
Wilkosz 2005).

<sup>3</sup>I define 'moral agency' as the ability to be morally responsible for (some of) one's action and  
'moral ignorance' as the inability to judge something as morally right or wrong.

Nevertheless, there are such arguments. So, before concluding that moral agency and moral ignorance are compatible, we must examine these arguments. I will distinguish two main kinds of arguments, each one corresponding to a certain kind of condition for moral responsibility. It is usual in the literature about freedom and moral responsibility to distinguish two categories of conditions for moral responsibility: the *epistemic* conditions (what one has to know for being responsible for one's action) and the *freedom* conditions (the type of control one has to exert on one's action for being responsible).

As far as the epistemic conditions are concerned, it is commonplace to say that one had to know and understand what one was doing for being morally responsible. Some have understood this condition as implying that one must be able to understand the moral significance of one's action to be morally responsible for them. But if this is true, then moral judgment is a condition for moral agency, and one cannot be a moral agent without the capacity to form moral judgment.

### 7.2.1 The Epistemic Conditions for Moral Responsibility

To give an example of the debate about the epistemic conditions for moral responsibility, let's have a look at the debate on the moral responsibility of psychopathic individuals. Psychological studies have recently suggested that psychopaths are complicated cases: it is not even sure that they are able to form genuine moral judgments. Studies suggest that they fail to distinguish between moral rules and conventional rules, a distinction 'normal' people master around the age of 4.<sup>4</sup> For this reason, people like Levy have suggested that we should not hold psychopaths responsible for their action: 'Moral responsibility requires moral knowledge; because psychopaths lack this knowledge through no fault of their own, we must refrain from blaming them. Psychopaths are victims, as well as victimizers' (Levy 2008, p. 136).

The principle according to which moral responsibility requires moral knowledge is widely endorsed. For example, a similar principle is part of the *Model Penal Code* of the American Law Institute:

A person is not responsible for criminal conduct if at the time of such conduct as a result of mental disease or defect he lacks substantial capacity either to appreciate the wrongfulness of his conduct or to conform his conduct to the requirements of the law.

Now, this principle is just a way among others to state that moral agency requires moral knowledge – but what reasons do we have to endorse this principle?

One reason I see is that this principle seems to correspond to a kind of excuses we use to accept for wrongdoing. For example, someone who has harmed another accidentally can say that 'he didn't know he was doing something wrong'. I don't think that this is enough to support this principle. Surely, if we imagine that someone

<sup>4</sup>See Blair (1995, 1997) and footnote 1 in this text.

offers a child a peanut butter sandwich without knowing that the child is allergic to peanut butter, and that this action results in the child being driven to the hospital, we would accept an excuse such as ‘I didn’t know I was doing something wrong’ (meaning: I didn’t know that my action would have such dreadful consequences). But we would be less willing to accept it if the person actually intended to kill the child by poisoning her with peanut butter and tried to escape blame by saying she didn’t know that poisoning was wrong.

Contrasting these two examples, we can draw a distinction already made by Aristotle (*Nicomachean Ethics*, III, 2) between two kinds of moral ignorance: ignorance of the particular circumstances of an act (what exactly has happened) and ignorance of the moral value of this act (if the act was permissible or forbidden). According to him, only the first kind of ignorance can be a valid excuse, the second kind of ignorance being the mark of vice. Accordingly, things are complicated because Aristotle also thinks that, ultimately, moral ignorance in this second sense is the result of the agent’s choice. So, if moral ignorance in this sense is not an exculpatory condition, that would be because the agent is ultimately responsible for it. I do not wish to enter in an exegesis of Aristotle’s conception of moral ignorance: all I’m interested in here is the distinction between moral ignorance of the particular circumstances of an action (*factual ignorance*) and moral ignorance of the moral value of this action (*moral ignorance* in the strong sense). This distinction allows us to see that even if it is intuitive to say of someone that he should be forgiven because he didn’t know what he was doing, such expressions are more commonly used in cases of factual ignorance, when we excuse an agent because he didn’t have a full understanding of the impact and consequences of his action.

Thus, I don’t think it is obvious that moral ignorance is a mitigating circumstance. Though this has been accepted as a valid principle by philosophers and courts, the idea seems to have no clear intuitive appeal. One explanation for its acceptance is that it is the product of an overgeneralization of the intuitive principle according to which ignorance of the nature of an action can be a valid excuse, that ended up including in the ‘nature’ of an action its moral value.

