Can Nonhuman Animals Be Moral Agents?

Virginie Simoneau-Gilbert

Pembroke College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Philosophy

Hilary Term 2025

Can Nonhuman Animals Be Moral Agents?

Virginie Simoneau-Gilbert
Pembroke College, University of Oxford
A thesis submitted for the degree of DPhil in Philosophy
Hilary Term 2025

Primary supervisor: Prof. Alison Hills
Secondary supervisor: Prof. Dominic Wilkinson
Former primary supervisor: Prof. Jeff McMahan

Thesis Abstract

This thesis tackles the following question: Can nonhuman animals (hereafter, animals) be moral agents? Chapter 1 offers a summary of the debate on animal morality and highlights how moral agency has been understood in regard to two types of moral capacity: *epistemic* and *self-control* capacities. *Contra* threshold views of moral agency, I argue that moral agency is best understood as a gradual and multi-faceted phenomenon and that it can be teased apart from the concept of moral responsibility. Chapter 2 highlights how even primary forms of empathy, like emotional contagion, are relevant to moral agency in an *epistemic* sense. In that chapter, I argue that emotional contagion, which many psychologists and philosophers consider the most basic type of empathy, enables animals and young children to have access to a morally relevant *evaluative* fact: the badness of others' suffering. Chapter 3 expands on the argument developed in Chapter 2 and argues that many animals possess a further epistemic capacity associated with moral agency. In that chapter, I stress how animals' capacity for emotional contagion and recognition of intentional action in others gives them access to an important *deontic* fact: the wrong-making features of intentionally causing suffering. Chapter 4 explores moral responsibility practices in animals and addresses animals' capacity for *self-control*. I posit a Strawsonian approach to moral responsibility and argue that animals' capacities (1) to recognise the wrong-making features of intentionally causing suffering and (2) to form interpersonal relationships with other animals (3) give rise to expectations about how they ought to be treated. These expectations find their expression in a specific emotion: anger. Finally, Chapter 5 briefly explores the practical implications of recognising animals as moral agents. I argue that we may be justified in holding some domesticated animals morally responsible for their actions. I also explore how recognising some animals as moral agents widens our understanding of how we can harm them, both subjectively and objectively.

Word count:
63,956 without bibliography
72,414 with bibliography

Acknowledgements

I feel immensely blessed and humbled to be able to express my gratitude to so many people for supporting me throughout my doctoral studies.

The first person I want to thank is the love of my life, *mon amoureux*, Maxime Laporte, for always believing in me. Every day of this intellectual adventure, he showed me, by his example, the meaning of the word *empathy*. Without his faith, completing this DPhil would have been much harder and longer, if not impossible. I dedicate my thesis to him, my favourite moral agent.

I also want to thank my mother, Chantal Simoneau, for her unconditional love and support and for accepting that her little baby had flown far away from the nest. I know it was not easy. Our weekly calls were deeply heartwarming. They reminded me of home.

I will be forever grateful to my supervisors for their kindness, patience, and demandingness. It was an honour to work with them. My first primary supervisor, Prof. Jeff McMahan, is one of my greatest influences. His brilliance, principledness, sheer love of animals, and unbreakable will to make a difference for those who suffer will inspire me for life. I wish to thank my second primary supervisor, Prof. Alison Hills, for her rigour, benevolence, attention to detail, constructiveness, and original thinking. As a woman in philosophy, I could have never asked for a better role model to look up to. My meetings with Prof. Dominic Wilkinson, my secondary supervisor, were always a great source of motivation and new ideas. I benefitted enormously from his philosophical sharpness, generosity, writing insights, and professional advice. I cannot thank him enough.

I am indebted to Prof. Will Kymlicka and Prof. Jonathan Birch, two wonderful mentors, for offering lengthy remarks on some of my chapters and discussing them with me. I can say with confidence that my thesis is much stronger thanks to their feedback. I look forward to working with Will for my postdoc next year and collaborating with Jonathan on future projects.

For their insightful questions and comments on my work, I owe gratitude to Dr. Joanna Demaree-Cotton, who was one of my Confirmation of Status assessors, the editing team who worked on *The Plant-Based and Vegan Handbook* (Springer, 2024), four anonymous reviewers at the *Journal of Ethics* and *Utilitas*, and the researchers who attended the following conferences: The *Études animales et engagement* Conference held in Strasbourg in 2024, the online talk organised in 2024 by the Observatoire de recherche sur la condition animale, the online talk organised in 2023 by the Society for the Study of Ethics and Animals, the Groupe de recherche en éthique environmentale et animale's early career scholars workshop held online in 2022, the *Estivales de la question animale* held in France in 2022, and the *Penser le spécisme aujourd'hui* Conference held at the École normale supérieure de Lyon in 2022.

I am also grateful to Prof. Marc Bekoff for answering my questions about social play and third-party intervention in animals. His genuine enthusiasm for animal cognition was contagious, and our thought-provoking conversation was one of the highlights of my DPhil.

I wish to thank my dear philosophy friends and mentors in Oxford and Montréal, without whom writing this thesis would have been a lonely enterprise: Jen Semler, Erin Blaire, Ronya Ramrath, Paige Massey, Akumjung Pongen, Lorenzo Elijah, Rhys Southan, Alice Winham, Prof. Christian Nadeau, Dr. Valéry Giroux, Prof. Christiane Gauthier, Prof. Bruno Lacroix, Christiane Bailey, and Dr. Martin Gibert. They always succeeded in reminding me why I had chosen to study philosophy all those years ago, and animal ethics in particular: because it is profoundly meaningful, it matters morally. I owe a great deal to Prof. Christian Nadeau and Dr. Valéry Giroux, my master's supervisors, for always inspiring me to become a better philosopher.

For their wonderful encouragement and cheerfulness, I wish to thank my amazing friends Luc Papineau, Alexia Renard, Élise Desaulniers, Rachel Lamoureux, Inès Benadda, Mathilde Lafortune, Prof. Alain Roy, Raphaëlle Élément, Félix Tremblay, Annie Veillette, Samuel Villeneuve, Clarisse Émond-Larochelle, Marcus Dahl, Diane Elijah, Chrystophe Letendre, Étienne-René Contant, Michaël Lessard, Marie-Andrée Plante, Victoria Emmanuelle Forest Briand, Cassidy Bereskin, and Anat Peled. I am also grateful to my father, Patrice Gilbert, my godfather, Jean-Denis Gagnon, my godmother, Brigitte Simoneau, and my grandmother, Lise Ouellet, for their support.

Finally, I want to thank everyone at the Rhodes Trust, especially Mary Eaton, for their invaluable generosity. Without the Trust's confidence in me, undertaking a DPhil at Oxford would have remained a distant dream.

Table of Contents

Table of Illustrations

À Maxime, mon agent moral préféré

Introduction

In 1964, Stanley Wechkin, Jules Masserman, and William Terris, from the Department of Neurology and Psychiatry at Northwestern University in Chicago, presented some rhesus monkeys with a very simple choice. The monkeys could operate a device that would give them food but cause painful electric shocks to another monkey, or they could refrain from operating the device. In that case, they would go hungry, but their conspecific would not suffer. Wechkin, Masserman, and Terris found out that most monkeys abstained from getting food. Two of them even refused to eat for five and twelve days.[1] These results were interpreted as providing evidence of altruism in a nonhuman species.

Over the last decades, ethologists and biologists have observed similar behaviour in numerous species of nonhuman animals, not just primates: dogs acting to free their trapped and distressed human caregivers,[2] rats choosing to first help a conspecific in need instead of eating chocolate chips,[3] and coyotes refraining from harshly biting their playmates to continue playing.[4]

At first glance, such behaviour seems morally significant. The reason for that intuition is simple: animals[5] who help others in need or adjust to their conspecifics' wants and needs seem to exhibit *concern* for the well-being of others, sometimes at the cost of their own well-being. Considerations over individuals' well-being are one of the central objects of morality, which covers values such as fairness, justice, well-being, autonomy, and rights, and the

[1] See Masserman, Wechkin & Terris (1964a and 1964b).
[2] Van Bourg, Patterson & Wynne (2020).
[3] Sato, Tan, Tate, & Okada, (2015).
[4] For one of the first overviews of social play in wolves, coyotes and dogs, see Bekoff (1974).
[5] I shall use "nonhuman animals" and "animals" interchangeably in my thesis.

obligations to which they give rise.[6] Moreover, animals who are attentive to the well-being of others could be motivated by empathy, which most psychologists and many philosophers now define as the capacity to feel and understand the emotions of others, whether it be through emotional contagion – the capacity to feel others' emotions –, or more sophisticated cognitive abilities like imagination and perspective-taking.[7] Basic forms of empathy like emotional contagion have been observed in all mammalian species and some avian species, such as pigeons, ravens, cockatiels, zebra finches, kea parrots and hens toward their chicks.[8]

Most philosophers agree that empathy can play a role in human beings' moral lives, although they do not agree on its importance. Kantians typically argue that empathy and other emotions or emotional capacities can suggest reasons for action but are not reasons in themselves.[9] Utilitarians, on their part, generally agree that empathy can lead us to maximise well-being and do what is morally right, although it often needs to be constrained by our capacity for moral reasoning. For instance, psychologist Paul Bloom and philosopher Jesse Prinz rightly highlight how empathy can be vulnerable to morally problematic biases.[10] These include "cuteness" effects, as human beings are more likely to be empathetic toward endearing animals, even though this characteristic is morally arbitrary.[11] Finally, some sentimentalists can go as far as to argue that empathy plays a more fundamental role. It could even underpin our capacity for moral judgment.[12]

Furthermore, recent empirical studies and anecdotes on animals' concern for the well-being of others put pressure on the largely held view in animal ethics that animals are moral

[6] For an overview of various definitions of morality, see Musschenga (2013), pp. 100-102.
[7] See Preston & de Waal (2002).
[8] For an extensive review of the empirical literature on emotional contagion in animals, see Pérez-Manrique & Gomila (2022).
[9] See, for instance, Korsgaard (2009), pp. 109-112.
[10] Bloom (2016) and Prinz (2011), pp. 225-227.
[11] Prinz (2011), p. 226.
[12] See Hume (1739), 469, for instance.

patients *only* – individuals who matter morally, toward whom moral agents have moral obligations, but who do not have obligations toward others. The distinction between moral patients and moral agents was put forward by philosopher Tom Regan, who provided in *The Case for Animal Rights* (1983) a highly influential framework that shaped the development of animal ethics. Regan sums up the differences between moral agents and individuals who are moral patients *only* as follows:

> In contrast to moral agents, moral patients lack the prerequisites that would enable them to control their own behavior in ways that would make them morally accountable for what they do. A moral patient lacks the ability to formulate, let alone bring to bear, moral principles in deliberating about which one among a number of possible acts it would be right or proper to perform. Moral patients, in a word, cannot do what is right, nor can they do what is wrong.[13]

Moral patients can act in a way that is detrimental to the well-being of others, Regan grants. But because they lack sophisticated capacities, such as the capacity to reflect on what ought to be done in light of impartial moral principles, individuals who are moral patients cannot do what is right or wrong.[14] Paradigmatic cases of moral patients include young children and nonhuman animals, among many.[15]

Three things can be noted here. First, Regan uses the term "moral patient" in a narrow way, to describe individuals who are moral patients *only*, but the categories of moral agents and moral patients are not mutually exclusive. In fact, most adult human beings are both moral agents (individuals who can fulfil moral obligations) and moral patients (individuals toward

[13] Regan (1983), p. 152.
[14] *Ibid.,* p. 153.
[15] *Ibid.*

whom moral agents have obligations). That said, recent philosophical developments on artificial moral agents have cast doubt on that view. Some artificial intelligence systems, largely taken not to be sentient, could be moral agents capable of moral actions but not moral patients.[16] We can think here of self-driving cars. Second, describing animals as *mere* moral patients comes with advantages, especially for animal rights theories. As Christiane Bailey points out, it allows philosophers to argue that animals do not need to have moral duties or possess high cognitive capacities to be rights-holders.[17] Third, the moral patiency paradigm may be the appropriate framework to describe the moral status of many animals. For example, reptiles are sentient animals who can feel pleasure and pain. Their well-being matters morally. However, they have not been found capable of empathy, and it is likely they cannot feel and understand others' suffering.[18]

In the last two decades, a new debate, known as the "animal morality debate," has developed. Several philosophers have challenged Regan's framework and have argued that the moral patiency paradigm fails to account for recent findings on the capacities of animals. These novel arguments raise the following pressing questions: If some animals are not moral patients, what are they? Could they be moral agents? My goal is to answer the latter question, which also gave this thesis its title. But before providing an overview of my main arguments and conclusions, we need to expand on Regan's remarks and shed light on moral agency – what it is, how it is related to moral responsibility, and which capacities are associated with it.

[16] See Floridi & Sanders (2017), especially pp. 317-320.
[17] Bailey (2014), p. 34.
[18] For a recent literature review on reptile sentience, see Lambert, Carder & D'Cruze (2019).

What Is Moral Agency?

According to philosopher Vinit Haksar, moral agents are individuals who are "expected to meet the demands of morality."[19] To put it more precisely, moral agents can be held morally responsible for their conduct. This means they can have moral duties and be praised or blamed for failing to fulfil their obligations. Moral agents are individuals who possess some capacities that enable moral actions for which responsibility responses are warranted.[20]

As noted by philosopher Timothy Nailer, we have very few theories of moral agency, but several of moral responsibility, and the literature on the latter can help us define the former.[21] According to Nailer, neurotypical adult human beings are paradigmatic moral agents, and they can be held morally responsible for their actions. Nailer sums up the link between moral agency and moral responsibility as follows: "Moral agents are those individuals with the ability to meet the relevant responsibility conditions."[22] Such responsibility conditions are best understood in terms of agential abilities.[23]

Philosophers disagree on the capacities that ground moral agency and, by extension, moral responsibility. These can be put into two broad categories: *epistemic* and *self-control* capacities, following philosopher Mark Rowlands' distinction between theories of moral responsibility grounded in moral understanding and those grounded in self-control.[24] *Epistemic capacities* refer to moral agents' capacity to have access to moral facts. Moral agents who possess such capacities can understand moral propositions, the reasons why some actions are right or wrong, and the moral principles that underlie them.[25] *Self-control capacities*, understood very

[19] Haksar (1998).
[20] *Ibid.*
[21] Nailer (2022), p. 7.
[22] *Ibid.*
[23] *Ibid.*
[24] Rowlands (2012), especially Chapters 4 and 7.
[25] *Ibid.,* p. 239. This corresponds to Rowlands' definition of moral understanding. I shall return to it in Chapter 1.

broadly, relate to agents' capacity to control their behaviour and do what is morally right. It concerns individuals' capacity to make moral decisions. Epistemic and self-control capacities could be more "intellectual" in nature, related to agents' capacity to understand reasons for action and control their behaviour in light of such reasons, or more emotional. For instance, it is plausible that moral agents' capacity for guilt helps them see that an action is wrong and refrain from doing it. Both types of capacities, intellectual and emotional, often interact with one another. As mentioned previously, more sophisticated forms of empathy are fostered by cognitive capacities like imagination and perspective-taking.

Intellectual and emotional capacities admit various levels of sophistication, and philosophers have different views on the degree of moral development that individuals must reach to be moral agents. According to more demanding views of moral agency, such as those put forward by Kantian philosophers like Tom Regan and, more recently, Christine Korsgaard, agents must meet very high standards of moral conduct to be described as moral agents.[26] They must be capable of formulating moral principles and reflecting on them. If they do not possess such capacities, they cannot act for moral reasons, and their actions cannot be moral.[27] I shall return to these accounts in greater detail and assess their robustness in Chapter 1. In contrast, other accounts of moral agency are more inclusive. Doing the right thing for the right reasons, such as helping someone in distress to alleviate that person's suffering, is praiseworthy. Agents do not need to scrutinise their reasons for action and show reflective

[26] Regan (1983), p. 152 and Korsgaard (2006).
[27] Regan (1983), p. 152.

self-control to be moral agents.[28] Of course, several theories of moral agency or moral responsibility fall between the two ends of that spectrum.[29]

Thesis Structure

As already noted, my thesis aims to answer the following question: Can animals be moral agents? This broad question raises, in turn, more specific ones to which my five thesis chapters are dedicated:

(1) How should moral agency be defined? *(Chapter 1)*

(2) Can empathy play an epistemic role? Can it allow animals to have access to the badness of others' suffering? *(Chapter 2)*

(3) Can empathy play a further epistemic role? Can it enable animals to have access to the wrong-making features of causing suffering? *(Chapter 3)*

(4) Can empathy also play a self-control role? Can empathetic animals hold each other morally responsible? *(Chapter 4)*

(5) What are the implications for animal ethics of recognising some animals as moral agents? *(Chapter 5)*

My answer to this puzzle is positive: some animals can be moral agents, although to a much lesser degree than most human moral agents. Without settling the debate between

[28] An influential account of that type is the one developed by Nomy Arpaly in *Unprincipled Virtue (*2002). That said, Arpaly denies that animals can be moral agents because they do not possess or track moral concepts such as *increasing utility* or *respecting persons*. They cannot respond to moral reasons, according to Arpaly. See Arpaly (2002), p. 146.

[29] An influential example of that type of theory is the account of moral responsibility developed by Fischer & Ravizza (1998). According to Fischer and Ravizza, agents must be capable of responding to moral reasons to be held morally responsible. This capacity is understood in regard to "guidance control", which is an individual's capacity to do otherwise for a moral reason.

theories of moral agency and moral responsibility, I argue that empathy can play both an epistemic and self-control role in allowing animals to *have access* to some moral facts and *adjust* to others' emotional reactions and the demands they express. This, in turn, has important implications for theories of moral responsibility, especially for those that take responsibility practices and emotions like moral anger as the starting point of any philosophical inquiry about the nature of moral responsibility. Paradigmatic examples of such accounts include the Strawsonian approaches.[30]

Chapter 1 provides an overview of the animal morality debate by summarising arguments developed both by philosophers and comparative psychologists on the moral capacities of animals and how moral agency should be defined. I assess the robustness of Christine Korsgaard's and Mark Rowlands' theories of moral agency, according to which animals cannot be moral agents because their capacities do not meet a particular line of sophistication. In that chapter and throughout this thesis, I refer to these accounts as *threshold views*, according to which agents must reach a very specific threshold of moral development to be moral agents. Drawing from an uncontroversial claim – that neurotypical adult human beings are moral agents – and the scientific literature on children's moral capacities, I pose a challenge to proponents of threshold views. I argue that threshold views find themselves in troubled waters when determining around what age human beings become moral agents, which renders them vulnerable to important epistemic and moral errors. I conclude that moral agency is best understood as a gradual and multi-faceted phenomenon, as this account can more easily answer the difficulties I identify for threshold theories of moral agency.

If moral agency should be defined as a phenomenon that develops over time and encompasses several moral capacities, this leaves the door open to recognising infants and

[30] See Strawson (1962). For a summary of this approach, see also Shoemaker (2015) and §4.1 of my thesis.

many animals as moral agents. Chapter 2 explores how empathy in young children and many nonhuman animals could be relevant to moral agency. I argue that even primary types of empathy, like emotional contagion, can play an *epistemic* role. More precisely, I argue that empathetic animals can have access to the badness of others' suffering if: (1) the intentional object of their empathy is another individual in a state of suffering, (2) they suffer in tune with the sufferer, and (3) the aversiveness of this experience allows them to have access to the badness of others' suffering. I contend that no other criterion, like altruistic motivation, is required for animals to access that moral fact.

Three things should be said here to clarify what I mean by "having access." First, the argument I develop in Chapter 2 should not be conflated with arguments about animals' possession of evaluative concepts such as "good" or "bad". I argue that animals do not need moral concepts to respond to moral facts and have access to them. Second, I do not claim that animals' access to the badness of others' suffering is akin to a form of moral *knowledge*. To support that claim, I would need to show how empathetic animals' beliefs about others' suffering are true and justified and can resist Gettier-type cases.[31] This strikes me as an overly ambitious task if not an impossible one. Third, I do not need to endorse a specific theory of the relationship between suffering's aversiveness and badness and, more broadly, between phenomenal and evaluative properties. My argument simply rests on the uncontroversial thesis that a tight relationship exists between suffering's aversiveness and badness. No more precise picture of the nature of moral facts is needed here.

Chapter 3 follows a similar line of argument and proposes an entirely novel argument on animals' access to wrongdoing. In that chapter, I build on the main conclusion of

[31] This is nowadays a relatively common definition of knowledge. For an overview of the philosophical literature on knowledge, see Ichikawa & Steup (2024).