### 7.2.2 Moral Knowledge and Acting for Good Reasons

Another reason to endorse this principle is that it might help us explain the difference between good actions done from good intentions (that are praiseworthy) and good actions that are done without good intentions (and often are not praiseworthy). Let’s take the following (paradigmatic) example<sup>5</sup>:

(Pond) As a man walks by a pond, he notices a young child drowning. He dives into the pond to save the child and brings the child back to the shore.

<sup>5</sup>This case is inspired by Singer (1972).

Now, should we praise this man for having saved the child? It depends on *why* he saved the child. As a philosopher will justly point out, we shouldn't (and wouldn't) praise this man if he didn't care at all about the child and saved it only because he hoped to receive a reward or to be considered as a hero. This means that, to be praiseworthy, our man must have acted *for the good reasons*.

But what does it mean to act *for the good reasons*? Following Kant, a philosopher could advance the following thesis:

(Reflectivism) An agent deserved to be praised (or blamed) for having performed a given action A only if he did so because he considered this action to be good (or thought this action to be bad).

This principle gives a very simple answer to our problem: the man who saves the child only because he cared about the reward does not deserve praise because he did not do it because he thought it to be good (i.e. by reason of respect for the moral law, to use a Kantian expression). Though very simple, Reflectivism has an important implication: that an agent can be morally responsible for an action only if he has the capacity to judge something as morally good (or wrong). Thus, the need to distinguish between praiseworthy good actions and good actions not performed for the good reasons might be a support for this principle, because it would be the best way to distinguish the two kinds of good actions.

But are we forced to endorse Reflectivism if we are to make a difference between these two kinds of good actions? I don't think so. To show why, I'll propose a number of counter-arguments. All of them rely on appeal to intuitions, and so are not strong enough to conclude that Reflectivism is false. Nevertheless, I'll argue they are sufficient to show that we have no reason to endorse Reflectivism.

Let's start with (i) the *argument from impulsivity*. Here is a slightly modified version of Pond:

(Pond\*) As a man walks by a pond, he notices a young child drowning. Understanding that the child will soon die, he immediately dives into the pond to save the child, without taking the time to figure whether it's the right thing to do. He succeeds in retrieving the child.

And here is a second variant:

(Pond\*\*) As a man walks by a pond, he notices a young child drowning. Due to his strict moral education, this man has taken the habit (and is motivated) to do what he thinks is right. Realizing that the children will soon die and that saving him would be something right, he dives into the pond to save the child. He succeeds in retrieving the child.

These two men have different reasons for acting. Our first man doesn't take time to think: he doesn't classify his action as 'right' before acting. If we asked him about his reasons for acting, he would answer something like: 'because the child would have died'. Our second man goes through some kind of moral reasoning, and saves the child because he categorizes this action as 'right'. If we asked him about his reasons for acting, he would answer: 'because it was the right thing to do'. According to Reflectivism, only the second one should be considered praiseworthy, because he's the only one to act on the basis of a moral judgment (the judgment that it is right to save the child and wrong to let him die).



But does this conclusion seem right? I don't think so. I rather think that, if we compare those two cases, the man in Pond\* is at least as praiseworthy as the man in Pond\*\*. In fact, there even seems to be something wrong with our second man: shouldn't he be more concerned about the child rather than whether it is right or wrong to save him?

Reflectivism, in the way I presented it, seems to discredit every good action that would look like a 'moral reflex'. If your friend is about to stumble, and you instinctively grab him to prevent his fall, then you shouldn't be praised for your action, since you did that only because you cared about your friend, and not because you judged that helping your friend was the right thing to do. And Reflectivism doesn't stop there: it also discredits actions that come from emotional reactions and are not mediated by moral reasoning. If a man, seeing a homeless person freezing in the winter is suddenly overwhelmed by compassion and gives him his coat, without wondering whether this is right or not, but just wanting to help this particular person, then Reflectivism should conclude that his action is not praiseworthy. However, we tend to praise and even to be moved by such actions. Finally, Reflectivism seems to lead to the conclusion that, when a friend or a parent helps us, he is all the more praiseworthy for helping us because it was the right thing to do. But, this doesn't seem right: we do not want our friends or our parents to help us because they think it's the right thing to do – we want them to help us because they actually *care* about us. So, to sum up, Reflectivism goes against most of our basic moral appreciations.<sup>6</sup>

One could object that we focused on cases in which we praise agents for right actions, and that Reflectivism is much more plausible when it comes to wrong actions: is it not intuitive that we shouldn't blame (and punish) those who didn't know that what they did was wrong?

This is far from clear. Let's imagine the following case (drawn from Pizarro et al. 2003):

(Smash) Because of his overwhelming and uncontrollable anger, Jack impulsively smashed the window of the car parked in front of him because it was parked too close to his.

Let's say that Jack had a bad day, was irritated, and smashed the window 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.<sup>7</sup>

<sup>6</sup>One might say that these appreciations are not really *moral* and that the praise we attribute agents for caring about their relative has nothing to do with moral praise. This is indeed a possibility; nevertheless it seems me very unlikely: a mother who doesn't care about her children elicits from us blame that is very likely *moral* blame.

<sup>7</sup>The results obtained by Pizarro et al. show that many participants considered Jack responsible and blameworthy for having smashed the window. That Jack had not taken the time to realize that what he was doing was wrong was no consideration.

Another argument that could be opposed to Reflectivism are the cases of ‘inverse akrasia’. By ‘inverse akrasia’, Arpaly and Schroeder (1999) mean cases in which the agent fails to do what he thinks is right and does what he thinks is wrong, but in which we actually consider wrong what he failed to do and right what he actually did. A famous example is the case of Huckleberry Finn, who keeps helping Jim the runaway slave while thinking that the right course of action is to return the slave to his lawful owner. In this case, Kantianism should predict that we shouldn’t praise Huckleberry for helping Jim, since he doesn’t think that it is the right thing to do.

Still, many people still judge that Huckleberry is responsible and deserves praise for his actions. So, it is still not intuitive that one has to know that he’s doing something right (or wrong) to be responsible for his actions.

One might object that inverse akrasia is no definitive objection to Reflectivism if we switch to a weaker version according to which for an agent to be morally responsible for his actions, this agent has just to assess the moral value of what he’s doing, but not to *correctly* assess this value. If we drop the ‘correctly’, then Reflectivism can accommodate these cases, because Huckleberry (incorrectly) assesses the moral value of their actions.

This less demanding version still can’t account for cases of impulsive actions. Surely, Reflectivism could once again lower its demands to become able to accommodate such cases. For example, it could say that people don’t actually have to act upon the basis of moral judgments to be responsible for their actions – but that they must only have the ability to judge that what they have done is right or wrong. In this version however, Reflectivism loses one of its main advantages and motivations: it can no longer explain the difference between the man who saves the child because he cares about it and the man who saves the child because he expects a reward – both have the ability to judge *a posteriori* that what they did (saving the child) was the right thing to do. Of course, Reflectivism can try to account for this difference by another feature of these cases, but if it does so, then we have no longer any reason to endorse it, because our reason to endorse it was to account for this difference.

Another objection to this weaker version of Reflectivism is that it is hard to see why the mere fact of having this ability should make a difference if the fact of using it doesn’t. We can compare this objection to Frankfurt’s famous argument against the Principle of Alternate Possibilities (Frankfurt 1969). The Principle of Alternate Possibilities is a principle that states that an agent cannot be morally responsible for what he did if he couldn’t have done otherwise. Against this principle, Frankfurt has us imagine cases in which an agent has no alternative (and thus cannot do otherwise), but in which this fact doesn’t motivate the subject (e.g. John decides by himself to kill his wife while, unbeknownst to him, Black has implanted in his brain a chip that would have compelled John to kill his wife if he hadn’t chosen to do it by himself). In this case, according to Frankfurt, the agent’s action is exactly the same as in a parallel case in which he has an alternative (the case in which John decides by himself to kill his wife but Black and the chip do not exist). So, Frankfurt asks, why should we make any difference between these two actions? If John is responsible when he has an alternative, he should also be responsible when he has

none, provided that the lack of alternative does not cause his action in any way. On 295  
this model, we can take our case Pond\* and imagine a parallel case Pond\*\*\* in 296  
which the agent has an innate psychological defect that prevents him to understand 297  
what is good and what is bad. Since in Pond\* such knowledge doesn't play any role 298  
in the production of the action, why should we treat the agent's action in Pond\*\*\* 299  
differently and deny praise to the agent? 300