Chapter 2, namely, that empathetic animals can have access to the badness of others' suffering, and I highlight how many animals can also recognise when someone intentionally causes suffering to another individual. These two capacities, I argue, enable them to have access to the wrong-making features of causing suffering: that another individual causes a bad state of affairs. Like the arguments I develop in my second chapter, the ideas I develop in Chapter 3 do not rely on a particular account of the relationship between properties – here, between evaluative and deontic properties. I rather presuppose that there is a close connection between the *badness* of suffering and the *wrongness* of causing it. That view is largely uncontroversial among philosophers, including non-consequentialists. I then close this chapter by responding to three possible objections: that animals cannot have access to the wrong-making features of causing suffering (1) because they do not explicitly possess the concept of what is *morally* wrong, as opposed to other types of wrong, (2) because the capacity to have access to the wrong-making features of intentionally causing suffering requires more than merely having an aversion for some actions and that (3) because it involves more sophisticated cognitive capacities according to some moral theories.

Chapter 4 examines the link between animal moral agency and moral responsibility. It addresses the possibility of recognising many animals as moral agents who engage in moral responsibility practices with each other. I argue that animals' capacities (1) to have access to the wrong-making features of causing suffering and (2) to form relationships with other animals (3) give rise to expectations about how they ought to be treated. These expectations are expressed through anger, which is typically aroused by the perception that an expectation was not met, and which is an emotion common to all social animals. To defend this thesis, I posit a Strawsonian approach to moral responsibility and expand on some of the key concepts put forward by Peter Strawson in *Freedom and Resentment* (1962). I then defend the three

proposed criteria by emphasising how empathy could play an important *self-control* role by enabling animals to adjust to others' anger and the expectations it expresses. I close my fourth chapter by defending my account against one possible objection: that animals' anger stems from the perception that a predictive expectation was violated but not a normative expectation that concerns how others ought to behave.

Chapter 5 is slightly different from the previous ones. In this last chapter, I do not aim to offer a sharp argument on animals' moral capacities. My goal is rather to outline the main practical implications of recognising some animals as moral agents and to briefly explore possible new arguments. These, however, deserve to be more developed in future research projects. In that chapter, I tentatively argue that recognising some animals as moral agents could justify our practice of holding domesticated animals morally responsible for some actions and could widen our understanding of how we can subjectively and objectively harm animals.

Theoretical Assumptions

Before turning to the arguments, I want to note the theoretical assumptions that underpin my work. These touch upon my meta-ethical framework and some debates in the science and philosophy of animal minds that I chose to put aside.

First, the arguments I develop in my thesis fit well with some meta-ethical theories but not with others. Indeed, the key thesis I defend in Chapters 2 and 3, according to which empathetic animals could have access to the badness of others' suffering and the wrong-making features of causing suffering, aligns with meta-ethical realism. According to proponents of such views, moral facts exist independently of what moral agents feel or decide,

though there is disagreement over the precise nature of these facts.[32] Empathy, then, could constitute one of the means by which human beings and other animals can have access to moral facts. My arguments could also align with simple subjectivism, according to which moral facts are a matter of moral agents' feelings of approval or disapproval. Although it is not defended in philosophy anymore, that meta-ethical framework has historically played an important role in fostering novel views on the moral capacities of animals, especially in David Hume's work.[33] I shall come back to it in Chapter 1.

The views I defend in my thesis are less straightforwardly compatible with other meta-ethical pictures, such as constructivism and norm-expressivism, according to which moral facts rely on human beings' capacity to rationally construct them or agree on moral norms.[34] Hence, a complete account of animal moral agency would require a more thorough defence of moral realism, which falls outside the scope of this thesis.

Second, several questions arise from any philosophical examination of animals' capacities and minds. These questions can be divided into two categories: methodological and philosophical. The former concerns the main methodological principles scientists must follow when studying animal minds and how we should interpret the empirical evidence we have on animal empathy. As noted by philosopher Kristin Andrews, there are three of them: (1) anti-anthropomorphism, which rejects the unjustified attribution of human capacities to animals; (2) Morgan's Canon, which states that animals' capacities should not be explained by appealing to higher psychological processes if they can be interpreted in terms of lower psychological processes, and (3) anti-anthropocentrism, according to which scientists ought to reject holding

[32] For a summary of these debates, see Sayre-McCord (2023).
[33] See Hume (1739), 469. For a summary of subjectivism and its criticisms, see Kauppinen (2022).
[34] For a summary of these meta-ethical theories, see van Roojen (2023) and Bagnoli (2024).

human minds as the highest standard for studying animal minds.[35] Over the last two decades, there has been important disagreement concerning the relevance of the first and second principles. For instance, ethologist Marc Bekoff has criticised the first principle by pointing out how using anthropomorphic language can allow scientists to better understand the lives, worlds, and capacities of animals without necessarily making science less rigorous.[36] Other comparative psychologists have criticised Morgan's Canon by putting emphasis on its lack of clarity over the definition of "higher" and "lower" psychological processes and on how it could potentially slow down scientific progress.[37]

The other family of challenges is related to the philosophy of animal minds. It concerns questions about the nature of animal consciousness and consciousness more broadly. As pointed out by Andrews, there are two types of theories of consciousness: representationalist and nonrepresentationalist theories. According to representationalism, conscious individuals have the capacity to represent certain things, such as beliefs in the case of first-order theories or metacognitive mental states in the case of higher-order theories.[38] According to nonrepresentationalist theories, consciousness can be explained in neurophysiological terms.[39] Similarly, there is a strong disagreement about how we can know that animals are conscious. According to *non-inferential* views, we can observe others' minds in their bodily movements and behavioural cues. Knowledge about other minds and others' mental states would thus be akin to direct perception.[40] On *inferential* views, more is needed to infer that animals are conscious,

[35] Andrews (2020a), p. 3.
[36] Bekoff (2000), p. 867.
[37] For a summary, see Andrews (2020a), pp. 11-13.
[38] Andrews (2014), p. 52.
[39] *Ibid.,* p. 53.
[40] *Ibid.,* p. 14.

such as evidence for flexible behaviour and learning.[41] We cannot directly perceive minds in action.

I shall not address these two types of challenges in my thesis. Three reasons support that choice. First, my background is that of a moral philosopher, not a comparative psychologist or philosopher of mind. I became interested in animal morality after writing a master's thesis on the philosophical underpinnings of legal personhood and the possibility of recognising nonhuman animals as legal persons. Second, insightful arguments and remarks on animal consciousness, animal concepts, and scientific scepticism regarding animal morality have already been offered elsewhere.[42] I shall leave them to the philosophers of mind. Third, the theses I put forward in my work are minimal. They do not presuppose any specific picture of animal consciousness. Instead, they rest on the simple claim that animals are sentient agents. As such, they have beliefs, desires and emotions with a content or intentional object. In other words, numerous animals are agents with an emotional and mental life, which can vary in degrees of complexity. This is all I need to examine animal empathy's relationship to moral agency.

Let us now dive into the debate.

[41] *Ibid.*, p. 56.
[42] In addition to Andrews (2014 and 2020a) and Bekoff (2000), see de Waal (2006a). On animal beliefs and concepts, see Allen (1998), Searle (1994), and Rowlands (2012), especially Chapter 2, which is a reply to Stich (1979).

Chapter 1

How Should Moral Agency Be Defined?

(§1.1 and §1.2 were published in a different version in *The Plant-Based and Vegan Handbook*, Springer. See Simoneau-Gilbert 2024).

As seen in my thesis introduction, many philosophers defend a theory of moral agency that is grounded in very specific moral capacities, such as the capacity to scrutinise reasons for action, that would exclude animals from the realm of agents who can be *moral* agents. Hence, the altruistic rhesus monkeys studied by Wechkin, Masserman, and Terris would not qualify as moral agents.

I think these views are too narrow. In this chapter, I defend a gradualist and multi-faceted understanding of moral agency and argue that such an understanding is the most robust. To do so, I first summarise the animal morality debate and provide an overview of the current competing views on how to interpret animals' behaviour after providing some preliminary remarks on my approach (§1.1.). Philosophers and scientists who have addressed the question of animal morality can be put into three categories: those who think that animals should be described as moral agents (§1.2.1), those who claim that animals are rather proto-moral agents (§1.2.2), and those who argue that animals cannot be moral agents, namely, Christine Korsgaard and Mark Rowlands (§1.2.3). Drawing from these different theories, I identify the main capacities that underlie moral agency and highlight how they admit different degrees of sophistication (§1.3). Finally, I argue that the two main philosophical arguments proposed against granting moral agency to animals in the animal morality debate face substantial challenges. First, Korsgaard grapples with making sense of children's moral capacities (§1.4.1). Rowlands, on his part, has proposed an argument that suffers from

substantial conceptual conflation and did not succeed in convincing me that the concept of moral agency should be closely linked to the one of moral responsibility (§1.4.2). Toward the end of this chapter (§1.4.2 and 1.5), we will see that a gradualist and multi-faceted view of moral agency can better face these challenges, in addition to putting pressure on the widely held view that moral agency and moral responsibility ought to be inseparable.

1.1. Some Preliminary Remarks

Before I examine the literature on animal moral agency in detail, two things should be said about the approach I will adopt throughout this chapter and thesis.

First, my goal is not to defend a specific definition of moral agency and thus try to settle the debate between philosophers on its very precise definition but rather to understand these theories as describing diverse phenomena that are part of the umbrella concept of moral agency. The reason for this choice is essentially methodological. It has to do with how some philosophers who believe that animals are moral agents defend a precise theory of moral agency and how they examine the empirical evidence to support their account. Philosophers could be right in taking an approach to moral agency that is theory-based, especially if specific accounts of moral agency have proven to offer some important advantages. Yet, in doing so, philosophers risk overlooking important empirical evidence that may fall outside their definition of moral agency and its moral implications for the way we ought to treat animals. For instance, philosophers who defend a specific conception of moral agency, like a Kantian one, tend to offer answers to the question of animal moral agency that save them from thoroughly examining the empirical literature on animals' capacities. Because they define moral agency as an "all-or-nothing" phenomenon that should be grounded above all in reflective self-control, their answer regarding animals' moral behaviour is simple: animals cannot be

moral agents. On the contrary, remaining agnostic on the priority that *should* be given to emotional or intellectual capacities may limit the occurrence of methodological biases related to how we interpret empirical evidence. Similarly, it does not lead us to overlook the ways in which we could harm animals that possess some capacities associated with moral agency. I shall expand on that claim in §1.4.2 and Chapter 5.

Second, philosopher Simon Fitzpatrick rightly highlights how authors who have addressed the question of animal morality have taken two entirely different approaches. One is "natural", and one is conceptual.[43] Scientists who have adopted the former have proposed an analysis of moral agency that is primarily descriptive and empirical. It is closely associated with views on the biological foundations of morality, the history of evolution, and the evolutionary function of morality. The other approach, more associated with philosophers like David Hume, Adam Smith, Christine Korsgaard, and Mark Rowlands, consists of answering the philosophical question of how we *should* define morality. Though some scientists and philosophers defend similar claims regarding animals' moral agency (§1.2.1) and proto-moral agency (§1.2.2), they defend their views on entirely different grounds. As Fitzpatrick notes, we should thus be careful about not conflating natural and conceptual analyses, and it is possible that in their attempts to define moral agency, the two groups "have largely been talking past each other" from different viewpoints.[44]

For instance, primatologist Frans de Waal offers evolutionary criticisms of Korsgaard's views. According to de Waal, it makes little sense to say that our capacity for reflective self-control "breaks with our animal past", as Korsgaard puts it.[45] Our moral maxims do not arise independently from our social organisation and cannot transcend the social instincts we have

[43] Fitzpatrick (2017), p. 1155.
[44] *Ibid.*
[45] Korsgaard (2006), p. 104.

in common with animals. De Waal writes that Korsgaard's theory is "like arriving at the top of a tower to declare that the rest of the building is irrelevant, that the precious concept of 'tower' ought to be reserved for its summit."[46] [47]

Fitzpatrick argues that in his response to Korsgaard, de Waal misses the target, since Korsgaard's project is not to define morality in biological terms.[48] Korsgaard can perfectly accept the Darwinian view that human morality is the product of animal social instincts and evolutionary forces while claiming that from a philosophical perspective, the word "moral agency" should solely be used for agents capable of reflective self-control.[49] There is no contradiction between the two assertions because Korsgaard's project is primarily a conceptual one. She does not aim to define morality from an evolutionary perspective, unlike ethologists and primatologists. People who work on animal moral agency with a natural approach thus work from a highly different perspective than the one adopted by philosophers.

This is not to say that empirical matters regarding the evolution of morality are not relevant *at all* for assessing the robustness of a philosophical theory of moral agency. I think they are. Yet de Waal fails to challenge Korsgaard's theory on philosophical grounds. Although

---

[46] de Waal (2006b), p. 181.

[47] Over the last decades, Kantian accounts of moral agency like Korsgaard's have faced other significant challenges that I shall not explore here. For instance, the link between scrutiny and self-control has received many criticisms, especially regarding akrasia cases. They must also face issues related to the infinite regress of second-order evaluations. On these two criticisms, see Mele (2001), Chapters 3 and 4, Tappolet (2017), and Rowlands (2012), pp. 174-178. Another important criticism has been raised by psychologists working on decision-making, such as Joshua Greene. According to these new developments, most of our moral decisions, especially those that are characterized by deontological judgements, could be largely automatic and driven by emotions or intuitions rather than reflective self-control. On that topic, see Greene (2008 and 2009) and Berker (2009) for criticisms of Greene. Finally, in the empirical literature on psychopathy, it is often suggested that secondary psychopaths show considerably underdeveloped capacities for self-control and often struggle with aggressivity and impulsivity issues, though they can have access to their reasons and scrutinize them. See Mealey (1995), cited in Aaltola (2014), p. 83. Some philosophers have taken these findings to indicate that the role of our capacity for reflective self-control and the link between scrutiny over our reasons for action and self-control could have been exaggerated.

[48] Fitzpatrick (2017), p. 1155.

[49] In fact, Korsgaard has argued that our capacity for reflective self-control may find its origins in evolution and, more precisely, in our mechanisms of dominance and social control. According to Korsgaard, we became rational when we started to exercise dominance on ourselves. On that point, see Korsgaard (2010), p. 21.

I agree with Fitzpatrick's analysis, I think empirical findings are helpful in assessing the robustness of moral agency theories, especially the empirical literature on the development of *human* moral agency. In this chapter, I intend to show how the empirical evidence on children's moral development places threshold views of moral agency like Korsgaard's in a difficult position. I will expand on this claim in §1.4. But before that, we need to examine the current debate on animal morality more closely.

1.2. The Debate on Animal Morality: An Overview

1.2.1. Animals as Moral Agents

Animals have been most often described as entirely incapable of moral behaviour. In the 1980s, several philosophers started to challenge these assumptions by adopting definitions of moral agency inspired by sentimentalism and virtue ethics. These new philosophical developments have also been supported by extensive scientific literature in ethology on empathy and cooperation in communities of apes, monkeys, cetaceans, canines, felines, rodents, and other species.[50] Animals' behaviour has attracted the attention of both philosophers and scientists.

The first family of views includes the work of authors who have claimed that animals can be moral agents. Within the contemporary debate on the definition of moral agency, Stephen Clark was the first philosopher to challenge the view that animals cannot be moral agents and has proposed a theory of animal moral agency inspired by virtue ethics. In *The Nature of the Beast* (1982), Clark argues that animals can be aware of some features of a situation and act in a way that reflects certain stable personality traits. In other words, they can respond

[50] For a summary, see Bekoff & Pierce (2009), pp. 24-54.

to the good- and bad-making components of a situation in the same way as a virtuous human being.[51] This capacity makes them moral agents, though to a lesser degree than neurotypical adult human beings. Clark thus departs from Aristotle's virtue ethics and puts forward a very minimal and non-intellectual criterion for virtuous action. Unlike virtuous agents described in *Nicomachean Ethics*, animals do not know they are performing virtuous actions and cannot decide to perform them for themselves, though they may satisfy Aristotle's criterion of acting from stable character traits.[52] For example, a "good" dog shows consistent concern for the well-being of others. Yet Clark argues that animals are not motivated to do the right thing as such and do not possess moral systems. Their behaviour may be best described as "ethical" rather than "moral."[53]

In the same vein, philosopher Steve Sapontzis acknowledges the fact that animals can act for the right reasons, in the sense that an animal agent can "recognize […] the moral value of the action that moves him to act"[54], even if he cannot do moral theory, grasp moral principles, or provide reasons for his actions. Moreover, animals exhibit flexibility and intentionality in their actions, and their behaviour can reflect some of their character traits like courage and empathy.[55] Even if we condition animals to act in a certain way, their actions possess moral value, for animals can still be responsive to the good- and bad-making features of a situation. As Sapontzis remarks in *Morals, Reason, and Animals* (1987), even animals who exhibit maternal instincts can have flexible "reactions to the needs of the baby"[56] that are not

---

[51] Clark (1984), p. 107.
[52] In *Nicomachean Ethics*, Aristotle claims that an action expresses a virtue when (1) the agent knows that she is performing a virtuous action, (2) chooses that action because it is virtuous and (3) that action expresses a stable trait in the agent. See Aristotle, *Nicomachean Ethics*, 11105a27-35. For a summary and comments, see Rowlands (2013), p. 17, and Dixon (2008a), pp. 71-75.
[53] Clark (1984), p. 107.
[54] Sapontzis (1987), p. 32.
[55] *Ibid.,* pp. 32-34.
[56] *Ibid.,* p. 34.

devoid of a moral dimension. Because animals can show "intentional, straightforward acts of kindness, courage, and the like"[57], they are capable of virtuous action.

In their book *Wild Justice* (2009), ethologist Mark Bekoff and philosopher Jessica Pierce have taken an entirely different approach to the animal morality debate by arguing that morality is "context-specific" and "species-specific".[58] According to them, morality should be defined as "a suite of other-regarding behaviors that cultivate and regulate complex interactions within social groups."[59] Morality's role is to facilitate cooperation among the members of a given community. In that sense, morality is species-specific because species members do have the same ways of regulating social interactions.

That said, Bekoff and Pierce warn their readers against the potential mistake of seeing their theory of animal moral agency as an endorsement of moral relativism, for animal communities show behavioural similarities and share common attitudes like empathy, altruism, cooperation, and perhaps a sense of fairness.[60] On the contrary, Bekoff and Pierce identify three moral "clusters" that are shared by most social mammals: (1) cooperation, which includes attitudes of altruism, reciprocity, honesty, and trust,[61] (2) empathy, in which we can find various emotions like sympathy, grief, and consolation,[62] and (3) justice, which they define very broadly as "a set of expectations about what one deserves and how one ought to be treated"[63] and which includes capacities for sharing, equity, fair play, and forgiveness.

More recently, several philosophers have claimed that animals can be moral agents and have argued that animals can manifest virtues in their actions and can act in a way that is not

---

[57] *Ibid.,* p. 147.
[58] Bekoff & Pierce (2009), p. 144.
[59] *Ibid.*, p. 7.
[60] *Ibid.,* pp. 147-149.
[61] *Ibid.,* p. XIV.
[62] *Ibid.*
[63] *Ibid.,* p. 113.

merely instinctive or conditioned[64] (David DeGrazia) or by arguing that some species, like great apes, can understand norms even if they do not possess a full-fledged theory of mind, which is the capacity to ascribe mental states to others that are different from our own, to understand that others can have false beliefs, to understand how others' emotions can be different from ours, and to understand how they can be caused by states of affairs (Kristin Andrews).[65] Authors like Lori Gruen and Kristin Andrews have also stressed the importance of empathy for moral conduct and have argued that great apes are capable of more cognitively complex forms of empathy.[66] Evelyn Pluhar, on her part, has defended a more limited attribution of moral agency to several species of animals. According to her, animals possess capacities necessary for full-fledged moral agency, like emotions, memory, and purposive behaviour.[67]

Some authors have also examined the possibility of recognising animals as moral agents who are morally responsible to some degree. As mentioned earlier, Bekoff and Pierce have argued that morality is "context-specific" and "species-specific." This entails that there is wolf morality, chimpanzee morality, lion morality, etc., and that animals who are moral agents could be morally responsible, but only in an intra-species context.[68] More recently, philosopher Dorna Behdadi has stressed the importance of animals' social contexts and interactions and has argued that animals engage in "moral responsibility practices" in which they hold each other morally responsible. Drawing from evidence on canines' social play and moral capacities, Behdadi has argued that these animals are moral agents who engage in moral

[64] DeGrazia (1996), p. 63.
[65] See Andrews (2009).
[66] Andrews & Gruen (2014), pp. 203-205.
[67] Pluhar (1995), p. 55.
[68] Bekoff & Pierce (2009), p. 144.

responsibility practices, such as praise and blame, both with each other and human beings.[69] In the same vein, philosopher Asia Ferrin has argued that animals can be morally responsible because they can manifest good or ill will, here defined as the capacity to act for moral reasons and that empathy could enable animals to care about the well-being of others.[70] This means that empathetic animals could be morally responsible to other members of their species but that human beings should refrain from holding animals morally responsible because of the "lack of overlapping social context" across species.[71] Philosopher Paul Shapiro has also argued that animals can be morally responsible, but only for actions they can morally understand to some degree. For instance, animals who are empathetic, who can care for other individuals and who can understand that others suffer but choose to harm others without a good reason, such as survival, can be morally responsible and blameworthy.[72] A dog who attacks her family members exhibits morally problematic behaviour because she is a moral agent to some degree and can even have obligations toward her family members. A dog may have the negative obligation not to attack them or even the minimal positive obligation to protect them against intruders.[73]

All these authors share two key theses, although they do not agree on the relationship between moral agency and moral responsibility and whether animal moral agents are morally responsible. These two claims can be summed up as follows: (1) moral agency comes in different degrees and forms, and (2) some animals can be described as moral agents, though they exhibit a kind of moral agency that is not identical or as sophisticated as that of neurotypical adult human beings.