To follow this idea further, let's imagine an individual who, due to very specific 301  
and focal brain damages he suffered at birth, is deprived of moral concepts: 302  
he doesn't understand what words like 'right', 'wrong', 'fair' or 'unfair' mean. 303  
This individual has no other cognitive issues whatsoever: he's fully capable of 304  
understanding what he's doing and the consequences of this action. Let's now 305  
say that this individual is also (and was before suffering from brain damages) a 306  
dangerous criminal: he enjoys kidnapping people and takes pleasure torturing them. 307  
Remember that he perfectly understands what he's doing: he knows what pain is, 308  
and he knows that his victims suffer from the treatment he inflicts on them. He 309  
also knows that they don't want to suffer and that torture is a highly traumatizing 310  
experience. He just doesn't know whether it is 'right' or 'wrong' to torture people. 311  
Is it obvious that this individual is not responsible for what he did? And that he 312  
doesn't deserve blame? I don't think so. We have the feeling that, even if he can't 313  
form moral judgment about his actions, this individual is a terrible person. He does 314  
something wrong (torturing people) for clearly vicious motives (because he enjoys 315  
watching people suffering). 316

Let's now imagine the reverse individual (what I call a 'reverse psychopath'): 317  
a man who does not have access to moral concepts but deeply cares about other 318  
persons (because he's deeply empathetic), so that he spends most of his time helping 319  
people. Should we be reluctant to praise him? Should we consider that the people 320  
he helps don't have to be grateful? That seems a wrong conclusion.<sup>8</sup> 321

Of course, these arguments are far from being inescapable: they are mostly 322  
appeals to our intuitions. But the fact that our intuitions go against Reflectivism is 323  
*prima facie* a reason to reject it, unless its adherents provide a convincing argument. 324  
And if we have no reason to endorse Reflectivism, then I can't think of any other 325  
reason to accept the thesis according to which moral ignorance is incompatible with 326  
moral agency. 327

---

<sup>8</sup>One could consider that these example are not sufficient because it is one thing to evaluate persons (as nice or vicious) and another to attribute them responsibility for their action. This is true. Nevertheless, I think we also have the intuition in those cases that these persons are responsible for their actions. Let's take the reverse psychopath and imagine that he helps you (by saving your life). It seems quite natural to feel grateful and consider that you owe him something. But this feeling cannot be accounted by a mere judgment of 'niceness' (you are not indebted to all people you find nice): rather, you would feel indebted in such a case because the reverse psychopath deserves credit for what he did – that is to say: because he was responsible for what he did.

### 7.3 Reasons and the Freedom Conditions for Moral Responsibility

328

329

In the previous section, I have examined what in the epistemic conditions for moral responsibility could make moral ignorance incompatible with moral agency. Nevertheless, considerations about the freedom conditions for moral responsibility could give support to the thesis that moral agency requires the capacity to form moral judgment.

There are many incompatible ways of framing the freedom conditions for moral responsibility – I will focus here on its definition as ‘reason-responsiveness’. Fischer and Ravizza (1998) have given ‘reason-responsiveness’ as a criterion for freedom: basically, the idea is that we can identify whether a given action is free by considering whether the agent would have acted differently had his reasons been different. If I would have had a good reason not to get out of bed this morning, for instance, I would not have got out of bed. Thus we can say that my getting out of bed this morning is something I did freely. To be free is to act on the basis of reasons.

The question is now what it means to act on the basis of reasons. In a strong reading, acting on the basis of reason R amounts to act because we consider R to be a good reason. But assessing whether a reason is a good reason or not seems to immediately drive us in the realm of practical reason and of morality, and thus to make moral evaluation of reasons a requirement for acting for reasons and moral responsibility. Such a strong reading can be found in Korsgaard’s conception of morality:

Our purposes may be suggested to us by our desires and emotions, but they are not determined for us by our affective states, for if we had judged it wrong to pursue them, we could have laid them aside. Since we choose not only the means to our ends but also the ends themselves, this is intentionality at a deeper level. For we exert a deeper level of control over own movements when we choose our ends as well as the means to them than that exhibited by an animal that pursues ends that are given to it by its affective states, even if it pursues them consciously and intelligently. Another way to put the point is to say that we do not merely have intentions, good or bad. We assess and adopt them. We have the capacity for normative self-government, or, as Kant called it, ‘autonomy’. It is at this level that morality emerges. The morality of your action is not a function of the content of your intentions. It is a function of the exercise of normative self-government. (De Waal 1997, p. 112)

According to Korsgaard, being moral implies to be reason-responsive in a particular way, i.e. to act according to our reasons and the judgments we make about them. Under such a strong reading of reason-responsiveness, the ability to form moral evaluations is a key component of moral agency.

However, I would like that a weaker account of reason-responsiveness is sufficient to capture our intuitions about moral agency. Let’s first return to the problem mentioned in the second section: why do we praise the man who saves the child ‘because the child would have died’ but not the man who saves the child ‘to become famous or get a reward’? The Reflectivist answer was that the man who saves the child for a reward does not act on the basis of a moral judgment. Nevertheless, we rejected this condition as too high: the man who impulsively dives

to save the child does not act upon the basis of a moral judgment either – still, he deserves praise for his actions.

Reflectivism might be on the right track when it says that a moral agent deserves praise when he acts *for good reasons*. But a good reason doesn't need to be the judgment that an action is morally good. The man who dives 'because the child would have died' acts precisely for a *good* reason, though 'the child would have died' is not a moral judgment.

So, what makes a reason a morally good reason? I think that what makes our man praiseworthy is that he's motivated by the fact he somehow *cares* about the child: he dives because he wants to prevent the child to be harmed. Thus, one acts for the good reasons when one actually cares about the person one is trying to help. In the same way, Huckleberry Finn deserves praise for helping Jim, since he helps Jim because he *cares* about him, whatever he might think of the moral appropriateness of such an attitude. Similarly, our 'reversed psychopath' also deserves praise for helping people because, though he doesn't realize that what he's doing is right, he *cares* about people.

The question is now: what does it take to be able to *care* about someone? One condition is to be able to understand that this person has *interests* – that some things have (positive or negative) values for this person. For example, if you understand that a person can suffer and doesn't want to suffer, then you understand that it is in his interests not to suffer. Nevertheless, understanding that persons have interests is not enough to actually care about them. For example, psychopaths totally understand that people have interests, but do not care about them. To care about someone is to give an intrinsic (i.e. non-instrumental) positive value to the fact that this person's interests are preserved and augmented, and to give an intrinsic negative value to the fact that this person's interests are damaged.

So, to deserve praise for a right action involving a moral patient, one just has to perform this action because one cared about this moral patient's interest. This is why, the case of a man that would help the child because he has been taught it's the right thing to do but not because he cares about the child is so disturbing.

These conditions also allow us to determine when an agent deserves praise for punishing a criminal that harmed a moral patient. One does not immediately deserve praise for 'having done justice'. For example, if I punish a criminal not because I care about what he did to the victim, but because I just like punishing people, I don't deserve praise – here again, I must care about the moral patient, the victim, to deserve praise for what I did.

Going further, it also allows us to understand what it takes to deserve blame for having done something wrong. To deserve blame, I must harm a person and, though I understand this person has interests, not care about them. For example, if I stomp on someone's toes, just because I'm in a hurry and don't care about harming this person, it's enough for me to deserve blame. Otherwise said, I deserve blame when my action is expressive of the fact I don't care about others.<sup>9</sup>

<sup>9</sup>For a development of the psychological theory underlying this account of moral responsibility, see Cova et al. ([submitted](#)).

My proposal is then that being a moral agent only requires understanding that others have interests and the capacity to be motivated by this understanding (that is: to act in accordance with how much I care about others). If I don't care about these interests, I have bad motives. If I care about these interests, I have good motives – but the main point is to understand that some entities have *interests*, and that makes them moral patients. All it takes to be a moral agent is thus to be able to act according to what we attach importance to and a bit of theory of mind.<sup>10</sup>

With this theory, we can construe a less demanding reason-responsiveness for the freedom condition: to be reason-responsive is to be able to be sensitive to the well-being of others, and thus able to act (or refrain from acting) on the reason that someone's well-being is at stake.<sup>11</sup> But this doesn't mean that agents are required to act because they consider that improving the well-being of others is a good thing. If it did, that would run against our intuitions about the Pond cases presented in Sect. 7.2.