---

[69] Behdadi (2020 and 2024).
[70] Ferrin (2019), pp. 137-138 and pp. 143-144.
[71] *Ibid.,* p. 146.
[72] Shapiro (2006), p. 365.
[73] *Ibid.,* p. 369.

1.2.2. Animals as Proto-Moral Agents

Other scientists and philosophers have taken a different approach, which consists of claiming that animals possess several emotional and social capacities related to moral agency, that there exists a continuity between animals' moral behaviour and full-fledged moral agency, but that animals cannot be considered moral agents because they lack more sophisticated cognitive capacities essential to moral agency. Animals are instead "proto-moral agents",[74] to quote primatologist Frans de Waal, though this concept is not explicitly used by the other authors that I include within this chapter section.

Today, two philosophers are cited as major sources of inspiration for authors who work on animals' moral behaviour. The first one is David Hume, according to whom sympathy, which he defines as our ability to render others' sentiments "present to us"[75], is central to moral agency. Indeed, Hume argues that our capacity for sympathy allows us to form judgements of approbation or disapprobation of others' actions.[76] Furthermore, he recognises that several animals are also capable of empathy and possess numerous "natural abilities" like love and friendliness.[77] However, animals lack reason and, thus, cannot perceive moral obligations or acquire a point of view that departs from their immediate situation.[78] Hence, animals possess some capacities that are related to morality, but some passages in Hume's work suggest that morality is exclusively human.[79] This explains why David Hume may be best classified within the group of authors who recognise a form of continuity between

[74] de Waal (2006a), pp. 54-55.
[75] Hume (1739), section 2.1.11.
[76] *Ibid.,* sections and 2.2.2 and 3.2.2.
[77] Hume (1739), section 2.2.12.
[78] Hume, David (1748), 9.5 to 9.7. See also Beauchamp (1999), p. 328.
[79] *Ibid.*

the moral capacities of human beings and other animals but without describing animals as moral.

The second one is Adam Smith, who, like Hume, stresses the central role of moral emotions in our capacity to respond to moral judgements of approval and disapproval. Smith also recognises that sympathy, which he associates with the cognitive capacity of placing oneself in another's position or "changing places in fancy with the sufferer"[80], can be limited in scope if it arises only in immediate situations. According to Smith, emotions qualify as moral when they reach a more abstract level and when they lead us to make judgements of approval or disapproval about how *anyone* should be treated.[81] Emotions become moral when agents can take the point of view of an impartial spectator. As philosopher Philip Kitcher explains, Smith defines sympathy as involving "reflecting upon – mirroring – the judgements of those with many perspectives around us, until we can combine each point of view, with its peculiar biases, into an assessment that expresses a genuinely moral sentiment."[82] Without such impartiality, animals can feel a basic type of sympathy, but the latter cannot be described as moral, although it may be necessary for the development of the more abstract and impartial type of sympathy we find in human beings.

Scientists have also highlighted how animals and human beings share several similarities in their moral capacities. According to Charles Darwin, human morality is deeply rooted in our social sentiments or, as he calls them, our "social instincts," which are common to all social animals. The function of these social sentiments may be to promote solidarity between the members of the same group and thus increase the group's overall fitness.[83]

[80] Smith (1759), section 1, chapter 1.
[81] *Ibid.,* section 1, chapter 5.
[82] Kitcher (2006), p. 132.
[83] Darwin (1871), pp. 149-150.

However, though social instincts may lie at the heart of morality, they are insufficient to make animals moral agents. That is because full moral agency requires two additional capacities: (1) the capacity to compare our past and future actions and motivations and (2) the capacity to approve or disapprove them.[84] According to Darwin, the emergence of these capacities is closely tied to the development of higher rational capacities, which animals do not possess. Nevertheless, there exists an evolutionary continuity between human and nonhuman animals in regard to their moral capacities. The recognition of this continuity led Darwin to famously write in *The Descent of Man* (1871) that "[a]ny animal whatever, endowed with well-marked social instincts, the parental and filial affections being here included, would inevitably acquire a moral sense or conscience, as soon as its mental powers had become as well, or nearly as well developed, as in man."[85]

Primatologist Frans de Waal also defends a Darwinian account of animal morality. De Waal describes morality as a "tower" that rests on three key building blocks: (1) moral sentiments, which include empathy, reciprocity, retribution, conflict resolution, and a sense of fairness; (2) social pressure, which he defines as a pressure "to contribute to common goals and uphold agreed-upon social rules"[86] and which encompasses attitudes of conformism, community concern, and conflict resolution, and (3) judgement and reasoning.[87] According to de Waal, the third building block may be uniquely human, but moral sentiments and social pressure are "evolutionary ancient"[88] and may be found in various species of social animals. However, humans have developed a more sophisticated sense of morality with the development of language and the influence of warfare, which induced greater solidarity among

[84] *Ibid.,* pp. 170-171 and 933.
[85] *Ibid.,* pp. 149-150.
[86] de Waal (2006b), p. 169.
[87] *Ibid.*, p. 165.
[88] de Waal (2006a), p. 7.

members of the same group.[89] Furthermore, human beings explicitly teach the importance of respecting moral standards and favouring the community's interests over selfish desires.[90] Human beings have moral systems, unlike animals.

Finally, Peter Singer also shares de Waal's views on the nonhuman roots of morality but stresses the importance of practical reason and universalisation for moral agency. In *Primates and Philosophers* (2006), Singer goes so far as to confess his "reluctant respect"[91] for Kant. Like him, Singer thinks that morality cannot be based on emotions alone. What distinguishes us as moral agents from other animals is our capacity to reason and abstract from our emotional responses. Unlike animals, we can reach a more impartial perspective that applies to members of our group *and* other group members. We can see that others "have interests similar to our own"[92] and that "there is no impartial reason why their interests should not count as much as the interests of our own group, or indeed as much as our own interests."[93] In that sense, human morality is unique in its scope, which is "new"[94] in evolution and "contrary"[95] to the emotional capacities we share with other social animals. Most animals, including human beings, can feel empathy. However, only human beings can abstract from their emotions, decide to follow or reject them and expand their concern to members of other groups.

To support this conclusion, Singer refers to Joshua Greene's and Jonathan Haidt's work on moral judgement.[96] The two psychologists have used neuroimaging to study what happens in the emotional and rational parts of the brain when some subjects are responding

[89] *Ibid.,* p. 55.
[90] *Ibid.,* p. 54.
[91] Singer (2006), p. 150.
[92] *Ibid.*, p. 145.
[93] *Ibid.*
[94] *Ibid.*
[95] *Ibid.*
[96] See Greene and Haidt (2002).

to different versions of the trolley problem. In one of these situation settings, a trolley will kill five people if nothing is done, but the participant can push a fat person standing on a footbridge above the track, and that person's body will stop the trolley. Greene and Haidt discovered that people who chose not to push the stranger showed activity in evolutionary older brain parts associated with social mammals' emotions. In contrast, people who decided to kill the fat person exhibited activity predominantly in brain parts associated with intellectual capacities.[97] For Singer, this proves that people who are ready to perform the morally right action (at least, from Singer's utilitarian perspective) can overcome their emotional responses to save as many lives as possible. Other participants had automatic emotional responses common to many social mammals but did not assess them.[98] According to Singer, it is reasoning that makes human beings moral agents.[99]

1.2.3. Moral Agency as a Uniquely Human Phenomenon

Finally, some philosophers have categorically denied that animals can be moral agents or even proto-moral agents. One of the most influential advocates of this option is Christine Korsgaard, for whom morality is not merely a matter of emotions. It rather rests on normative reflective self-control. According to Korsgaard, animals can be guided by perception[100] and are capable of intentional action[101] but cannot be guided by rational principles. Human beings are the only animals who can formulate maxims and universalise them, and the Kantian universalizability criterion enables us to adopt or reject actions and purposes. In other words,

---

[97] Singer (2006), pp. 148-149.
[98] *Ibid.*, p. 149.
[99] It is worth noting that these experiments and the normative conclusions that Greene and Haidt have derived from them have raised substantial criticisms in the last decades. On that point, see Berker (2009).
[100] Korsgaard (2006), p. 108.
[101] *Ibid.,* p. 110.

agents can formulate maxims. These can become universal laws when they do not admit contradictions in their conception and the agents' will.

Hence, what distinguishes human agents from other agents is that, although their goals can arise from their emotions,[102] humans can rationally assess them as universalisable and see them as grounds for their actions. Korsgaard writes:

> [A] nonhuman agent may be conscious of the object of his fear or desire, and conscious of it as *fearful* or *desirable*, and so as something to be avoided or to be sought. That is the ground of his action. But a rational animal is, in addition, conscious *that* she fears or desires the object, and that she is inclined to act in a certain way as a result. That's what I mean by being conscious of the ground as a ground.[103]

The difference between human and nonhuman animals is not merely a matter of degree but rather all or nothing. According to Korsgaard, our rational capacities "break with our animal past"[104] and make us truly moral, while animals' empathetic behaviour *cannot* be described as moral, nor can it be seen as participating in morality in some way. Animals do not possess one iota of morality.

It is worth noting that Korsgaard's theses on animal moral agency come with *internalist* and *constructivist* views on the nature of normativity and moral facts.[105] According to Mark Rowlands, Kantian philosophers like Korsgaard think that moral agency takes what he calls "the ASCNM route" (Access-Scrutiny-Control-Normativity-Moral).[106] In other words, an agent's access (A) to her purposes and desires enables her to scrutinise (S) them, and this

[102] *Ibid.,* p. 112.
[103] *Ibid.,* p. 113.
[104] *Ibid.,* p. 104.
[105] Fitzpatrick (2017), p. 1170.
[106] Rowlands (2012), p. 152.

capacity is a necessary condition for the agent to have control (C) over her inclinations. This capacity for reflective self-control, or, as Korsgaard calls it, normative self-government, then provides a normative (N) and moral (M) status to these desires, which exercise power on the agent. Korsgaard's meta-ethical theory is thus *internalist* in the sense that the rightness or wrongness of an agent's actions is determined by her psychological processes.[107] Similarly, her theses on animals' lack of moral agency are heavily loaded with *constructivist* views on moral facts, according to which moral facts are not independent of agents' practical deliberation process.[108] They are rather constructed by moral agents. Hence, human beings' capacity for reflective self-control plays an important role in underpinning both Korsgaard's accounts of moral agency and moral facts. It further sheds light on the reasons why animals cannot be moral agents. Their actions cannot be described as morally right or wrong.

Mark Rowlands also denies that animals can be moral agents but defends this thesis on grounds entirely different from Korsgaard's. According to Rowlands, animals cannot be moral agents because the concept of moral agency has been too closely tied to the notions of moral responsibility, praiseworthiness, and blameworthiness.[109] A moral agent is an individual who can be held responsible and praised or blamed for her actions. More precisely, Rowlands argues that moral responsibility and, by extension, moral agency are not grounded in the ASCNM route, for the route can likely break down at the Scrutiny-Control step.[110] Scrutiny can guide human beings' actions and lead to self-control, but not always. Rowlands contends that we do not need to be capable of reflectively controlling ourselves to be held morally responsible.

---

[107] Fitzpatrick (2017), p. 1170.
[108] *Ibid.*
[109] Rowlands (2012), p. 175.
[110] *Ibid.* p. 208.

Rather, moral responsibility and moral agency can be best understood in relation to moral understanding. Moral agents can understand moral facts and why they matter, and this understanding can come in various degrees. Rowlands takes moral understanding to rest on four capacities: (1) to make qualitative distinctions between positive and negative states like others' happiness and pain, (2) to grasp moral facts, (3) to understand why an action is right or wrong, and (4) to understand the moral principles that underwrite moral facts.[111]

Rowlands does not deny that animals possess some moral capacities. Yet he argues that animals' capacities are best captured by the concept of "moral subjecthood" rather than "moral agency". Rowlands defines the notion of a "moral subject" as follows:

> X is a moral subject if X possesses (1) a sensitivity to the good- or bad-making features of situations, where (2) this sensitivity can be normatively assessed, and (3) is grounded in the operations of a reliable mechanism (moral module).[112]

We need to further clarify what Rowlands' means by sensitivity, normatively assessed, and moral module. First, Rowlands claims that animals can show a form of sensitivity to moral features that is emotional and experiential. Animals can have emotions, which can be qualified as moral when (1) they are intentional, in the sense that they have an object, (2) there exists a proposition p that expresses a moral proposition like "Suffering is bad", and (3) an agent can track the proposition p.[113] Indeed, Rowlands claims that a proposition p* can track another proposition p if and only if the truth of p guarantees the truth of p* because there is a reliable, asymmetrical connection between the concepts in the two propositions.[114] Taking as an

[111] *Ibid.,* p. 239.
[112] *Ibid.,* p. 230.
[113] *Ibid.,* p. 69.
[114] *Ibid.,* pp. 58-63.

example the beliefs of his dog Hugo, who likes to chase squirrels, Rowlands argues that the proposition that "there is a squirrel in the tree" (proposition p) guarantees the truth of "the chaseable thing is up there" (proposition p*), which Hugo is capable of forming. Hugo's proposition p* thus tracks proposition p.[115]

We can further illustrate these three conditions with a familiar example of dog empathy. Suppose a young boy is drowning in a lake, and his dog jumps into the water to help him. In that situation, the dog's empathy (1) takes the child's distress as content, (2) there exists the proposition that the child's suffering is bad, (3) and the dog's empathetic concern tracks that proposition. The truth of "The child's suffering is bad" guarantees that the dog's empathy is not misguided.

Second, Rowlands' remarks on animals' sensitivity to the good- and bad-making features of situations fit well with *externalist*, *consequentialist*,[116] and *realist* or *quasi-realist* accounts of moral facts.[117] According to Rowlands, there exist features of situations that make them good or bad, and these features are objective moral facts that are external to an agent's psychological processes. Thus, an animal's action is morally right if the animal's sensitivity detects the good- and bad-making features of a situation and if the action enhances the good-making features.

Third, animals' moral sensitivity should be grounded in the operations of a reliable mechanism, which Rowlands calls a "moral module." The idea here is quite simple: a dog's sensitivity should not be contingent or accidental but should rest on a mechanism that

[115] *Ibid.,* pp. 53-62.
[116] *Ibid.,* pp. 222-223.
[117] Susana Monsó has argued that Rowlands' theory would also be compatible with Blackburn's quasi-realism. Blackburn admits that we can talk of moral claims as being true or false but raises doubts about the ontological status of moral facts. On that point, see Monsó (2015), p. 676.

guarantees that this sensitivity is always aroused when the dog is put in similar circumstances.[118] Such a mechanism could be akin to a stable disposition. Though Rowlands does not provide an exact definition or example of such a reliable mechanism, philosopher Susana Monsó, who follows Rowlands in describing animals as moral subjects who can feel what she calls "minimal moral empathy", has recently suggested that the perception-action model (PAM) proposed by Frans de Waal and Stephanie Preston could play the role of a moral module in human beings and other animals.[119] The PAM is a biological mechanism that always ensures that social mammals detect motor movements and emotions in other individuals.[120]

Shortly after the publication of his book *Can Animals Be Moral?* (2012), Rowlands provided a second argument in favour of animal moral subjecthood, which partly rests on the ones he developed in his book. It is useful to reproduce that second argument at length here. Rowlands writes:

> I think it is a mistake to characterize animals as moral agents. The concept of agency is inseparable from the concept of responsibility and hence from the concepts of praise and blame. If animals are moral agents, they are responsible for what they do and so can be praised or blamed for this. At one time, courts of law—both nonsecular and secular—set up to try (and subsequently execute) animals for perceived crimes were not uncommon (see, e.g., Evans, 1906; Dinzelbacher, 2002). I assume few would wish for a return to this practice. At the core of this unwillingness is the thought that animals are not responsible—and so cannot be blamed—for what they do. If this is correct, then their characterization in terms of moral agency should be resisted.
>
> It is not helpful, in this context, to talk of degrees of responsibility. Suppose a pig—let us call her "Babe"—has been tried and is to be executed for the heinous crime of stealing the neighboring farmer's turnips. Babe, the court has decided, is a moral

[118] Rowlands (2012), pp. 145-146.
[119] See Preston & de Waal (2002).
[120] Monsó (2015), pp. 683-684.

> agent and so is responsible for what she does. This does seem silly. But the attempt to save the idea of moral agency by invoking the idea of degrees of responsibility scarcely redeems it. To claim that Babe is less responsible for what she does—less than, say, an average adult human—seems to imply that the punishment should be mitigated. Instead of death, perhaps a good flogging would be appropriate. This punishment is as silly as the original. Ultimately, we might retreat to the idea that Babe is responsible for stealing the turnips but should not be punished. But given this retreat, what is left of the concept of responsibility? I suspect very little. I do not want to rule out the possibility of moral agency, and hence responsibility, in animals *tout court*. However, I suspect that if it exists at all in animals, moral agency is restricted to a number of small, highly idiosyncratic cases. In the vast majority of cases, animals cannot plausibly be regarded as moral agents.[121]

Rowlands' more recent argument thus relies on a series of new claims: (1) moral responsibility is inseparable from the concepts of praise and blame, (2) moral responsibility seems to be closely linked with punishment, and (3) we cannot invoke the idea of degrees of moral responsibility, for it entails that punishment should be mitigated. In the extract cited above, Rowlands also seems to conflate the notions of moral and legal responsibility by referring to animal trials as a consequence of attributing moral agency and moral responsibility to animals. I shall come back to this problem in §1.4.2.

To sum up Rowlands' views, most social mammals are moral subjects who exhibit a moral sensitivity that can be normatively assessed and grounded in a moral module, but they lack sufficient intellectual capacities to deeply understand moral facts. As a result, they cannot be held morally responsible for their actions and cannot be described as moral agents.

Hence, we can take both Korsgaard and Rowlands to agree on the following general thesis: animals lack the necessary intellectual capacities to be considered moral agents:

[121] Rowlands (2013), pp. 18-19.

reflective self-control in the case of Korsgaard and moral understanding in the case of Rowlands. Animals can rather be described as moral patients, according to Korgsaard, or as moral subjects, according to Rowlands.

1.3. Capacities Underlying Moral Agency

Philosophers who have contributed to the animal morality debate have identified several capacities they take to be essential or even necessary components of moral agency. These capacities can be put into two simple categories: emotional capacities, with a clear focus on empathy, and intellectual capacities, with an emphasis on reflective self-control and moral understanding. With the summary of the animal morality debate in mind, we can now fully see how these capacities can play the *epistemic* and *self-control* roles that have been associated with moral agency.

Emotional capacities are one of the key capacities that have been associated with moral agency, especially in sentimentalist theories and in the work of more recent authors who are sympathetic to recognising animals as moral agents, proto-moral agents, or moral subjects. Indeed, some emotional capacities like empathy could play the epistemic and self-control roles I put forward in my thesis introduction. Empathy could be *epistemically* important because it could be essential to human beings' capacity for moral judgement (Hume and Smith), it could enable animals to track moral propositions like "Suffering is bad" (Rowlands), and it could suggest reasons for action without being a reason for action in itself (Korsgaard). Emotional capacities like empathy could also have a *self-control* role to play if we define the notion of self-control in the weaker, non-reflective sense. It could allow agents to act in a certain way, for instance, to help others in need, refrain from doing what is wrong and adjust to others'

emotional states.[122] According to some philosophers, actions driven by empathy could even reflect more stable character traits and be a sign of a virtuous character (Clark and Sapontzis).

Intellectual capacities, on their part, play an important *epistemic* role in the moral lives of human moral agents. Indeed, they allow human beings to understand moral propositions and the principles that underwrite them (Rowlands), like the principles of utility and impartiality (Singer) or universality (Korsgaard). Such capacities also depend on other intellectual ones, like the capacity to use propositional language. Furthermore, intellectual capacities can enable moral agents to exercise *control* over their own actions in a much stronger sense. As noted by Singer and Korsgaard, moral agents' capacity to access and scrutinise their reasons for action allows them to choose certain motives as grounds for their action and act in accordance with moral principles, whether these be utilitarian or deontological.

However, emotional and intellectual capacities do not constitute an all-or-nothing phenomenon, and they are likely to interact with each other. The literature on child moral development may prove to be helpful in identifying the different levels of sophistication that these capacities admit, to which I now wish to turn. I aim to flesh out the main developmental stages these capacities admit, which will be helpful in assessing the robustness of Korsgaard's and Rowlands' accounts.

[122] In fact, the link between empathy and moral motivation has been well-studied. Several studies in psychology, mostly on children's moral development, establish either a low, moderate, or strong positive link between empathy and helping behaviour. The correlation's strength varies according to various factors such as experimental settings, methods for assessing empathy, the precise definition of empathy with which scientists work, and the age of children. For an extensive literature review, see Eisenberg & Miller (1987). See also Tangney, Stueig & Mashek (2007), p. 363 and Batson (2010).

1.3.1. Emotional Capacities: The Case of Empathy

As noted earlier, it is widely agreed in child psychology and ethology that empathy comes in various types and degrees of sophistication. Several influential theories have been proposed to account for the development of empathy in children and nonhuman animals.

One of them was put forward by child psychologist Martin Hoffman, who has identified five stages of sophistication in children's empathy: (1) reactive newborn cry, in which very young toddlers react to others' distress by crying; (2) egocentric empathetic distress, in which children react to others' suffering as if they were themselves in pain, and do not sharply distinguish others' emotions from their own, (3) quasi-egocentric empathetic distress, in which children grasp the fact that others' distress is not their own, but still lack a robust distinction between themselves and others, and thus help others by providing a form of comfort that they personally find reassuring, (4) veridical empathetic distress, in which children can feel a similar emotion to the others' feelings but without confusing the distinct sources of these emotions, and (5) empathy for others that extend beyond immediate situations.[123]

Hoffman's key thesis is that for empathy to become more extensive, children must possess a sharp self-other distinction, both in the sense that they must be able to see themselves as physically distinct from others and to take their own self as an object (stage 4).[124] They also have to acquire role-taking capacities (stage 5).[125] According to Hoffman, a more robust self-other distinction appears between 18 and 24 months when children can recognise themselves in a mirror. The mirror test shows that they possess a certain self-knowledge and that they can take their own self as an object of attention.[126] I shall come back to these claims

[123] Hoffman (2000), p. 6.
[124] *Ibid.,* pp. 7-8.
[125] *Ibid.,* p. 7.
[126] *Ibid.*, p. 70.

in Chapter 2, as recent findings on newborn infants' capacities put pressure on the view that emotional contagion relies on a lack of self-other distinction.

Another influential theory of empathy is de Waal's "Russian doll" model, according to which there exist three building blocks or levels of empathy in human and nonhuman animals: (1) emotional contagion, (2) sympathetic concern, which involves consolation and help, and (3) cognitive empathy. The last stage requires perspective-taking and may be common only to human beings and highly intelligent animals like great apes, dolphins, and elephants.[127] However, very young children and many animals possess capacities associated with the two first levels. I shall expand on these distinctions, the various degrees of sophistication empathy admits, and their relationship with moral agency in the next chapter. For now, it is sufficient to stress that empathy develops gradually, just like the other capacities associated with moral agency.

1.3.2. Intellectual Capacities: The Case of Moral Understanding and Reflective Self-Control

As we have seen, philosophers who deny that animals can be moral agents do so by appealing to the fact that animals lack moral understanding (Rowlands) or reflective self-control (Korsgaard). These capacities have been thoroughly investigated by psychologists working on child moral development.

I divide this section into two sub-sections. The first explores the way moral norms can be integrated into an agent's psychology and guide their actions and how reflective self-control relies on prior development stages. The second sub-section is dedicated to the way agents can grasp the different types and contents of norms, which is required for full-fledged *moral*

[127] de Waal (2008), pp. 287-288.

understanding. While philosophers often refer to moral facts, propositions, or principles, psychologists typically use the vocabulary of *norms* to account for moral facts about moral agents' actions. Psychologists are more specifically concerned with *rules of behaviour* and not norms about the goodness and badness of some states of affairs. I shall concentrate on such rules of behaviour here, as the psychology literature can help us shed light on the phenomenon of moral understanding.

Reflective Self-Control

It is likely that the type of reflective self-control endorsed by Korsgaard and Kantian philosophers admits prior stages of development. Indeed, there are various ways in which agents can control themselves and be guided by norms when acting. I shall mention three of them here: by being internalised, accepted, or endorsed. Let us summarise these three degrees of sophistication.

First, a norm can be *internalised*. An internalised norm is typically not entertained *as a norm* by agents but nevertheless motivates them to act. Norms such as crossing the street on a pedestrian light or waiting in line at the bank are typical examples of norms that are internalised by human beings and guide their behaviour. We very often follow them without entertaining them as norms when acting. Moreover, norms can become internalised under the influence of different factors and through using various means: punishment, understanding of others' expectations of behaviour and moral education in the case of very young children, or after being explicitly accepted and endorsed by older human beings.[128] For instance, when being asked not to look at or play with a toy that is not theirs, a certain proportion of children could resist that desire even in the absence of experimenters, which suggests that their behaviour

[128] See Campbell (1964), p. 392.

was not merely informed by the presence of an external authority. Moreover, children who resisted the impulse to play or look at a toy often talked to themselves. Participants were observed reminding themselves not to touch the toy.[129]

Moreover, it is often taken for granted that internalised norms bear a strong link with agents' motivation and that this relation is direct and robust.[130] Internalized norms provide "an intrinsic normative motivation," according to philosopher Daniel Kelly.[131] However, there exists a debate about the source of internalised norms' motivational strength, which could be found in emotions like fear of punishment. It could also stem from cognitive states like prior beliefs – the belief that we ought to follow norms.[132] Similarly, other definitional issues remain regarding the depth of internalisation, the content of internalised norms, and the type of norms that can be internalised.[133]

Second, norms can be *accepted*. Once individuals have acquired the capacity to see norms as norms, they can accept or reject them. As philosopher Peter Railton notes, when an agent accepts a norm, she "reflectively considers norm N and freely decides to treat it as action-guiding or reason-giving."[134] When accepted, norms bear authority on the agent in the sense that the agent can consciously choose to act in conformity with them and can consider norms as counting in favour of doing a precise action. Norms give them *reasons for action*. In other words, when an agent accepts a norm, she is aware that this norm exists and that she can act on the basis of that norm, which will then become a reason for action.

---

[129] *Ibid.*
[130] Kelly (2020), p. 37.
[131] *Ibid.,* p. 38.
[132] *Ibid.,* p. 39.
[133] For a summary of these questions, see Campbell (1964), p. 395.
[134] Railton (2006), p. 15.

It is likely that this capacity gradually appears with the development of metacognition, which is an individual's capacity to think about her own mental states. Metacognition appears during the first year of life[135] and is well-acquired at the age of three.[136] In the moral domain, Nancy Eisenberg and Michael Hand have found that four-year-old children can make decisions when presented with various moral dilemmas. They can also provide justifications for their choice that are either selfish or "needs-oriented," which means that these reasons are related to the various needs of others.[137] Young children can explicitly express emotions and desires, ask and answer "Why?" questions, and give basic justifications for their past or future actions. This indicates that very young children gradually understand norms and how these can give them reasons for action. They thus seem to possess the capacity to think about norms and make decisions or act on the basis of these norms, though they are still incapable of scrutinising a large variety of norms and principles, at least in the moral domain.

As children's emotional and cognitive capacities develop, they acquire a broader knowledge of norms and can scrutinise a wider variety of them. Eisenberg and her colleagues[138] have identified different sequences in children's reasoning. While very young children use hedonistic and needs-oriented justifications, older children in elementary school use a form of reasoning that reflects their desire to be approved by their parents, teachers, and classmates and to behave in ways that are socially seen as good. Toward the end of their studies in elementary school, children then start to incorporate more abstract moral principles and self-oriented moral emotions like guilt in their reasoning. Their reasoning becomes increasingly based on various moral principles and emotional responses. Unlike younger children, they can

[135] Goupil & Kouider (2016).
[136] Escolano-Prérez, Herrero-Nivela & Anguera (2019).
[137] Eisenberg, Fabes & Spinrad (2006), pp. 659-661.
[138] *Ibid.*

scrutinise a variety of considerations that are not merely hedonistic, needs-oriented, or related to social approbation. The current evidence indicates that this competence starts to be more robustly acquired by the age of nine. Children can thus take into account a wider variety of norms in their reasoning process and consider that some norms may override others in specific moral dilemmas.

Third, agents can *endorse* norms. According to Peter Railton, endorsement is distinct from acceptance in the sense that it involves an evaluative judgment. More precisely, endorsing a norm is different from simply choosing a norm "by an act of will."[139] Endorsement, unlike acceptance, requires an evaluative judgement that includes reasons and grounds, and the source of norms' authority comes partly from these reasons and grounds.[140] Hence, endorsement involves higher-order acts and second-order thoughts. For that reason, it has often been linked to moral agents' capacity to identify with their rational nature.[141]

In Kantian theories of moral agency like Korsgaard's, agents' capacity to endorse norms is essential. Moral agents can scrutinise norms and choose them as grounds for action if they can be universalised and chosen by all rational creatures. Moral agents' capacity for reflective self-control enables them not only to reach moral judgements about what one ought to do, but also to act in accordance with these judgements. In other words, agents capable of self-control can regulate their behaviour in response to moral reasons. These capacities may be possessed only by older children, teenagers, and adults, for they are the only beings who can scrutinise various reasons and see moral norms as universally applying to all rational beings. This capacity is taken to be uniquely human, as rightly pointed out by Korsgaard.[142]

---

[139] Railton (2006), p. 22.
[140] *Ibid.*
[141] *Ibid.*
[142] There is currently no empirical evidence that moral reasoning is possessed by animals, including highly intelligent mammals. That said, some animals show a relatively high level of self-awareness and are capable of

Moral Understanding

Moral agents are often described as individuals who can *understand* moral norms or *facts*, in Rowlands' words. This means that they can understand the diversity of norms, their content, and their underlying principles. Since Elliot Turiel's work on the differences between conventional norms and moral norms,[143] psychologists who work on the moral development of children have focused their attention on this dichotomy and how children acquire the capacity to distinguish moral norms from conventional ones.

More recently, primatologists Claudia Rudolf von Rohr, Judith M. Burkart, and Carel P. van Schairk have argued that another type of norm might be relevant for studying precursors to human social and moral norms: personal norms. According to von Rohr, Burkart, and van Schairk, *personal norms* state "the personal expectations of how an individual wants to be treated."[144] According to the authors, it is highly implausible that individuals adopt social norms about how others should be treated before forming norms about "how oneself ought to be treated."[145] These can concern expectations related to food sharing or playing, for instance. It is thus likely that children can form very early in life expectations about how they ought to be treated. In Chapter 3, I shall come back to the appropriateness of describing such expectations as "norms."

Norms can also be of another type: *conventional*. Turiel claims that conventional norms are norms that are defined by consensus from the members of the same group and/or from their social organisation.[146] They apply in societies where some implicit or explicit agreements

---

metacognition. We can think of great apes, monkeys, and dolphins, but dogs have also demonstrated some limited metacognitive capacities. For a summary of the current evidence and for further references, see Ferrigno, Kornell & Cantlon (2017), p. 1.

[143] See Turiel (1982).

[144] Rudolf von Rohr, Burkart & Shaik (2011), p. 9.

[145] *Ibid.*

[146] Turiel (1982), p. 38.

regarding these norms can be found. Their validity depends on a given social context. This is the case with human norms regarding clothing and etiquette, which vary across social contexts.

Norms can also be *moral*. Researchers working on the nature of moral norms either define moral norms as a particular type of social norm that governs beneficial or harmful behaviour,[147] as social norms that have been moralised[148], and as norms that are highly different from conventional ones. These moral norms can include prohibitions, obligations, and more general moral principles. Furthermore, according to Turiel and his colleagues, the validity of moral norms is seen by study participants as universal in the sense that it does not depend on the social context or social agreements of human groups.[149] Moreover, moral norms deal with issues regarding harm (to others and the group's interests), rights, and fairness, and their violation usually triggers a "signature response" like stronger emotional responses from study participants.[150]

Elliot Turiel and Larry Nucci have found that children as young as two years and 10 months already start grasping the distinction between moral and conventional norms.[151] Indeed, they consider that violations of moral norms are more serious than violations of conventional ones. By the age of four years, children recognise that conventional norms are contingent[152] and, between the ages of six and nine, they show greater competence at distinguishing the two types of norms in complex situations, though children aged between

147 FeldmanHall, Son & Heffner (2018), p. 2.
148 Van Schoelandt (2018), p. 231.
149 Turiel (1982), p. 156. See also Kelly, Stich, Haley, Eng & Fessler (2007), p. 118.
150 Turiel (1982), pp. 46-48. However, Turiel's work has received criticisms in the last decades. For instance, researchers have found that some conventional norms like religious and etiquette norms are often seen as independent from authority. Conversely, some moral norms may be seen as authority dependent or as more loosely related to the signature emotional response of moral norms. Finally, some philosophers have argued that moral norms encompass greater concepts like harm/care, fairness/reciprocity/, in-group/loyalty, authority/respect, purity/sanctity, and liberty/oppression. For an overview of these criticisms and further references, see O'Neil (2017) and Kelly (2022).
151 Nucci & Turiel (1978), p. 406.
152 Smetana & Braeges (1990), pp. 329-346.

three to nine years tend to see harming out-group members as being more similar to infringing conventional norms than universal, moral norms.[153] According to researchers Marjorie Rhodes and Lisa Chalik, these findings can be explained by the fact that social categories, such as being part of a specific group, greatly influence children's moral reasoning even though they can make basic distinctions between moral and conventional norms. Hence, young children do fully understand the universal and impartial dimension of moral norms – a capacity that lies at the heart of Korsgaard's theory of moral agency.[154]

1.4. Defending a Gradualist and Multi-Faceted View of Moral Agency

In sum, the core capacities that have been associated with moral agency in different theories of moral agency admit degrees of sophistication, and I sought to offer a brief summary of children's moral development in the last section. The gradual development of human beings' moral capacities raises important questions for Korsgaard's and Rowlands' accounts of moral agency, which will have to face substantial problems in identifying when exactly we become moral agents.[155]

---

[153] Rhodes & Chalik (2013).

[154] In the case of animals, it is highly improbable that they can grasp the distinction between social or conventional and moral norms, though social mammals may exhibit a sensitivity to norm violations related to harm and cooperation and, arguably, fairness. Nevertheless, stronger empirical evidence on animals' norms responsiveness is needed to study how various norms are integrated into animals' psychology and to what type of norms they are sensitive. Studies on the subject have been carried out primarily on primates so far. See, for instance, van Wolkenten, Brosnan & de Waal (2007), Proctor, Williamson, de Waal & Brosnan (2013), Range, Horn, Viranyi & Huber (2009), and Wascher & Bugnyar (2013).

[155] I shall not raise criticisms regarding the work of recent thinkers who have defined animals' behaviour as moral or proto-moral, for the simple reason that these philosophers and scientists all agree that moral agency is a matter of degree and do not reject the vocabulary of "moral agency" to describe animals' behaviour. Nevertheless, this does not mean that I find myself agreeing with these authors. I think the criticisms I raise against Korgsaard's theory could also be applied to Singer, to some extent, since both agree on the importance of reflective self-control for moral agency. Furthermore, I could have criticized Singer for his methodology, which consists of giving empirical support to his definition of moral agency by citing only one study that has been thoroughly criticized since its publication. Other criticisms regarding the value of special obligations to in-group members and duties of loyalty have also been raised by de Waal, among others. See, for instance, de Waal (2006b), p. 165.

1.4.1. Problems with Korsgaard's Theory of Moral Agency

Perhaps surprisingly, I think Korsgaard describes a phenomenon, namely, reflective self-control, that is real and unique to adult human moral agency. It also plays an important role in grounding both moral and legal responsibility, for individuals who are capable of reflective self-control are moral agents who can be held morally and legally responsible for their behaviour.

Yet her account raises important questions regarding the inclusion or exclusion of children from the category of moral agents. Indeed, I mentioned earlier the fact that most children younger than nine years old see the norm of not harming members of other groups as a conventional one.[156] This shows that they do not grasp the universal scope of moral norms and do not see moral norms as applying to all rational beings. Furthermore, we have seen that evidence on the development of reasoning in moral children indicates that children start to incorporate diverse moral principles in their reasoning and can act in accordance with the outcome of a robust moral deliberation process only toward the end of their studies in elementary school. Such empirical evidence suggests that *only* older children may satisfy the Kantian requirement of reflective self-control and thus be included in Korsgaard's definition of moral agents.

Korsgaard's theory implies that young children's behaviour falls outside the scope of morality, a conclusion that most psychologists[157] and several philosophers[158] would find implausible. A five-year-old who shows great concern for her classmates could not be seen as exercising moral agency, even though young children already possess significant moral

[156] Rhodes & Chalik (2013).
[157] For a summary, see Montreuil, Noronha, Floriani & Carnevale (2018). See also the various works on children's moral development cited in this chapter.
[158] See, for instance, Andrews (2009), pp. 433-436, Burroughs (2020), Dixon (2008b), Traina (2009), and Dwyer (2003).

capacities like the capacities to feel empathy, grasp moral norms *as* moral norms and understand what is good and bad. Because children are not moral agents and lack the capacity for reflective self-control, their actions are simply *amoral* in Korsgaard's Kantian framework. Parents and teachers may describe children's behaviour as participating in morality in a certain way. For example, the child might be described as a "good student". Yet this label could only be justified by appealing to instrumental reasons, for instance, to favour children's moral development as individuals who will eventually be capable of reflective self-control and develop into full-fledged moral agents.[159] But young children do not possess the capacities to think about their emotions and desires, to act in accordance with the judgement that resulted from this deliberation process, and to see moral norms as applying universally to all rational creatures. They are not moral agents, according to Kantians.

Faced with this criticism, Korsgaard has three choices: (1) accept that most younger children who are incapable of reflective self-control can be moral agents, (2) reject this first possibility but admit that older children *can* be moral agents, or (3) deny that both younger and older children can be moral agents.

(1) The first option is just not a plausible one for a Kantian, since it profoundly threatens the view that reflective self-control lies at the heart of moral agency. Furthermore, it would force Korsgaard to concede that children's other moral capacities, like empathy or moral understanding, may play a more fundamental or equally fundamental role than reflective self-control. According to that picture, empathy and moral understanding would be important capacities *on their own* - not just because they can give rise to motives for action. It would also force her to put aside her threshold account of moral agency in favour of a more gradualist one. It is thus easy to see why the first response is not one a Kantian philosopher can accept.

[159] On this claim, see Schapiro (1999), especially pp. 734-737.

(2) However, the second option, which consists of denying that young children can be considered moral agents while including older children in that category, comes with its own set of problems. One challenge for Korsgaard will be to explain why different degrees of moral competence still exist within the category of moral agents who possess reflective self-control. For instance, although adult human beings, teenagers, and older children are capable of reflective self-control, it still seems like adults possess a more robust form of moral agency and have greater moral capacities. Korsgaard could provide two answers here to explain this phenomenon.

The first solution consists of admitting degrees of reflective self-control and assuming that adults are capable of a more robust type of self-control. Korsgaard could claim that a full moral agent possesses greater knowledge of moral norms and is thus capable of integrating more elements in her deliberation process. This, in turn, may explain why adults show more robust capacities for self-control.

Second, Korsgaard could appeal to emotional capacities to explain why some moral agents, even within the category of self-reflective agents, exhibit greater moral competence than others. For example, though older children, teenagers, and adults may all have reached a certain level of reflective self-control, they may not possess the same capacity for empathy. Studies on toddlers and young children indicate that the affective dimension of empathy is already present from a young age but that the cognitive components of empathy increase as children develop more sophisticated intellectual capacities and acquire greater knowledge of the moral and social world. In fact, Nancy Eisenberg's work shows that empathetic justifications in children's moral reasoning are more preponderant as children grow up, as already noted. These are also correlated with other justifications related to the positive or negative consequences of one's actions on others' emotional states. Eisenberg also highlights

how teenagers' justifications reflect more "substantial amounts of reasoning" that have a strongly empathetic orientation.[160] Thus, it is possible that adults' greater empathetic capacities explain their more robust self-control. Empathy is more extensively included in their moral reasoning, and this robust form of empathy may provide a motivational force to their moral judgements.

In fact, Eisenberg's findings are compatible to some extent with Korsgaard's views on emotions. In *The Constitution of Agency* (2008), Korsgaard recognises that emotions can play an important role in our moral lives. In her book, she argues that emotions are a form of perception of a reason that can "help us to form and act on correct *conceptions* of the good".[161] Therefore, emotions can be relevant to the exercise of moral agency insofar as they inform the deliberation process, fall under the control of reason, and strengthen agents' tendencies to act in accordance with their best moral judgements.

However, adopting these two solutions could lead Korsgaard to concede that moral agency, even when grounded in reflective self-control, is a gradualist phenomenon that also interacts with other moral capacities that develop simultaneously with agents' capacity for self-control. To put it in another way, she will have to argue that reflective self-control may admit different degrees of sophistication or robustness and that these different levels of development in reflective self-control may be informed by a complete scrutiny process, which, in turn, relies partly on other moral capacities such as empathy.