Thus, we think there is a plausible version of the epistemic and freedom conditions for moral responsibility according to which moral responsibility might not require the ability to form moral judgment. In the last section of the paper I return to the implication of such a possibility.

## 7.4 Conclusion: Moral Animals and Twice-Moral Human Beings

I began this paper by presenting a conflict between the continuists, who think we share with animals the psychological requirements for morality, and discontinuists, who think that we are endowed with unique moral capacities. In this paper, I've argued that both perspectives could be reconciled by distinguishing two components of our moral life: moral agency (we are morally responsible for our actions) and moral judgments (we are able to evaluate our behaviour and that of others).

To side with the discontinuist, it is hard to deny that moral life is much richer in human beings than in any other moral animals: we are able to ask tough moral questions and reason about difficult moral situations (such as moral dilemmas). Nevertheless, the continuist can also cite cases in which it is hard to deny that other species can have a real moral life. Let's consider the following story:

<sup>10</sup>Some (e.g. Knobe 2006) have argued that our theory-of-mind is already suffused with moral considerations and evaluations, which goes directly against our argument that presupposes that theory of mind is independent from the faculty of moral judgment. However, there are reasons to doubt this claim. See Cova et al. (2010) for a rebuttal.

<sup>11</sup>Does it mean that empathy is necessary to be a moral agent? No, empathy is what makes us care about others' interests and be *good* moral agents. But, a *bad* moral agent, one who doesn't care about others, is still a moral agent, so empathy is not necessary to be a moral agent. Note also that empathy might not even be necessary to be a good moral agent: there might be other emotional or cognitive ways to care about others' interests.

The Arnhem chimpanzees spend the winters indoors. Each morning, after cleaning the hall and before releasing the colony, the keeper hoses out all the rubber tires in the enclosure and hangs them one by one on a horizontal log extending from the climbing frame. One day Krom was interested in a tire in which water had been retained. Unfortunately, this particular tire was at the end of the row, with six or more heavy tires hanging in front of it. Krom pulled and pulled at the one she wanted but could not move it off the log. She pushed the tire backward, but there it hit the climbing frame and could not be removed either. Krom worked in vain on this problem for over ten minutes, ignored by everyone except Otto Adang, my successor in Arnhem, and Jakie, a seven-year-old male chimpanzee to whom Krom used to be the 'aunt' (a caretaker other than the mother) when he was younger. Immediately after Krom gave up and walked away from the scene, Jakie approached. Without hesitation he pushed the tires off the log one by one, as any sensible chimpanzee would, beginning with the front one, followed by the second in the row, and so on. When he reached the last tire, he carefully removed it so that no water was lost and carried the tire straight to his aunt, where he placed it upright in front of her. Krom accepted his present without any special acknowledgment and was already scooping water with her hand when Jakie left. (De Waal 1997, p. 83)

In this case, it seems that Jakie understood what Krom wanted, and helped her get it, and it is hard not to perceive this story in moral terms. According to the account developed in Sect. 7.3, it is the fact that Jakie understood that Krom had interests that makes him look like a moral agent.<sup>12</sup>

By stressing that moral agency can exist without moral judgment, I hope to have contributed to diminishing the gap between continuists and discontinuists. Indeed, it is interesting to note that continuists often emphasize action and emotions, while discontinuists stresses the uniqueness of human moral reflexion. These two insights can be reconciled in a more complex and fine-grained view of moral life, a view that opens the interesting possibility that moral agency can have evolved independently from moral judgment.

<sup>12</sup>This claim immediately raises several questions. (i) First, if Jakie is really a moral agent, does that immediately make him a moral patient (i.e. someone it is wrong to harm). I was tempted to say 'yes' until an anonymous reviewer gave me the following argument I found quite convincing: 'Are all moral agents moral patients? Probably, but: one could imagine someone who has racist beliefs and thinks, for example, that black people are inferior to white people, and thus think that they do not bear rights in the same way and to the same extent as white people, yet think that they are equally morally responsible for their actions, and are bound by the same norms. In this case, being a moral agent would not be sufficient for being a moral patient. This option does not seem conceptually incoherent.' (ii) Second, if Jakie can be responsible for his action, does it necessarily entail that he can be punished? I am not sure either, for punishment seems to require something more than moral responsibility. For example, it seems to me that we want the people we punish to understand why they are punished and that it is essential for punishment that the punished one understands it as such. This intuition is supported by experimental studies showing 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 (Gollwitzer and Denzler 2009). If it is the case, then one has to be both a moral agent and a moral judge to be an appropriate target of punishment. (Even if you take a consequentialist stance on punishment, then you must admit that an agent that cannot understand why he has been punished and on whom deterrence won't work is not a suitable target of punishment.)