This objection is a mild one and is not fatal to Korsgaard's views on reflective self-control. Indeed, she could reply to this reservation by simply biting the bullet and claiming that reflective self-control is a phenomenon that develops gradually but one that needs to meet

[160] Eisenberg, Fabes & Spinrad (2006), p. 661.
[161] Korsgaard (2008), p. 18.

a certain line of development to be described as "moral agency". Yet *explicitly acknowledging* that reflective self-control is a matter of degree rather than an all-or-nothing capacity and that these degrees are related to other capacities relevant to moral agency may better explain why, even among rational moral agents who meet a certain threshold of normative self-government, there exist different degrees of moral competence.

(3) Finally, a third option is available to Korsgaard: raise the criteria for reflective self-control higher and deny altogether that both young and older children can be moral agents. This would mean that only adults could be moral agents because they possess a sufficiently high degree of reflective self-control. As philosopher Michael Tiboris notes, one could defend this thesis by arguing that older children lack a capacity for reflective self-control that reflects a *stable* and robustly *authoritative* perspective on their own deliberation process.[162] In other words, older children lack a *consistent* deliberative perspective on their own emotions, desires, and purposes and *lack authority* in two ways: they cannot choose a moral principle over another and cannot translate moral principles into action. Thus, being capable of *some degree* of reflective self-control is not sufficient for older children to be included in the category of moral agents. The capacity for reflective self-control must meet a certain threshold of robustness.

However, even if we accept that agents can have a stable perspective on their own motives for action, the solution of denying moral agency to non-adults will have to face two problems. First, it is likely that children can meet such a more robust threshold of reflective self-control. Tiboris highlights how several studies on children's moral development suggest that children already have personality traits around the age of two or three years and that several remain stable over the years.[163] These personality traits may lead children to act in very

[162] Tiboris (2014), p. 83.
[163] Tiboris (2014), p. 82. See also Fu, Evans, Xu & Lee (2012), Kochanska & Aksan (2006), and Mischel, Shoda & Rodriguez (1989), cited in Tiboris (2014), p. 82.

specific ways and may partly provide the "stable deliberative perspective" that children are often shown to be lacking in the Kantian perspective.[164] Furthermore, children show a very stable conception of themselves starting from the age of eight years and are capable of "pervasive interpersonal comparisons". As Tiboris writes, "these comparisons are essential for principled decision-making since they provide context for deciding the value of one action over another."[165] Toward the end of elementary school, children can also take into account a variety of moral principles in their reasoning, as we have seen.[166] Thus, the claim that children lack a sufficiently stable perspective on their own deliberation process and cannot meet a certain threshold of reflective self-control is subject to counterexamples.

The same thing can be said regarding children's lack of authority over their own motivations, for the evidence I mentioned previously indicates that children can already scrutinise moral principles and act in accordance with them. This suggests that children can possess a sufficiently robust degree of authority over their own behaviour to be defined as moral agents, as they have a stable deliberative perspective that is provided by consistent self-conceptions. They are also aware of various principles, can see moral norms as being universal, can choose one norm over another, and can act in accordance with that rule. Hence, it seems like older children can show a form of reflective self-control that is sufficiently stable and authoritative to meet a higher threshold of moral agency.

The option of excluding older children from the realm of moral agents faces a second problem. Even though we do not accept that older children can meet the criterion of reflective self-control, the third solution to the puzzling case of older children faces an epistemic

---

[164] That said, I suspect that Kantians are searching for a stronger criterion than personality traits to ground the stability of self-control.
[165] Tiboris (2014), p. 83. See Fu, Evans, Xu & Lee (2012), Kochanska & Aksan (2006), and Mischel, Shoda & Rodriguez (1989), cited in Tiboris (2014), p. 83.
[166] See Eisenberg, Fabes & Spinrad (2006), pp. 659-661.

problem: the challenge of identifying when exactly, in their life, agents become *moral* agents and thus pass from the first category to the other. Though I suspect that the answer could be something like "adulthood" for Kantians, it remains unclear when, precisely, in adulthood, agents acquire such robust and well-established capacity for reflective self-control. In fact, establishing such a sharp line between agents may simply be empirically difficult. It could render Kantian theories of moral agency vulnerable to mistakes in their way of distinguishing between moral agents and moral patients.

To answer these criticisms, proponents of threshold views of moral agency like Korsgaard could argue that although drawing a line between moral agents and moral patients or moral subjects is hard, it is not an impossible or vain task. After all, some moral capacities can be examined with the help of thresholds. This is especially true of legal responsibility, which relies on assessments of agents' capacity for moral understanding, self-control, and knowledge of legal systems and whether agents' capacities meet the degree of development or robustness required for legal responsibility.

I do not wish to deny the usefulness of threshold approaches for criminal punishment here. That said, I contend that in moral matters, we might have strong reasons not to buy into such a picture of moral agency. These reasons can be *epistemic* in the sense that they concern how arduous and arbitrary establishing a threshold of moral agency can be. But these reasons can also be *moral,* that is, related to the morality of how we currently treat and ought to treat children and animals. Indeed, describing several animals merely as moral patients runs into the risk of overlooking the specific ways in which we can harm them, such as disappointing their expectations of good will and giving rise to anger. These types of harm, as we will see in Chapter 5 and my thesis conclusion, are not fully captured by the moral patiency and moral

subjecthood paradigms. They also raise new questions related to how we should bolster theories of animal well-being.

In contrast, proponents of a gradualist and multi-faceted view of moral agency do not have to respond to the puzzling case of young and older children. Indeed, their answer is starkly simple: children are moral agents to some degree. They are moral agents for two reasons: (1) they possess some capacities related to moral agency, and (2) these capacities admit different stages of development. Furthermore, these philosophers do not have to respond to the empirical challenge of determining when exactly an agent becomes a moral one: many individuals, including newborn infants, can be moral agents because they are capable of emotional contagion, and that capacity is relevant to moral agency in an epistemic way. Yet they occupy the lowest degree of moral agency among many social mammals and might not be held morally responsible. I shall expand on this claim in Chapter 2.

1.4.2. Problems with Rowlands' Theory of Moral Agency

Mark Rowlands also contends that animals cannot be moral agents but on grounds entirely different from those of Korsgaard. In *Can Animals Be Moral?,* Rowlands does not deny that moral agency is a matter of degree but still argues that moral agents are morally responsible and that moral responsibility rests on moral understanding. Because animals cannot understand moral facts, they cannot be held morally responsible for their actions and be described as moral agents. They can only be moral subjects. Even though Rowlands acknowledges that moral understanding and, by extension, moral agency are a matter of degree, I think his arguments are flawed for several reasons.

Let us recall the example of Babe, the naughty pig who steals turnips. According to Rowlands, Babe should be punished and even legally prosecuted, but in a mitigated way, if we

take her to be a moral agent. I think there is one problem with that case: moral responsibility does not entail full legal responsibility, especially if we acknowledge, like Rowlands, that moral agency is a matter of degree. There is no contradiction between holding an individual *morally* responsible to some degree and denying that this individual cannot meet some additional criteria or threshold to be held *legally* responsible. We may argue that even though children cannot be held legally responsible, they can be held morally responsible to a certain extent.[167] Legal responsibility is a distinct concept that can require a certain high degree of moral responsibility and some basic knowledge of laws and legal systems.[168] Animals will *never* be able to meet such a high standard. The example of animal trials used by Rowlands seems to be out of order.

That said, Rowlands's remarks on the case of Babe could be interpreted differently and not just in light of legal responsibility. The example of Babe could be read as presupposing a retributive theory of punishment. According to retributivists, punishment constitutes a form of hardship imposed on *morally responsible* agents who *deserve* to be punished. Still, this second interpretation is problematic. A form of partial or mitigated moral responsibility to moral agents does not necessarily lead to mitigated punishment for *all* actions. When philosophers argue that a moral agent possesses a lower degree of moral agency and moral responsibility, they usually do not simply mean that the agent should be punished in a softer way for *all* her actions. Rather, several of them argue that moral agents have moral responsibility for the actions they can morally understand and that they cannot be held morally responsible, nor

[167] The question of children's moral responsibility has been a contentious one among philosophers See, for instance, Burroughs (2020), Dixon (2008b), Traina (2009), and Dwyer (2003).
[168] For an overview of the differences between moral responsibility and legal responsibility, see Hart (2008), Ripstein (1999), and Moore (2009), Chapter 3.

punished, for actions that fall outside the scope of their moral understanding.[169] To illustrate the point, we can take the following example proposed by philosopher Paul Shapiro:

> [W]hile it would be appropriate to hold a four-year-old human responsible for hitting his sister, it would not be appropriate to hold him responsible for publicly calling attention to a disabled person on the street (assuming he couldn't be expected to grasp the potential for hurt feelings). In short, the less mentally developed a moral agent, the fewer obligations she will have.[170]

A four-year-old child cannot understand how calling attention to a disabled person publicly can be hurtful to that individual. In the same vein, the child is not aware of the discrimination and challenges disabled persons must face and does not understand how her behaviour may contribute to creating a more hostile social environment for such people. It would thus be inappropriate to hold a young child responsible and punish her in a *retributivist* sense, though a parent may explain to her child why she should not do this in the future.

Rowlands does not make these distinctions between moral understanding, moral responsibility, and punishment and conflates mitigated moral responsibility with mitigated punishment for *all* actions. However, if we argue that animals can have some degree of moral responsibility and thus moral agency, this will not lead to recognising Babe as partially responsible for stealing turnips and punishing her in a milder way. It rather means that Babe will be held morally responsible for the things she can understand or "have access to", to use the vocabulary I put forward in Chapters 2 and 3. Eating someone else's turnips is perhaps not an action she can understand as morally wrong, but maybe one could argue that other

[169] See, for instance, DeGrazia (1996), p. 107, and Shapiro (2006), p. 365. It is worth noting here that a four-year-old child could is less morally responsible and blameworthy than adult human beings. The difference between these two moral agents could be explained by a difference in their moral obligations and blameworthiness.
[170] Shapiro (2006), p. 365.

actions, like biting human beings for reasons that are not related to self-defence, fall under her moral understanding. Therefore, granting partial moral responsibility to animals does not lead to mitigated punishment for *every* action but rather to moral responsibility for a more limited range of actions. In fact, as I will argue in my next chapters, many animals can have access to the badness of others' suffering (Chapter 2), recognise intentional harm, and have access to the wrong-making features of causing suffering (Chapter 3). This, I contend, makes them liable to moral responsibility practices (Chapter 4).

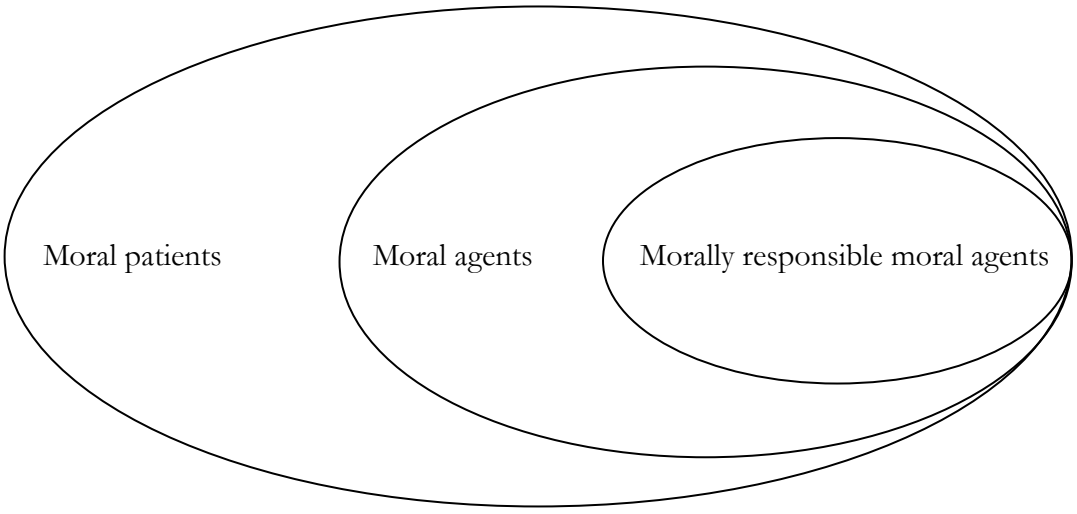
More broadly, the older and newer arguments proposed by Rowlands suffer from three additional problems. The first issue is that Rowlands does not offer any justification in favour of tying moral agency so closely to moral responsibility. He rather takes it as a fact and simply asserts that moral agency is "inseparable" from the concept of moral responsibility. The link between the two notions could be explained by two things. First, agential capacities, such as epistemic and self-control capacities, have been taken by philosophers to ground moral responsibility. Second, neurotypical human adults are paradigmatic examples of moral agents who can be morally responsible. These two reasons may explain why we now consider moral agency and moral responsibility to be closely related.

However, this does not need to be the case. Findings about infants' and animals' capacity for emotional contagion could support the view that emotional contagion is relevant to moral agency in an epistemic sense while being insufficient, *alone*, to ground moral responsibility. As we will see in Chapter 2, basic forms of empathy like emotional contagion could enable infants and animals to access evaluative facts – facts that concern states of affairs. Yet, as I argue in Chapter 3, mere emotional contagion cannot grant animals and young children access to deontic facts – facts that concern actions. Similarly, it does enable them to engage in moral responsibility practices. Hence, emotional contagion could already play an

important epistemic role in allowing infants and animals to have access to evaluative facts, but without being sufficient for access to deontic facts and being held morally responsible. I shall come back in more detail to these arguments in my next chapters.

Moreover, the development of some artificial agents like robots and self-driving cars invites us to think of moral agency as being more loosely linked to moral responsibility.[171] By the same token, these new cases could lead us to argue that moral agency is a phenomenon that is best described as a multi-faceted phenomenon. Hence, the relationship between moral patiency, moral agency, and moral responsibility may look like this if we concentrate here on animal moral agents and put aside the thorny case of non-sentient artificial moral agents, which could be moral agents without being moral patients:

Moral Patients, Moral Agents, and Morally Responsible Moral Agents



[171] For an overview of the debate on artificial moral agency, see Misselhorn (2022) and Coeckelbergh (2009).

Rowlands must also face a second problem: explaining why we should sharply make a distinction between moral understanding and moral subjects' capacities and why the two concepts cannot fall on the same continuum. It is likely that before agents acquire moral understanding, they already have some capacities that are necessary for its development. For instance, six-month-old toddlers, when put in situations where they have to choose between helping and hindering individuals and then between neutral and hindering ones, show very clear preferences for others who are either helping or neutral.[172] Hamlin, Wynn and Bloom, the researchers who conducted that study, conclude that such evaluative capacities may "serve as a foundation for a developing system of moral cognition" and may be a basis "for any system that will eventually contain more abstract concepts of right and wrong."[173] Hence, it remains unclear why exactly we should sharply distinguish between moral subjects' capacities from a more robust type of moral understanding, since the latter seems to rely on other pre-reflective capacities, like the capacity to "track" moral propositions without the direct possession of moral concepts. We can best think of these two capacities as falling on the same spectrum.

In fact, Rowlands acknowledges this possibility in *Can Animals Be Moral?* and does not argue against it. He rather claims that if we define moral agency in relation to moral understanding, and if we see that capacity as admitting degrees of development, we could think of animals as "occupying a position on the far left" of that spectrum, "with normal, adult hams clustering toward the right."[174] He also concedes that "we may think of the distinction between a (mere) moral subject and a moral agent as one of degree rather than kind."[175] Hence, given

[172] Hamlin, Wynn & Bloom (2007), pp. 557-558. However, their findings were recently criticized by Lucca, Yuen, Wang *et al.* (2025).
[173] Hamlin, Wynn & Bloom (2007), pp. 558-559.
[174] Rowlands (2012), p. 241.
[175] *Ibid.*

the empirical evidence on children's moral development and Rowlands' tendency to admit that the difference between moral subjects and moral agents may be merely a matter of degree, it becomes hard to see why one should remain strongly committed to the view that there exists a sharp distinction between moral subjects and moral agents.

Finally, Rowlands will have to respond to a third challenge: the same as Korsgaard, which consists of identifying *when* precisely an individual becomes a moral agent or, in the case of Rowlands, passes from the category of a moral subject to that of a moral agent. As mentioned before, this may be an extremely difficult task. It could also give rise to important epistemic and moral mistakes, and Rowlands does not provide such a precise standard or threshold that would allow us to distinguish between moral subjects and moral agents. He claims that moral understanding rests on the four capacities I summed up earlier: the capacities (1) to make qualitative distinctions between positive and negative states like others' happiness and pain, (2) to grasp moral facts, (3) to understand why something is right or wrong, and (4) to understand the moral principles that underwrite moral facts.[176] However, as highlighted in §1.2, moral understanding develops gradually, and it remains unclear when children could be said to possess it. Would, for instance, a three-year-old child who can grasp *some* differences between conventional and moral norms but not others be a moral agent?

On the contrary, philosophers who argue that moral agency should be understood as a gradual phenomenon do not get tangled up in these epistemic issues, which also have consequences for the way we currently treat or ought to treat animals. That said, proponents of gradualist and multi-faceted understandings of moral agency must clarify the relationship between moral agency and moral responsibility. This requires, in turn, adopting a very specific approach to moral responsibility, examining whether animals could be morally responsible

[176] *Ibid.,* p. 239.

according to such theory, and identifying specific actions for which animals could be held morally responsible. I will attempt to fulfil this objective in Chapter 4, in which I argue that many animals could be morally responsible according to Strawsonian accounts of moral responsibility. For now, it is sufficient to stress that philosophers who defend the view that animals can be moral agents put forward an account of moral agency that is more straightforward, as it does not have to establish a clear threshold of moral agency that would run into the problems highlighted here. This offers an obvious theoretical advantage to gradualist and multi-faceted accounts of moral agency while raising new questions and challenges pertaining to the relationship between moral agency and moral responsibility.

1.5. Summary of Chapter 1 and Concluding Remarks

The different views that have developed in the debate on animal morality and that I have explored in this chapter (§1.2) can be classified under the following table. Here, I have put aside older philosophical works like Hume's and Smith's and more empirical views like Darwin's and de Waal's to concentrate on the recent philosophical arguments.

The Debate on Animal Morality

Author	**Can animals be moral agents?**	**Capacities underlying animal moral agency**	**Can animals be morally responsible?**
Stephen Clark (1984)	Yes, or "ethical agents"	Responsiveness to the good- and bad-making features of situations, capacity to act virtuously[177]	Unspecified
Steve Sapontzis (1987)	Yes	Intentionality, flexibility, responsiveness to the good- and bad-making features of situations[178]	Unspecified

[177] Clark (1984), p. 107.
[178] Sapontzis (1987), pp. 32-24.

Marc Bekoff and Jessica Pierce (2009)	Yes	Cooperation, empathy, and justice[179]	Yes, in the context of their own community[180]
David DeGrazia (1996)	Yes	Manifest virtues, act in a way that is not instinctive or conditioned[181]	Yes, but for a limited range of actions and according to their moral capacities[182]
Kristin Andrews (2009)	Yes	Empathy and capacity to respond to norms[183]	Unclear
Evelyn Pluhar (1995)	Yes	Emotions, memory, and purposive behaviour[184]	Unspecified
Dorna Behdadi (2021 and 2024)	Yes	Having the participatory competence to engage in exchanges that involve reactive attitudes[185]	Yes
Asia Ferrin (2019)	Yes	Manifest good will or ill will, here defined as responsiveness to moral reasons[186]	Yes, where there is an overlapping social context[187]
Paul Shapiro (2006)	Yes	The capacity for virtuous behaviour[188]	Yes, but for a limited range of actions and according to their moral capacities[189]
Peter Singer (2006)	Unclear. They are perhaps proto-moral agents	Reflective self-control and the capacity for moral circle expansion[190]	Unspecified
Christine Korsgaard (2006)	No, they are moral patients	Reflective self-control[191]	No
Mark Rowlands (2012)	No, they are moral subjects who can feel	Moral understanding[193]	No

[179] Bekoff & Pierce (2009), p. XIV and p. 113.
[180] *Ibid.,* p. 144.
[181] DeGrazia (1996), p. 63.
[182] *Ibid.,* p. 107
[183] Andrews (2009). See also Andrews & Gruen (2014), pp. 203-205.
[184] Pluhar (1995), p. 55.
[185] Behdadi (2020 and 2024).
[186] Ferrin (2019), pp. 137-138 and pp. 143-144.
[187] *Ibid.,* p. 146.
[188] Shapiro (2006), p. 365.
[189] *Ibid.,* p. 369.
[190] Singer (2006), p. 145 and p. 150.
[191] Korsgaard (2006), pp. 112-113.
[193] *Ibid.,* p. 239.

	morally laden emotions[192]		
Susana Monsó (2015 and 2017)	No, they are moral subjects who can feel minimal moral empathy[194]	Moral understanding	No

Moreover, the arguments proposed in this chapter suggest that a gradualist and multi-faceted conception of moral agency better takes into account the different degrees of development that capacities like reflective self-control and moral understanding admit (§1.3). Unlike Christine Korsgaard's account of moral agency, a gradualist theory can better address the theoretical tension that the various findings on children's moral development have introduced for threshold accounts of moral agency (§1.4.1). We have also seen that the arguments provided by Rowlands fail to show that animals cannot be considered moral agents and cannot explain why we should see moral agency grounded in epistemic capacities as *always* linked to moral responsibility (§1.4.2).

Hence, this chapter leaves open the question of where animals can be placed on the continuum of moral agency, which moral capacities they possess, and to what extent these are relevant to the exercise of moral agency, either in an epistemic or self-control way. As mentioned previously, many philosophers agree that empathy is an important moral capacity. The role it can play will be the topic of Chapter 2. In the next chapters, we will also see that my account of animal moral agency builds on the following general ideas:

(1) Animals' empathy can allow them to have access to moral propositions such as "Suffering is bad" (Mark Rowlands and Susana Monsó).

[192] Rowlands (2012), p. 69.
[194] Monsó (2015), pp. 674-676 and Monsó (2017), p. 340 and pp. 350-353.

(2) Empathy allows animals to be responsive to moral reasons pertaining to others' well-being (Asia Ferrin and Paul Shapiro).

(3) Animals could be held morally responsible for a limited range of actions (David DeGrazia and Paul Shapiro).

(4) Animals' anger can be described as a form of blame, as a moral responsibility practice (Dorna Behdadi).

(5) Moral responsibility practices partly rely on moral agents' social context and relationships with each other (Marc Bekoff and Jessica Pierce, Paul Shapiro, Asia Ferrin, and Dorna Behdadi).

The next chapters examine and question some components of Monsó's account of empathy, especially its motivational outcome (Chapter 2). They also offer a more detailed theory of empathy's role in allowing animals to have access to the wrong-making features of causing suffering (Chapter 3), to engage in moral responsibility practices with other moral agents, and to adjust to others' emotional states such as suffering and anger. (Chapter 4). It is now time to develop my own account.

Chapter 2

Empathy's Epistemic Role: Can Animals Have Access to the Badness of Others' Suffering?

In *Germinal* (1885), French writer Émile Zola describes not only the ordeal of coal mine workers, but also that of horses. In the 19th century, many horses were taken down into the mines and forced to live, work, and die underground. Interestingly, the novel features two equine characters, Trompette and Bataille, who form a very strong bond. When Trompette dies of exhaustion or illness, Bataille is inconsolable. Zola describes his distress very poignantly:

> Since he [Trompette] had gone underground, he had never been able to become acclimatized. He remained listless, taking no pleasure in his work, as if he were tortured by the thought of the missing daylight. In vain Bataille, the longest-serving horse in the mine, nuzzled up to his ribs in friendly fashion, and nibbled his neck, to pass on a little of his ten years' experience of underground resignation. These caresses only aggravated Trompette's melancholy, and his skin shivered as he felt the confidences of this comrade who had grown old in the darkness; and both of them, every time they nuzzled each other in passing, *seemed to be lamenting*, the old one for not being able to remember, the young one for not being able to forget. In the stables they were neighbours at the same stall, and they spent their time with heads hung low, breathing into each other's faces, endlessly exchanging their reminiscences of daylight, visions of green grass, white roads, yellow light, and so on. Then when Trompette was drenched in sweat, and lay dying in agony on his bed of straw, Bataille started *to nuzzle him desperately, with short sobbing sniffs. He felt him growing cold*, the mine was taking away his last pleasure, this friend who had fallen from above bringing fine fresh smells reminding

him of his youth in the open air. *And he had broken his tether, whinnying with fear*, when he realized that the other horse had stopped moving.[195]
(emphasis added)

Although Zola describes Trompette's regret and Bataille's distress in anthropomorphic terms that presuppose the possession of long-term memory, he was right in describing Bataille as capable of feeling the emotions of another individual. Horses are emotional "sponges." They are happy when others are happy, feel nervous when others are nervous, and suffer when others suffer.[196] As we have seen, many mammals and birds can "catch" others' emotions. This capacity, called "emotional contagion" in the psychology literature, is often included in the umbrella concept of empathy.[197]

Emotional contagion, although basic, is also potentially relevant to moral agency, even when it does not lead to altruistic motivation or is not supported by more sophisticated moral capacities like the one for moral judgment. Suffering because another individual suffers seems to play a role in our moral lives. However, many philosophers have been reluctant to acknowledge this. They have often claimed that emotional contagion is fundamentally self-absorbed and rests on a lack of self-other distinction.[198]

In this chapter, I argue that these philosophers have dismissed the importance of emotional contagion too quickly, and I take up the challenge of providing an account of its importance. I propose and defend an account of empathy's epistemic role grounded in emotional contagion. More precisely, I argue that basic forms of empathy like emotional

[195] Zola (1885), p. 421.
[196] On emotional contagion in horses, see, for instance, Keeling, Jonare & Lanneborn (2009). See also Pérez-Manrique & Gomila (2022), section 2.1.6.
[197] Preston & de Waal (2002).
[198] See, for instance, Aaltola (2018), p. 82, Nussbaum (2013), p. 145, Deonna (2007), p. 111, Solomon (2008). p. 68, and Blum (2011), p. 172. I shall come back to these arguments in §2.4.

contagion enable animals to have access to the badness of others' suffering even though they cannot entertain a moral proposition like "Suffering is bad" and directly evaluate the suffering of others as bad. A creature's emotional contagion is epistemically relevant by virtue of three criteria: (1) the intentional object of her empathy is another individual in a state of suffering, (2) she suffers in tune with the sufferer, and (3) this feeling allows her to have access to the badness of others' suffering. No other criterion, like motivation to help, is needed to access this moral fact.

To defend this thesis, I first expose several difficulties that come with developing an account of empathy's epistemic importance. I also clarify the different possible definitions of concepts such as "moral emotions" and "empathy," and I specify the conceptual links between emotional contagion and empathy. I narrow the scope of my account to one type of empathy: empathy based on the *perception* of suffering in others and whose *intentional object* is another individual in a state of suffering (§2.1). I then defend the three proposed components (§2.2). Finally, I shall expand on the notion of "having access" (§2.3) and on the literature on emotional contagion in infants. (§2.4).

2.1. Defining Empathy: Some Clarifications

Before developing my account, it is worth highlighting some important difficulties that come with the challenge of defining empathy and developing an account of the role its basic manifestations can play in our moral lives. Two questions need to be discussed: (1) What is a *moral* emotion? (2) What is empathy?

2.1.1. What Is a Moral Emotion?

There are different ways in which an emotion can be linked to moral agency and can be called a *moral* emotion. As philosophers Florian Cova, Julien Deonna, and David Sander note, we can identify five definitions of moral emotions in the philosophical literature on emotions and morality. First, some emotions can be *moral* in virtue of their way of evaluating an object as "instantiating a certain moral value."[199] Feeling guilty about a previous action usually involves an evaluation of that action as *wrong*. Guilt thus instantiates a moral value and can be said to be a moral emotion precisely because of its way of evaluating its object. Second, emotions can be moral in an epistemic way, which means that they may play a role in the formation of our moral judgements and can even enable us to have access to some moral facts. Third, emotions can be moral because they can motivate us to act in a way that is morally praiseworthy. Fourth, some emotions can be qualified as moral because their cultivation is essential to morality and can "contribute to fostering morality" in individuals and groups.[200] Fifth, emotions can be moral because they can be morally evaluated. Agents can be praised or blamed for their emotional reactions or failures to emotionally react to certain situations.[201] In the same vein, philosopher Kevin Mulligan notes that emotions can be moral by virtue of various components, like their object, their nature, the motives they can give rise to, their functions, or their effects, among many.[202]

When we keep these distinctions in mind, it is easy to see how much of the debate on the importance of animal empathy rests on a certain conceptual confusion between these different possible definitions. For example, in *Animals, Emotion, and Morality* (2008),

[199] Cova, Deonna & Sander (2015), p. 397.
[200] *Ibid.,* p. 398.
[201] *Ibid.*
[202] Mulligan (2009a).

philosopher Beth Dixon denies that animals can feel a type of empathy that is "morally thick." She argues that in order to be strongly related to morality and be praiseworthy, empathy must come with three judgements: that the suffering of others is serious, that it is undeserved, and that the empathetic individual is vulnerable to similar suffering.[203] Such empathy can be praiseworthy because "it is directed at the right individuals and in the right circumstances."[204] Similarly, some authors argue that only human beings are capable of moral emotions because they possess greater self-awareness,[205] moral concepts like culpability and responsibility,[206] and evaluative language.[207]

However, these accounts of empathy and moral emotions seem to be committed to either the first or fifth definition of a moral emotion identified by Cova, Deonna, and Sander. Yet there are other ways in which one can construe and defend an account of empathy's relationship with moral agency. Therefore, it is possible that philosophers have been largely talking past each other when developing their theories of animal and human empathy. The debate may rest on a lack of clarity when it comes to defining in *what sense* an emotion can be moral. There can be many answers to that question.

My own answer is that empathy is epistemically relevant to moral agency for three reasons: (1) because its intentional object is a morally relevant feature and (2) because of its felt dimension, (3) which allows animals to have access to the badness of others' suffering. In other terms, even basic forms of empathy like emotional contagion have an important epistemic role to play. However, I do not defend the view that emotional contagion *alone* is

[203] Dixon (2008a), pp. 66-67.
[204] *Ibid.*, p. 68.
[205] Hauser (2000), p. 224.
[206] Roberts (1996), pp. 154-155.
[207] Joyce (2005), pp. 80-85.

sufficient or sophisticated enough to be the target of moral responsibility practices such as praise and blame. I shall come back to this claim.

2.1.2. What Is Empathy?

A second difficulty comes with the definition of empathy itself. As philosopher Frédérique De Vignemont and psychologist Tania Singer note, "there are probably nearly as many definitions of empathy as people working on this topic."[208] Broadly, empathy can refer to everything that falls between mere emotional contagion to more cognitively complex capacities like perspective-taking and understanding the emotions of others,[209] although some researchers have labelled the two last capacities as "compassion" and have argued that they are psychologically distinct from the affective components of empathy.[210] For instance, Daniel Batson has identified eight uses of the word "empathy" in the psychology literature: (1) knowing another person's internal state, which includes both her thoughts and emotions,[211] (2) "adopting the posture or matching neural responses of an observed other,"[212] (3) feeling the same emotion as another person,[213] (4) projecting yourself in the situation of another,[214] (5) "imagining how another is thinking and feeling",[215] (6) imagining what we would think or feel if we were in the other's situation, which often refers to Smith's concept of the impartial spectator,[216] (7) feeling personal distress when seeing the pain of others,[217] and (8) feeling for

[208] De Vignemont & Singer (2006), p. 435.
[209] Decety & Cowell (2014), p. 337.
[210] Jordan, Amir & Bloom (2016), p. 1107.
[211] Batson (2009), p. 4.
[212] *Ibid.*
[213] *Ibid.*, p. 5.
[214] *Ibid.*, p. 6.
[215] *Ibid.*, p. 7.
[216] *Ibid.*
[217] *Ibid.*, pp. 7-8.

another distressed person, which is often called "empathic concern".[218] These eight definitions thus cover different emotional and cognitive responses to the pain and joy of others, ranging from emotional contagion to imagining another's mental states and *knowing* what a person thinks and feels. Furthermore, Batson notes that these phenomena can be related to each other in some ways, but disagreements persist on the different psychological sources of these states and their conceptual labelling under the word "empathy".[219]

Similarly, several philosophers have argued that empathy should not be defined as a singular emotion. Lawrence Blum claims that we should not think of empathy as a unitary phenomenon but as a collection of sensitivities.[220] In the same vein, Robert Solomon argues that empathy may be not an emotion but a "capacity to have any number of emotions depending on the emotions of others."[221] Andrew Ortony, Gerald Clore and Allan Collins hold a similar view and contend that there are "empathic emotions" but not one emotion that could be labelled as "empathy."[222] Finally, Richard Lazarus argues that empathy is "an emotional capacity and a process" rather than an emotion.[223] These definitions are also consistent with the general claim in the debate on animal morality that empathy should be defined very broadly as an umbrella notion.

Another difficulty that comes with defining empathy concerns the influential view in psychology that there may be two types of empathy: affective empathy, which is characterised by strong emotional arousal at the perception of signs of suffering in others, and cognitive empathy, which is colder. The latter is fostered by perspective-taking, mind-reading, and

---

[218] *Ibid.*, p. 8.
[219] *Ibid.*, p. 3. See also Jordan, Amir & Bloom (2016).
[220] Blum (1994), p. 717.
[221] Solomon (2008), p. 66.
[222] Ortony, Clore & Collins (1988), p. 94.
[223] Lazarus (1991), p. 287.

imaginative capacities, which allow individuals to have a deeper understanding of others' emotional states. However, this type of empathy does not always come with a strong felt component. Children with autism are often mentioned in the literature to illustrate how one can have the affective dimension of empathy but not its cognitive dimension, while primary psychopaths are considered paradigmatic examples of individuals capable of cognitive empathy without affective empathy.[224] It is suggested in the psychology literature that primary psychopaths, unlike secondary psychopaths, are cold-hearted and emotionally detached but often score higher than average in tasks that require taking the perspective of others, evaluating their emotional states, and assessing their emotional vulnerability. Secondary psychopaths, on their part, are capable of emotional contagion but struggle with emotional regulation, which renders them prone to impulsive and harmful behaviour.[225]

The distinction also had a certain influence on animal ethics and the animal morality debate. For example, de Waal claims that great apes are capable of "cognitive empathy", which he defines as the "ability to picture oneself in the position of another individual."[226] Similarly, philosopher Lori Gruen discusses four types of human and nonhuman empathy: (1) emotional contagion, (2) emotional empathy, which is more sophisticated and less automatic than emotional contagion and allows individuals to recognise that others are experiencing specific emotions, (3) cognitive empathy, and (4) entangled empathy.[227] The latter notion is defined as a form of empathy that includes both affective and cognitive components and entails the perception of our relationships with others.[228] Finally, in her account of empathy and its link to animal ethics, philosopher Elisa Aaltola refers to cognitive and affective empathy, in

[224] Aaltola (2014).
[225] Mealey (1995), cited in Aaltola (2014), p. 83.
[226] de Waal (1996), pp. 47-48 and p. 70, and de Waal (2006a), pp. 36-40.
[227] Gruen (2019), pp. 485-486.
[228] *Ibid.,* p. 488.

addition to discussing other forms of empathy, such as embodied empathy, which is aroused by the bodily expressions of others and gives us embodied access to others' emotional states. Aaltola also addresses reflective empathy, a second-order level of empathy that allows us to reflect on our first-level empathy.[229]

We may thus wonder if empathy is best defined as an emotion. On the one hand, empathy does not involve a relatively simple and straightforward evaluation of its intentional object, unlike several emotions. For example, an individual who feels fear evaluates the intentional object of her fear as *dangerous*. In the philosophical literature on emotions, it is often suggested that the evaluative dimension of emotions individuates them.[230] What defines fear as an emotion is precisely that it entails an evaluation of its intentional object as *dangerous*, unlike other emotions. However, sophisticated forms of empathy often come with a diversity of evaluations. These evaluations can be positive or negative and related to the emotional states of others, their well-being, the importance of some individuals in one's life, etc. Hence, we may struggle to identify what is specific about empathy and what individuates it if we want to define it as a singular emotion. This is an important challenge, especially since several authors have argued that empathy can be directed at others' happiness or fear and does not seem to involve only a negative evaluation of others' suffering, as is the case with sympathy and compassion.[231]

On the other hand, empathy seems to share several components of emotions. To start with, they both have an intentional object. They are *about* something.[232] Some forms of

[229] Aaltola (2018).
[230] See, for instance, Deonna & Teroni (2014), pp. 16-17, Lyons (1985), p. 81, Teroni (2007), p. 399, and Naar (2019a), p. 562.
[231] See, for instance, Prinz (2009), p. 532, Singer & Klimecki (2014), p. R875, Goldman (1992), p. 32, Solomon (2008), p. 66, and Lazarus (1991), p. 289.
[232] Nussbaum (2001), p. 302 and p. 27.

empathy are also *felt* and come with various bodily changes, just like many emotions. Finally, philosophers who work on emotions agree that emotions always come with various cognitive states like perceptions, beliefs, rememberings, or imaginings. These various states have been described as the *cognitive base* of emotions in the philosophical literature.[233] Empathy, like emotions, may be grounded in perceptions, beliefs, rememberings, and imaginings. For instance, we may feel empathy when directly seeing a loved one suffer or when thinking about a friend's past suffering. We may also feel empathy that is directed at the suffering of novel and movie characters, though I shall not address the complex link between fiction and emotions here.[234] Even though empathy may not be a unitary emotion, some key concepts of the philosophy of emotion may be helpful in highlighting the structure of empathy and the differences between human empathy and empathy in nonhuman animals.

Furthermore, the notion of "cognitive base" is highly relevant for the study of animal empathy, and this approach has not been put forward so far in the debate on the moral capacities of animals. Reflecting on this notion leads us to narrow down the scope of the type of empathy we can study in animals. Although many animals have memory, including long-term memory and imagination, we currently have no strong evidence of empathy in animals that would be grounded in memories and imaginary scenarios.[235] In fact, this type of empathy may be impossible or extremely difficult to study in animals, at least in animals that cannot be language-trained, understand human language, and explicitly express their emotions regarding past events and imaginings. Such distinction allows us to acknowledge several differences between adult human empathy and animal empathy in regard to their cognitive base and to

[233] For a discussion of this notion, see Mulligan & Scherer (2012), p. 348, Tye (2008), p. 29, and Naar (2019a), p. 569.
[234] For an overview of the literature on fiction and emotion, see Friend (2016).
[235] On memory in animals, see Bruck (2013), Carruthers (2013), Shaw & Harvey (2020) and Jo, McCune, Jablonski & Lee (2023). On imagination in animals, see Peña-Guzmán (2022), Chapter 3.

address the important limitations from which animal empathy suffers. In short, some types of empathy may simply not be available to animals by virtue of their different mental lives.

Given the great variety in empathy's cognitive base and intentional object, my goal in this chapter is modest. I aim to explore the importance and limitations of empathy grounded in *perceptions and beliefs* and empathy that is directed at *another individual in a state of suffering*. This type of empathy relies on direct experience and mental states such as perceptions and beliefs. Unlike other forms of empathy, it takes others in a state of suffering as an intentional object and not in more positive emotional states like joy and happiness, although my account could easily be adapted to account for the goodness of happiness.

Before developing my argument, a few things also need to be said about the notion of "suffering," which is different from the concept of pain. Indeed, suffering can be defined in a *subjective* sense as a negative and aversive experiential state that can be caused by various things: physical pain, tiredness, coldness, intense hunger and thirst, remorse, shame, heartbreak, loss of a loved one, loneliness, depression, etc.[236] Philosophers do not agree on how to describe suffering precisely and on whether mild forms of physical discomfort can be included in the definition of suffering. But they generally agree that it is an unpleasant affective state.[237] Suffering feels bad, is aversive. I shall later expand on these remarks. For now, it is sufficient to stress that the examples of animal empathy I will explore are grounded in a *narrow* definition of suffering. According to this approach, suffering is a *highly* aversive emotional state, not just any unpleasant experience.

Subjective suffering is often distinguished from more objective definitions of suffering, according to which suffering does not need to be felt to be described as suffering. Individuals

[236] Svenaeus (2014), p. 409, Brady (2018), p. 13, and Andrew (2021), p. 23.
[237] Brady (2018), p. 5. See also DeGrazia (2014), pp. 135-139.

do not need to be aware of their suffering or to be in a negative affective state to be said to suffer. According to objective theories, suffering is rather tantamount to the absence of objective flourishing.[238] However, since empathy is usually defined as the capacity to feel and understand the *emotions* of others, the notion of "suffering" in this second chapter should be understood as suffering in the subjective sense. This may also explain major differences between empathy and emotions, such as compassion and sympathy, which may be directed at others' plights and bad situations more broadly rather than mere subjective suffering.[239]

My goal in this chapter is to propose and defend an account of empathy's epistemic role which can successfully respond to the challenges of assessing the importance of emotional contagion, the role that altruistic motivation plays or does not play in grounding empathy's epistemic relevance, and how we can make sense of animals' subjective experience of empathy. This minimal account allows me to argue that the most basic forms of empathy, like emotional contagion, are related to moral agency in some way while making room for different degrees of complexity. In this chapter, I argue that an animal can show a form of empathy that has an epistemic role to play by virtue of three components:

(1) The intentional object is another individual in a condition of suffering.

(2) She suffers in tune with the sufferer, which

(3) allows her to have access to the badness of others' suffering.