**Acknowledgments** This research was supported in part by a Grant from the French Agence 474  
Nationale de la Recherche (ANR) (ANR Blanche: SoCoDev). I thank François Jacquet and an 475  
anonymous reviewer for useful comments on previous versions of this paper. 476

## References

477

- Aristotle. 2011. *Nicomachean ethics*. Chicago: University of Chicago Press. 478
- Arpaly, N., and T. Schroeder. 1999. Praise, blame and the whole self. *Philosophical Studies* 93(2): 479  
161–188. 480
- Baumard, N. 2010. *Comment nous sommes devenus moraux: Une histoire naturelle du bien et du* 481  
*mal*. Paris: Odile Jacob. 482
- Blair, R.J. 1995. A cognitive developmental approach to morality: Investigating the psychopath. 483  
*Cognition* 57: 1–29. 484
- Blair, R.J. 1997. Moral reasoning and the child with psychopathic tendencies. *Personality and* 485  
*Individual Differences* 26: 731–739. 486
- Cova, F., E. Dupoux, and P. Jacob. 2010. Moral evaluation shapes linguistic reports of others' 487  
psychological states, not theory-of-mind judgments. *The Behavioral and Brain Sciences* 33: 488  
334–335. 489
- Cova, F., P. Jacob, and E. Dupoux. Submitted. The more you deserve blame, the more you deserve 490  
blame: The “valence matching” heuristic. 491
- Damasio, A. 1995. *Descartes' error*. New York: Putnam. 492
- De Waal, F. 1997. *Good natured: The origins of right and wrong in humans and other animals*. 493  
Cambridge, MA: Harvard University Press. 494
- Fischer, J.M., and M. Ravizza. 1998. *Responsibility and control: An essay on moral responsibility*. 495  
Cambridge, MA: Cambridge University Press. 496
- Frankfurt, H.G. 1969. Alternate possibilities and moral responsibility. *Journal of Philosophy* 497  
66(23): 829–839. 498
- Gollwitzer, M., and M. Denzler. 2009. What makes revenge sweet: Seeing the offender suffer or 499  
delivering a message? *Journal of Experimental Social Psychology* 45: 840–844. 500
- Kant, I. 2011. *Groundwork of the metaphysics of moral*. Cambridge, MA: Cambridge University 501  
Press. 502
- Knobe, J. 2006. The concept of intentional action: A case study in the uses of folk psychology. 503  
*Philosophical Studies* 130: 203–231. 504
- Levy, N. 2008. The responsibility of psychopaths revisited. *Philosophy, Psychiatry and Psychology* 505  
14(2): 129–138. 506
- Marit, R.S., and P.A. Wilkosz. 2005. Disorders of diminished motivation. *The Journal of Head* 507  
*Trauma Rehabilitation* 20(4): 377–388. 508
- Mikhail, J. 2011. *Elements of moral cognition: Rawls' linguistic analogy and the cognitive science* 509  
*of moral and legal judgment*. Cambridge, MA: Cambridge University Press. 510
- Pizarro, D., E. Uhlmann, and P. Salovey. 2003. Asymmetry in judgments of moral blame and 511  
praise. *Psychological Science* 14(3): 267–271. 512
- Singer, P. 1972. Famine, affluence and morality. *Philosophy and Public Affairs* 1(3): 229–243. 513
- Turiel, E. 2002. *The culture of morality: Social development, context and conflict*. Cambridge, MA: 514  
Cambridge University Press. 515

AQ2

AQ3



AUTHOR QUERIES

- AQ1. Please confirm the author affiliation.
- AQ2. Please provide in-text citation for the references Aristotle (2011), Kant (2011).
- AQ3. Please update the reference Cova et al. (submitted).

UNCORRECTED PROOF