[238] See Van Hooft, Stan (1998), pp. 125-131. For comments, see Edwards (2003), pp. 59-66 and Kious (2022), pp. 621-627.
[239] Blum (1994), pp. 173-175 and Nussbaum (2001), pp. 301-302.

As we will see, all mammals, including Zola's horses and some species of birds, can meet the three criteria because they are capable of emotional contagion. Let us now explore these.

2.2. Empathy's Epistemic Role: A Minimal Account

2.2.1. The intentional object of empathy is another individual in a state of suffering.

As previously mentioned, philosophers working on emotions agree that emotions are intentional. They are *about* something. Furthermore, there is a growing consensus among philosophers that concepts and propositional thought are not required for intentionality, and I here take for granted that animals can have intentional states. For example, a deer who feels fear is afraid *of something*, probably a predator. However, the intentional object of an emotion should not be confused with its cause. Though intentional objects can also be the cause of emotions, it is not always the case.[240] For instance, being hungry often leads me to feel angry or "hangry," but my hunger is not the intentional object of my anger. My anger is not *about* my hunger. I am not angry *that I am hungry,* even though my hunger can be the cause of my emotion. The intentional object of my anger is usually my old laser printer.

In the case of empathy, I want to suggest that its intentional object is other individuals in a state of suffering. When an animal feels empathy, she perceives some behavioural cues associated with suffering in another individual, forms the belief that this individual is suffering, and this situation constitutes the intentional object of her empathy. That type of empathy is directed at other individuals in a condition of suffering.

[240] For a discussion of this distinction, see Solomon (1973), p. 25, and Zagzebski (2003), pp. 112-114.

We can already anticipate two objections or reservations to this basic claim. A first objection to my account can be stated as follows: without a full-fledged theory of mind, which is the capacity to ascribe mental states to others that are different from our own, to understand that others can have false beliefs, to understand how others' emotions can be different from ours, and to understand how they can be caused by states of affairs, animals are incapable of empathy. Because most animals do not possess a full-fledged theory of mind, they cannot have a deeper insight into the mental states of others, which is required for empathy. Animal empathy may be best labelled as another emotion or emotional capacity rather than empathy, which requires insight into the mental states of others and more sophisticated mind-reading capacities.

However, this account of empathy's intentionality is problematic for three reasons. First, it seems that a sophisticated theory of mind is not required to recognise that another individual suffers. Indeed, it is likely that animals are capable of recognising suffering in others because they possess two relevant capacities: a physical sense of self that allows animals to be aware that others are separated from themselves and some capacities that enable them to recognise that others are suffering on the basis of behavioural features.[241] Whether animals recognise others' emotions by directly *perceiving* them or by *inferring* them from bodily cues is of no importance here. As mentioned in my thesis introduction, I do not need to settle that complex debate in regard to animal consciousness, which I leave to philosophers of mind.[242] It is sufficient to claim that animals possess the capacity to recognise basic emotions in others.[243] Such emotions include suffering, joy, and anger, to name a few.

---

[241] On animals' physical self-awareness, see DeGrazia (2009), Lage, Wolmarans & Mograbi (2022) and Lei (2023).
[242] For an overview of the debate, see Andrews (2014), pp. 82-84.
[243] I here follow a trend in the animal morality debate according to which animals do not need to possess sophisticated mind-reading capacities or to represent emotions *as* emotions to recognise emotions in other

In the philosophy of emotion, it is widely agreed that we can recognise emotions in other individuals on the basis of some behavioural cues that are closely associated with them, though these may not be essential to emotions and are not identical to them.[244] As Joel Smith notes, smiling is often a manifestation of joy. The latter is an emotion that involves a tendency to cause smiling, and from seeing someone smile, we can recognise that she is in a state of joy.[245] This view is also shared to some extent by philosophers like Quassim Cassam, Rowland Stout, and Joel Krueger, who argue we can directly *perceive* emotional states or *infer* them on the basis of bodily features, depending on what model of emotion recognition one adopts.[246] In short, it seems that we can recognise anger or joy by relying on their expressions, and these are evidence of others' mental states.[247]

The same thing can be said regarding animal empathy and the recognition of suffering in others: from seeing someone crying or screaming, from smelling stress pheromones, or from touching another individual's tense body, an animal can *perceive* or *infer* that the individual is suffering.[248] If animals were merely reacting to superficial cues without recognising them as "belonging" to someone suffering, they would probably not act to alleviate the individual's plight. They would prefer to escape the situation and avoid these aversive cues, and this reaction has been little observed in animals so far.[249] For example, when given the choice to help a drowning conspecific or escape the situation to eat chocolate chips, most rats chose to

individuals. It is sufficient that they discriminate between emotions by relying on bodily features. See, for instance, Monsó (2015 and 2017).

[244] For an overview of the debate on bodily changes and emotions, see Solomon (1973), pp. 23-24, Krueger (2012), pp. 160-161, Roberts (2013), p. 115, and Goldie (2007), p. 931.

[245] Smith (2018), pp. 146-147. For a similar view applied to anger, see McNeil (2012).

[246] Cassam (2017), Chapter 5, Stout (2010), and Krueger (2012).

[247] It should be said that these authors do not only argue that we can recognise emotions in bodily features, but that we can *know* the mental states of others by either directly perceiving them in these bodily features or by inference. They thus attempt to solve the problem of other minds. I am not committed to a stronger view here.

[248] For a philosophical discussion of how animals gather information about others' emotional states through touch, see Monsó & Wrage (2020).

[249] For a similar view applied to human beings, see Nichols (2001), p. 436, who extensively discusses Daniel Batson's work on altruistic motivation. See Batson (1991).

help the distressed cage mate.[250] This allows me to claim that animals react to other individuals in a state of suffering, not mere superficial cues of suffering, and that a full-fledged theory of mind does not seem to be required to recognise that others are suffering. More basic capacities for emotion recognition are likely sufficient.

Second, defending a demanding view of empathy's intentionality may lead us to deny that infants who do not possess a sophisticated theory of mind can be moral agents, even though they clearly seem to feel empathy when hugging or touching other children in a state of distress.[251] This comforting behaviour suggests that a full theory of mind, which emerges at the age of three but is well-acquired only by five or six years of age, may not be required to feel empathy and to be moved to act to alleviate others' suffering.[252] Nevertheless, a full-fledged theory of mind, along with some awareness of causal relationships, provides children with a deeper understanding of others' mental states toward the end of the preschool period. Three- and four-year-old children understand that an emotion like sadness involves a loss or failure to obtain something that is desired, and thus grasp the relationship between others' desires, beliefs, some states of the world, and sadness.[253] However, such a complex understanding does not seem to be necessary for empathy.

Third, even if we accept that a sophisticated theory of mind is required for an individual to feel empathy, this objection cannot exclude all animals from the realm of beings capable of feeling empathy. Indeed, great apes may possess such cognitive capacities, though the empirical evidence is still debated.[254] At least, studies indicate that it might be too early to

[250] Sato, Tan, Tate, & Okada, (2015).
[251] For an overview of the development of empathy in children, see Hoffman (2000), p. 6.
[252] Harris & Saarni (1989), p. 7 and p. 11.
[253] *Ibid.*, p. 7.
[254] For an overview of the debate, see de Waal (2006a), p. 70 and O' Connell (1995). For more recent reviews and, see Call & Tomasello (2008), Andrews (2013), and Royka & Santos (2022).

entirely deny some mind-reading capacities and a basic theory of mind to some animals like great apes, especially chimpanzees. Hence, the objection is likely to fail on both sides. On the one hand, a theory of mind may not be required to be aware that another individual suffers. On the other, some animals may possess it and could be included in this more demanding account of empathy's intentionality, depending on the threshold of mind-reading capacities animals ought to meet to be described as empathetic and whether such account of empathy could resist similar objections to the ones I raised in Chapter 1 toward threshold views of moral agency.

Yet, I think this objection captures some important limitations of animal empathy. Because animals lack a deeper understanding of others' mental states, they are vulnerable to false beliefs about the suffering of others.[255] To illustrate this, let us consider two examples which will be familiar to anyone who has lived with a dog:

> *The Swim:* Two siblings are swimming and having fun in a lake on a lovely summer day. They shout joyfully while playing in the water. Hearing children's high-pitched screams, the family dog jumps into the lake and rushes to rescue the siblings.

> *The Prank*: Valerie decides to prank her dog by pretending to feel sad just because she is curious to see how her dog will react. She starts to make loud crying sounds and sob while burying her face in her hands. As soon as her dog hears her cry, he tries to comfort Valerie by putting his head on her lap and licking her face and hands.

[255] This limitation is briefly mentioned by Angus Ross but not thoroughly addressed. See Ross (1983), pp. 204-218.

These examples illustrate two ideas that are partly captured by the objection on the possession of a full-fledged theory of mind. In the first case, the family dog wrongly believes that the children are in a state of distress. Without a deeper understanding of the children's emotional states, the detection of suffering is imprecise and relies on cues and bodily features that *can be* associated with suffering, like screams. But these screams can also be the manifestation of other emotions, for instance, joy and surprise. Hence, animals may be more vulnerable to false beliefs about the suffering of others because their capacity to discriminate emotions on the basis of bodily features is less sophisticated than that of human beings. It is not supported by more contextual information about others' mental states.

In the second case, the detection process is not flawed. The dog correctly associates behavioural cues like crying and sobbing with the belief that Valerie is suffering. However, he lacks a sophisticated understanding of Valerie's mental states. This prevents him from being aware that Valerie's suffering is not caused by some unfortunate event and is only a simulacrum. In normal circumstances, Valerie's expression of suffering would have indicated that she is suffering, but her dog lacks knowledge of the circumstances in which emotions arise. Without a sophisticated theory of mind, he cannot rule out the possibility that Valerie is not suffering, for he simply does not know about the causal relationships between states of the world and emotions.

Hence, most animals could be compared to very young children in terms of their capacity to detect and understand others' emotions. Although young children can feel empathy, they often make detection and discrimination mistakes until the age of 6, and these errors could be explained by the fact that very young children lack a full-fledged theory of mind.[256] For example, studies suggest that three-month-old infants can distinguish happy from

[256] On that point, see Bischof-Köhler (1991), p. 245.

surprised faces and that children's capacity to discriminate positive and negative expressions is well-acquired by the age of five or six months.[257] At the age of two years, children increasingly understand behaviour such as crying and laughing and can label such expressions as being related to negative or positive emotions.[258] Nevertheless, it is not before the age of four years that children can use various concepts such as happy, sad, mad, angry, and scared and more or less correctly apply them to facial expressions.[259] Such evidence indicates that most animals could be compared to toddlers in their capacities to recognise facial expressions and detect others' emotions.

These limitations apply to a lesser degree to animals like great apes, who seem to show a much greater awareness of causal relationships between others' mental states and the world. They are also capable of targeted helping, which is a form of help that is directed at the precise situations of others.[260] To illustrate this idea, we can refer to the case of Kuni, a female bonobo who once tried to help an unconscious bird that had fallen in her zoo enclosure. Frans de Waal sums up the anecdote as follows:

> The event concerns Kuni, who had found a stunned bird that had hit the glass wall of her zoo enclosure. Kuni took the bird up to the highest point of a tree to set it free. She spread its wings as if it were a little airplane, and sent it out into the air, thus showing a helping action geared to the needs of a bird. Obviously, such helping would not have worked for another bonobo, but for a bird it seemed perfectly appropriate. Kuni's reaction was probably based on what she knew about birds, seeing them fly by every day.[261]

---

[257] For an overview of the literature, see Michalson & Lewis (1985), p. 121.
[258] *Ibid.,* p. 122.
[259] *Ibid.*
[260] de Waal (2009), p. 92.
[261] *Ibid.*

This story suggests that Kuni knew that flying was part of the bird's normal behaviour and that something was wrong with the bird.[262] It seems like Kuni could understand the bird's plight beyond the mere expression of suffering, unlike most empathetic animals. This could be explained by the fact that great apes might be capable of more sophisticated mind-reading and imaginative capacities.[263] That said, I argue that these cognitive capacities are not necessary for an individual to feel empathy that is epistemically relevant to moral agency.

A second objection can be made against my minimal account: empathy requires imagination, a capacity that enables individuals to take the perspective of others and gain some deeper insight into their emotional states. For instance, proponents of the simulation theory of mind-reading argue that imaginative projection is required to understand the mental states of others. To feel empathy, we must imagine what it is like to suffer, put ourselves in the shoes of the sufferer, and adopt their perspective.[264] In the same vein, philosophers like Mitchell Green, Alex Neill, and Berys Gaut argue that we must imagine how an emotion or experience feels to feel empathy. Simply recognising that someone is suffering is not sufficient.[265] Similarly, according to Amy Coplan, empathy needs a complex process in which "one constructs another's subjective experience by simulating the experience of being in the other's situation."[266] Finally, Martha Nussbaum argues that empathy requires a reconstruction of another's experience grounded in imagination.[267] Hence, we may wonder if "empathy" is the right notion to describe animals' response to others' suffering and if animals are capable of

[262] *Ibid.*
[263] On great apes' capacity for imagination, see Mitchell (2002) and Suddendorf & Whiten (2007).
[264] For an overview, see Goldman (1992), p. 21, Gallagher (2012), and Kögler & Stueber (2000), pp. 7-12.
[265] Green (2007), pp. 187-190, Neill (1996), and Gaut (1999).
[266] Coplan (2011), p. 9. Also cited in Simmons (2013), pp. 98-99.
[267] Nussbaum (2001), p. 302 and p. 328. See also Nussbaum (2013), p. 145.

such imaginative displacement. If the answer to the latter question is negative, we must search for another concept to describe animals' emotional reactions and altruistic behaviour.

It is true that imagination greatly influences our experience of empathy, for example, by making others' emotions more vivid to us. It also underpins cognitive empathy. However, this alternative account is vulnerable to the same problems that the objection on the possession of a full-fledged theory of mind faces. First, it may lead to the exclusion of very young children from the category of moral agents, as the literature on the cognitive development of children suggests that it is not before the age of three that children can put aside their own stance and imagine others' perspectives.[268] Second, just as it is possible that some animals possess a sophisticated theory of mind, numerous animals may possess some imaginative capacities that could, in part, foster empathy. Indeed, animals could be aware of others' perspectives, such as when they send dishonest signals to other individuals, engage in playful pretence, and mimic other individuals.[269] Although such behaviour could be evolutionarily selected and consistent with other explanations like learning to "fool" other individuals because it increases animals' chances of survival or access to resources, it still seems to require some minimal awareness of other individuals as agents with various mental states.

Moreover, some studies on rats seem to show that these animals are capable of imagination and perspective-taking. Preliminary evidence from a study conducted by Nobuya Sato and his colleagues indicates that rats who had to free a distressed conspecific were more likely to help and did it faster when they had been subjected to the same unpleasant experience in the past.[270] The scientists have thus proposed that the rats may be capable of "experience projection," a form of sensitivity to the painful experience of others that is mediated by similar

[268] Goldman (1992), p. 26.
[269] Peña-Guzmán (2022), pp. 135-136.
[270] Sato, Tan, Tate, & Okada, (2015).

painful past experiences and which was at work in the role-reversal experiments.[271] Studies on rats' cognitive maps also suggest that rats can replay spatial sequences that they have never experienced or that do not exist and can even project themselves in new spatial sequences.[272] This capacity enables them to experience the world from a different perspective. As philosopher David Peña-Guzman notes, the capacity for self-projection and perspective-taking can be related to empathy, though it has been very little studied in situations involving suffering in other rats.[273] Such evidence indicates that it might be too early to conclude that some animals cannot take the perspective of others and show imagination. Unless one could defend the idea that these imaginative capacities must be complex and meet a certain threshold of sophistication, this objection does not stand on firm grounds when it is used to categorically deny that animals cannot feel empathy. Furthermore, imaginative capacities may not even be required to recognise suffering in another individual and have an emotion whose intentional object is another individual in a state of suffering.

To sum up, I argue that any animal with a certain physical sense of self and others, who can detect behavioural cues of suffering in others and who can recognise suffering in another individual is suffering, is capable of empathy that is epistemically relevant to moral agency. These capacities are sufficient to account for the intentionality of empathy. Although the intentionality of empathy allows different degrees of cognitive sophistication, I have argued that higher cognitive capacities, such as a full-fledged theory of mind and imagination, are not required for an individual to feel empathy that is directed at another individual in a

[271] *Ibid.* For comments, see also Andrews & Monsó (2022), pp. 393-394.
[272] See Foster & Wilson (2006), p. 680, Davidson, Kloosterman & Wilson (2009), p. 504, Karlsson & Frank (2009), p. 2, and Gupta, van der Meer, Touretzky & Redish (2010). (2010), pp. 695–96. Cited in Peña-Guzmán (2022), pp. 140-143.
[273] Peña-Guzmán (2022), p. 143.

state of suffering. Having a self-other distinction, being capable of detecting manifestations of suffering in others and recognising that someone is in a state of suffering are sufficient.

2.2.2. The empathetic animal suffers in tune with the sufferer.

As already noted, philosophers agree that most emotions involve feelings. Emotions are felt as pleasant or unpleasant and come with various bodily changes. For example, when we feel afraid, our heart beats faster, and our muscles are tense. Overall, fear is felt as unpleasant. Furthermore, as philosopher Peter Goldie points out, these feelings are part of an emotion that has an intentional object. They are "feeling towards" and involve "extraspective emotional engagement with the world beyond the body."[274] We do not feel afraid merely because we notice bodily changes in our body. We feel afraid *at* the sight of a thing that we evaluate as dangerous.

The same thing can be said of empathy, which often comes with a feeling similar to the one felt by the sufferer. When we feel empathy directed at others' suffering, we often suffer in tune with the sufferer. This feeling does not need to be identical to the suffering of others, but it needs to be responsive to others' suffering and changes in others' emotional states. If we see someone in excruciating physical pain, we do not need to feel the exact same form of suffering, though we may feel distressed and psychologically suffer. Hence, empathy often entails suffering in tune with someone else, and this may partly distinguish empathy from emotions like compassion or sympathy. The latter is often said to be colder and more

[274] Goldie (2004), p. 91 and p. 96.

detached, as it involves feeling *sorry for* someone else instead of "catching" the emotions of others.[275]

Making this similar feeling an essential component of empathy also allows us to account for the link between empathy and helpful behaviour but without claiming that such behaviour partly explains its epistemic role. Although I shall not address the precise structure of motivation and its link to emotions, most philosophers working on emotions agree that the felt dimension of emotions is "pushy."[276] It seems to bear a relationship with other mental states like desires and may motivate us to act. In the case of empathy, it can lead us to act altruistically.[277] Indeed, without that similar feeling directed at the suffering of others, an individual can easily witness suffering in others and use others' vulnerability to manipulate them or even take pleasure in seeing them suffer.[278] Giving such an important place to the felt dimension of empathy allows us to preserve a certain link between empathy and helpful behaviour, but without claiming that such behaviour entirely explains empathy's importance in our moral lives.

Some philosophers have taken empathy's motivational outcome to partly explain its relevance to moral agency, including its epistemic role. For example, Susana Monsó has developed an account of animal empathy that rests on the helpful behaviour resulting from empathy's unpleasant felt dimension. Monsó sums up her account as follows:

---

[275] See, for instance, Solomon (2008), p. 65 and p. 68, Lazarus (1991), pp. 287-289, Simmons (2013), pp. 98-99, Nichols (2001), p. 431, Prinz (2009), pp. 532-533, Malti & Ongley (2014), p. 166, Tangney, Stuewig & Mashek (2007), pp. 362-363, and Crisp (2008), p. 234.
[276] Zagzebski (2003), p. 116 and Roberts (2013), p. 116.
[277] Eisenberg & Miller (1987).
[278] This claim is also supported by the literature on psychopathy. As already noted in this chapter, most researchers agree that primary psychopaths can understand that others are suffering and are capable of cognitive empathy, but they do not seem to emotionally resonate with others. On empathy defects in some psychopaths, see Marsh (2014).

> Creature C possesses minimal moral empathy (MME) if: it has (1) an ability to detect distress behaviour that, (2) due to the action of a reliable mechanism, results in an emotion that is directed towards the distress behaviour, and built into which is (3) an urge to change the situation that, together with the emotional reaction, (4) tracks a relevant moral proposition.[279]

Monsó takes her account to be sufficient, but not necessary, for minimal moral empathy (MME). To put it differently, other forms of MME might be plausible, in Monsó's view. Here, I want to develop an alternative account to hers that also builds on the strengths of her definition. Monsó's proposal aims to explain the moral relevance of emotional contagion and does not presuppose any sophisticated mind-reading capacities. I am highly sympathetic to that goal. However, I think her account is both too minimal and not minimal enough.

On one hand, it is too minimal in regard to its intentional object. Indeed, it is hard to see how empathy can have distress *behaviour* as an intentional object and not an individual in a state of distress. Young children and animals capable of emotional contagion already possess a minimal self-other distinction and capacities for emotion recognition that allow them to feel empathy toward an individual in a state of suffering, and not just suffering *cues* or suffering as a *mental state*. Moreover, it is hard to see how distress behaviour renders empathy moral, since distress behaviour is not a morally relevant feature in itself. As noted by Monsó, who anticipates that objection, distress behaviour is rather a marker of a morally relevant feature – suffering.[280] Monsó takes it to be a reliable marker. However, it is not always the case, as illustrated by examples such as *The Swim* and *The Prank*.

[279] Monsó (2015), p. 681.
[280] *Ibid.*, p. 687.

To more directly and accurately account for animals' experience of empathy and its epistemic role, I suggest that the intentional object of animals' empathy be a different one. If animals can recognise that another individual suffers, it follows that the intentional object of their empathy is not someone's behavioural cues but another individual in a state of suffering. That option also enables us to explain why animals' empathetic reactions can easily be mistaken: they do not possess a full-fledged theory of mind or sophisticated mind-reading capacities that would allow them to have a more fine-tuned understanding of others' emotions.

On the other hand, I think her account is not minimal enough in regard to its motivational outcome. Indeed, the third criterion could potentially be too demanding, for it seems that empathy's epistemic link to moral agency cannot partly depend on the urge to change the situation, for instance, to help the sufferer. As philosopher Lawrence Blum notes, there are several counterexamples to the view that the importance of other-directed emotions rests on the motivation to which they give rise.[281] To illustrate the idea, let us examine the following example:

> *The Reactive Twin:* Emma and Liam, two four-month-old twins, are kept in the same bedroom. Emma suffers from recurrent ear infections and cries frequently. When Liam sees and hears his sister crying, he always watches her for a few minutes, feels increasingly uncomfortable, and starts crying in turn.

In this case, Liam possesses an implicit self-other distinction and the discrimination capacities that allow him to recognise emotions in others on the basis of behavioural cues.

[281] Lawrence Blum discusses the case of compassion in *Moral Perception and Particularity* (1994), but his argument can also be applied to my account of empathy. See Blum (1994), pp. 180-192.

That said, it is unclear that Liam feels the urge to help his sister here, as helpful behaviour and altruistic motivation have not been observed in infants younger than six months.[282] Indeed, six-month-old children sometimes react to the cries of another baby by gesturing, leaning toward the other child, or touching the baby. Their other-oriented behaviour suggests that they might have the desire to help distressed others. However, this behaviour has not been observed in younger children like Liam. Although Liam could be motivated to help without actually helping, altruistic behaviour serves as the main marker for studying altruistic motivation in children. Without such evidence, it is hard to argue that infants as young as four months can be motivated to help others in distress.

Hence, Liam's case indicates that simply catching others' emotions without being moved to help could already be relevant to moral agency in the epistemic sense. That said, the urge to help others may interact with emotional contagion, and it may be hard to disentangle the felt dimension of empathy and its motivational outcome in most moral agents. That motivation may also show that the empathetic individual who acts to help the sufferer evaluates that suffering as "something to be alleviated" or as "something bad" and hence is capable of a cognitively more sophisticated form of empathy. Monsó's third criterion rightfully accounts for empathy's plausible *self-control* role and the relationship between empathy, motivation to help, and moral action. Yet I argue that this urge is not needed to account for the *epistemic* role that empathy plays in our moral lives, with which Monsó is also concerned.

To respond to that objection, Monsó can either propose another criterion or simply put it aside. The third criterion could be replaced by the *evaluation* that distress should be alleviated rather than resting on an urge to change the situation. This would make her account more minimal in terms of motivational outcome but would presuppose some basic evaluative

[282] Hay, Nash & Pedersen (1981).

capacities, thus rendering it more demanding in other respects. I suggest that we simply put the third criterion aside to include individuals who may not possess the required cognitive capacities, such as an understanding of causality, a more robust self-other distinction, and a minimal sense of the future, to have the desire to see the sufferer's distress alleviated or to be able to evaluate suffering as something to be alleviated. These individuals may already have access to the badness of suffering in others. I shall expand on this claim in the following section.

Before turning to the last part of my chapter, I want to address one last possible objection to my account. One could argue here that my argument leaves the door open to recognising helpful but selfish behaviour stemming from empathy as still relevant to the exercise of moral agency. Indeed, there are many ways in which one can react to the suffering of others: by merely suffering in tune with another individual, by having the desire to help, by altruistically helping the sufferer, or by helping the sufferer for selfish reasons. For instance, one could act to alleviate the sufferer's distress to reduce one's own suffering, not for the sake of the sufferer. Because I deny that altruistic motivation grounds or partly grounds the importance of empathy, I must acknowledge that non-altruistic forms of empathy have a role to play in our moral lives. This is a consequence that some philosophers may find problematic, especially philosophers who might want to propose an account of empathy's epistemic role that would be grounded in its altruistic motivational outcome.

Two things are worth noting here. First, this objection also applies to Monsó, for having the desire to change the situation does not always imply that the underlying motivation of that urge is altruistic. Moral agents' behaviour can be other-directed without being fostered by an altruistic motive. Second, one response to that objection consists of disentangling the felt dimension of empathy and its motivational outcome and of appealing to the fact that

suffering at the sight of others' suffering can be a fitting or appropriate reaction to its object, even though a resulting selfish motivation is not. In other words, emotional contagion is fitting when there is a match between emotional contagion, which allows individuals to have access to the badness of others' suffering, and the property of others' suffering – its badness. In the two cases of misguided empathy I proposed earlier, *The Swim* and *The Prank,* the dog's empathy is not fitting. What the dog mistakenly perceives as suffering and feels as aversive is not really suffering and, hence, cannot be bad.[283] Empathy can also be said to be fitting if it is of the right size, if it is not an overreaction.[284] I shall come back in more detail in the next section to what I mean by "having access to the badness of others' suffering."

To sum up, suffering in tune with the sufferer already has an epistemic role to play, whether it leads to an altruistic motivation to help the sufferer or not. It is an essential component of my account of empathy's importance. The pushy dimension of this similar feeling also allows me to preserve the link between emotions and altruistic behaviour, but without being committed to more precise views about that relationship or to stronger claims about the role such behaviour plays in explaining empathy's epistemic role. Altruistic behaviour is *not* necessary to explain how empathy can be related to moral agency in the epistemic sense.

2.2.3. This similar feeling allows animals to have access to the badness of others' suffering.

The core idea of the chapter is the following: when an animal suffers in tune with the sufferer, this similar feeling allows her to have access to the badness of others' suffering, even though

[283] For an overview of the literature on the fittingness of emotions, see Naar (2021).
[284] For this distinction between the two types of fittingness, see D'Arms & Jacobson (2000), pp. 73-74.

she cannot entertain the proposition "Suffering is bad." However, before developing this idea more at length, we need to examine the various ways in which emotions or emotional capacities can be related to moral properties such as *good, bad, right,* and *wrong*. Empathy directed at the suffering of others seems to stand in a relationship with the notion of "badness" in multiple ways, and not only because of the similar feeling it involves. I shall mention two of them here.

One way in which empathy can be linked to badness is through its evaluative component. As noted in §2.1.2, empathy can entail various evaluations pertaining to suffering and the sufferer's value. This evaluative dimension is a common characteristic of empathy and emotions. Indeed, it is widely agreed in the philosophy of emotion that emotions come with evaluations and stand in a relationship with values, though philosophers do not agree on how we should define emotions and how they instantiate values. Philosophers like Martha Nussbaum and Robert Solomon argue that emotions are judgements, Christine Tappolet claims that emotions are akin to perceptions of value, Richard Lazarus proposes a definition of emotions as a set of appraisals about one's environment, and Jesse Prinz argues that emotions are embodied appraisals. Yet almost all philosophers working on emotions are committed to evaluationism, which is the view that emotions come with evaluations and a certain way of instantiating their object.[285] For example, feeling fear involves evaluating the intentional object of our fear as *dangerous*, anger entails an evaluation of its object as *offensive*, guilt comes with the evaluation of a past action as *wrong,* and so on. Similarly, empathy may come with various evaluations of others' suffering.

[285] For an overview of the different theories of emotions, see Solomon (1999), p. 12, Naar (2019b), pp. 469-470, Roberts (2009), pp. 568-572, Debes (2009), p. 3 and p. 6, and Mulligan (2009b), p. 482.

Human beings seem to be the only animals who can *directly* evaluate the suffering of others as *bad*. Furthermore, unlike other animals, human beings can understand *why* suffering is bad: because it negatively affects the well-being of others, because it is undeserved, etc. This is not to say that other animals are incapable of evaluating others' suffering negatively. Animals make all sorts of evaluations about their environment. For example, a deer can evaluate a predator as something to be avoided or berries as things to be eaten.[286] To use Christine Korsgaard's words, animals' responses to various objects of their environment can be "primitively normative."[287] When animals feel emotions, they already stand "in the presence of a normative fact", such as the dangerousness of danger when they feel fear, according to Korsgaard.[288] When feeling empathy, animals can evaluate the suffering of others negatively or even as something that should be alleviated. This evaluative dimension may partly explain, besides the felt and "pushy" dimension of empathy, why many empathetic animals act to alleviate the suffering of others. Moreover, animals may positively evaluate some individuals who suffer, as confirmed by their tendency to help individuals who are important to them and have some emotional significance in their lives. We can think of in-group members in the case of wild animals or members of their household in the case of domesticated animals.[289]

One way of accounting for the importance of animal empathy and its relationship with evaluative properties such as badness would be to argue that a negative evaluation of suffering, or the evaluation of suffering as something to be alleviated, tracks the property of *badness* even though most animals do not directly possess this concept. An animal's evaluation of suffering as something to be alleviated could track the concept of badness because the evaluation of

---

[286] Korsgaard (2009), p. 109.
[287] *Ibid.,* p. 111.
[288] *Ibid.,* p. 112.
[289] This important limitation of animal empathy is discussed at length by Peter Singer in *Primates and Philosophers* (2006), as we have seen in Chapter 1. See Singer (2006).

suffering as something to be alleviated entails that suffering is bad, even though most animals cannot use the notion of badness. As we have seen, Mark Rowlands claims that a proposition p can track another proposition p* if and only if the truth of p guarantees the truth of p* because there is a reliable and asymmetrical link between the concepts in the two propositions.[290] If we apply Rowlands' strategy to the case of empathy and animals' evaluations, we could perhaps claim that the truth of p, "Others' suffering is something to be alleviated", guarantees the truth of p*, "Others' suffering is bad."[291] If this argument is successful, it could allow us to claim that empathy in animals can reliably track the concept of badness. Highlighting the different evaluations that can be part of empathy also allows us to flesh out the various degrees of sophistication that empathy admits in human and nonhuman animals.

However, I shall put aside this strategy in this chapter for three reasons. First, arguments on animal concepts have already been offered elsewhere, including in the debate on animal morality, as I have just noted.[292] Second, grounding empathy's epistemic role in its evaluative component runs the risk of overlooking the importance of more basic forms of empathy, like emotional contagion. Indeed, it is widely agreed that emotional contagion often interacts with our natural capacity for mimicry, a capacity that is already present at birth[293], although mimicry alone is not sufficient to *feel* others' emotions. According to the psychology literature, emotional contagion does not seem to depend on evaluations, even implicit and

[290] Rowlands (2012), pp. 58-63.
[291] It is worth noting that this tracking strategy allows Rowlands to answer philosopher Beth Dixon's argument according to which animals cannot feel "morally thick" emotions because their emotions do not include the following three judgements: (1) the judgement that the suffering of others is serious, (2) the judgement that the sufferer's distress is undeserved, and (3) the judgement that others share a similar vulnerability to suffering. I shall not address these arguments here, as Beth Dixon is concerned about how we ought to define morally thick emotions but does not offer a more comprehensive picture of moral agency. For more details, see Dixon (2008a), pp. 66-68.
[292] See Allen (1998), p. 66, Searle (1994), and Rowlands (2012), Chapter 2.
[293] Goldman (1992), p. 29.

quick ones, and is an automatic process.[294] Third, even though young infants like Liam, the reactive twin, might be able to form evaluations about others' suffering and evaluate others' suffering as "aversive" or as "something to be alleviated," these evaluations are hard to study in pre-verbal children. To my knowledge, we do not have studies pertaining to four-month-old infants' evaluations of others' suffering.

Another strategy to account for the relevance of animal empathy was put forward by Mark Rowlands and Susana Monsó. It consists of grounding the link between empathy and moral agency in its phenomenal character, which tracks the proposition "Suffering is bad." According to Rowlands and Monsó, even though animals cannot form moral judgements and directly describe others' suffering as bad, their aversion for suffering and their motivation to help others allow us to argue that the proposition "This creature's distress is bad" is present in their empathy.[295] Indeed, there seems to be what Rowlands and Monsó call a "truth-preserving relation" between an animal's empathy and the moral proposition that "Suffering is bad," in the sense that if an animal feels empathy and is not misguided in feeling empathy, then the moral proposition must be true.[296] Hence, animals' empathy can be moral because it "has the adequate phenomenal character" and tracks a moral proposition.[297]

I think this account is a promising one, especially if we put aside Monsó's third criterion on empathy's motivational outcome. Yet, I think more could be said to account for how the *felt* dimension of empathy can allow animals to have access to some moral properties. My answer to this question is that the relationship between empathy in animals and morality is provided by two of its components: its intentional object and its felt dimension. In this

[294] See Monsó (2015), p. 680, Preston & de Waal (2002), p. 1, Bekoff & Pierce (2009), p. 89, Jordan, Amir & Bloom (2016), p. 1107, Prinz (2011), p. 212, and de Waal (2008), pp. 282-283.
[295] Rowlands (2012), pp. 68-69 and Monsó (2017), p. 351.
[296] Monsó (2017), p. 351.
[297] *Ibid.*

section, I want to argue in favour of three claims: (1) empathy's intentional object is a morally relevant feature, (2) the similar feeling that empathy entails allows animals to access the key properties that explain suffering's badness, and (3) that feeling is directed at another individual in a condition of suffering, which enables animals to have access to the badness of *others' suffering*, not just of their own suffering. Let us explore these three elements.

First, what partly explains the importance of empathy in our moral lives is that its intentional object is another individual in a state of suffering, which is a *morally relevant* feature. This has already been put forward by Susana Monsó, who rightly points out that most moral theories agree that distress is a morally relevant feature. Monsó writes: "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."[298] However, feeling distressed at the sight of non-morally relevant things is not relevant to moral agency. For example, Monsó notes that even though dogs can experience the noise of a vacuum cleaner as distressing, that emotional state is not a moral one because "vacuum cleaning noises are not morally relevant features of situations […]."[299] Hence, what partly explains empathy's epistemic role is its intentional object, which, in my account, is another individual in a condition of suffering and not just distress behaviour.

Second, suffering is generally bad for the person who experiences it. Indeed, most moral philosophers agree that suffering is *intrinsically* bad for the individual who suffers, although it could be *instrumentally* good in some circumstances.[300] It seems that suffering can sometimes lead to a greater good, for example, when someone undergoes painful surgery for

[298] *Ibid.*
[299] *Ibid.*
[300] Kahane (2016), p. 210 and Kauppinen (2020), p. 21.

a new organ, or that it can play a positive transformative role in one's life.[301] Furthermore, philosophers generally agree on the following idea: suffering is bad because *it feels bad*, it is negative well-being.[302] Suffering is highly *unpleasant*. We can suffer because of various things, but all forms of suffering are aversive. Philosophers explain suffering's unpleasantness in many ways, but defending one precise account is not required for my argument. Suffering could be unpleasant because it has a negative hedonic tone, it is disliked, it is undesired, it thwarts preferences, or because the sufferer is disposed to act to alleviate it.[303] I shall leave this question open here.

It thus seems that any sentient creature can have access to the badness of suffering to some extent simply by suffering. The fact that suffering *feels bad,* is unpleasant*,* is phenomenally bad, is a property that explains the badness of suffering. Hence, the novelty of my account lies in the following idea: animals can access the fact that suffering is bad, either in themselves or others, simply because suffering feels bad. In other words, animals can have access to the badness of suffering by directly experiencing *the key properties that make suffering bad.* They do not possess the concept of "badness" and cannot entertain a proposition like "Suffering is bad," but they can experience what makes suffering bad. Having access to the badness of suffering does not require entertaining a moral proposition or possessing some moral concepts, which are often taken to be important capacities associated with moral agency, as pointed out in Chapter 1. Animals' access to the badness of suffering is not provided by colder and intellectual capacities, like the ones for propositional thought and introspection, but

[301] For a Nietzschean challenge to the badness of suffering, see, for instance, Delon (2022).
[302] Sinhababu (2024) and Andrew (2021), pp. 26-28.
[303] For an overview, see Bain (2013), Kahane (2009), and Brady (2018), pp. 34-47.

simply by their capacity to suffer and thus to directly experience the properties that explain suffering's badness.[304]

Third, we need to stress the importance of the "other-directedness" of empathy to distinguish the suffering that empathy entails from other forms of suffering. Indeed, sentient animals can have access to the badness of their own suffering. For instance, horses forced to pull heavy loads of coal like Trompette and Bataille in *Germinal* can directly experience suffering as bad for themselves. What distinguishes the suffering that accompanies empathy from this form of suffering is that empathy is directed at others in a state of suffering. An empathetic animal like Bataille suffers when *he sees Trompette suffer*. His suffering is *directed* at another individual's suffering and not merely at his own suffering. Suffering in tune with other individuals involves feeling their suffering almost like it was our own, and this thesis is confirmed by studies in human beings that have shown that feeling empathy arouses the same brain regions that are solicited when we feel pain.[305] By the same token, suffering in tune with the sufferer allows human and nonhuman animals to have access to the badness not only of their own suffering, but also that of others.

That said, animals like Bataille cannot understand *why* others' suffering is bad or, conversely, why suffering could be instrumentally good if it leads to greater happiness. Animals' access to suffering's badness is thus limited to its *intrinsic* badness, to how it is bad for the individual who suffers. Seeing someone suffering may feel bad for a nonhuman animal,

[304] For instance, some authors have argued that we can have phenomenological evidence of suffering's badness through experience and introspection. Introspection and reflecting on our own experience of suffering can ground *moral knowledge* about the badness of suffering. However, these accounts have been thoroughly criticized, since introspection is vulnerable to all sorts of distortions. In this chapter, I am not committed to the view that animals can *know* suffering is bad, or that experience of suffering and introspection *ground* moral knowledge, but simply that there is something with the experience of suffering that explains its badness, and that animals can have access to that fact. For an overview of the debate on suffering, introspection, and moral knowledge, see Sinhababu (2024) and Delon (2022).

[305] Panksepp & Panksepp (2013), p. 492.

but that feeling is short-sighted. Animals cannot understand how suffering could be both intrinsically and instrumentally bad, or intrinsically bad but instrumentally good. Animals thus exhibit a much less robust form of moral agency. These limitations also have important implications for animals' access to the wrong-making features of causing suffering. I shall come back to these in Chapter 3.

2.3. Some Remarks on the Notion of "Having Access"

Before closing this chapter, more details on the notion of "having access" are in order here. It might be tempting to see my argument as being committed to a reductionist account of the relationship between suffering's aversiveness and badness. According to such an account, the fact that suffering feels bad would be identical to its badness. Animals would thus have access to the badness of suffering by experiencing its aversiveness. Both dimensions of suffering would be identical. In other words, suffering's badness would be reducible to its aversiveness.

However, such a reductionist picture would be vulnerable to Moorean "open question" arguments, according to which reductionist accounts lead us to conflate the meaning of aversiveness and badness. When asking a question such as "Is it true that X (aversiveness) is Y (bad)?", in which aversiveness and badness would be identical, we would be asking an absurd question, in which X and Y have the same meaning.[306] Moreover, a reductionist account of suffering's aversiveness and badness could be guilty of a naturalistic fallacy and violating Hume's Law, according to which we cannot derive a moral fact from a natural one.[307] Reductionist theories have faced substantial criticisms and are now quite unpopular among philosophers, although some still argue that reductionist accounts might be plausible when it

[306] Moore (1903), p. 21.
[307] For a summary, see Echeverri (2019), p. 551.

comes to explaining *some* moral facts like suffering's badness, which seems to bear a close connection with its aversiveness.[308] I shall not address these new developments here.

My argument is also consistent with a non-reductionist account of the relationship between phenomenal and evaluative properties. For instance, suffering's aversiveness, the fact that it feels bad, could ground, track or entail its badness but without being reducible to it.[309] Another option, put forward by philosophers Adam Lovett and Stefan Riedener, would consist of arguing that moral agents can be in contact with moral values when such values are "manifest" to them.[310] This picture could be applied to my argument as follows: when empathetic animals suffer in tune with others, they come "in contact" with the badness of others' suffering, which is manifest to them through the aversiveness of others' suffering. In sum, several accounts are on the table. If I adopt a non-reductionist theory of moral facts, animals could not be said to have direct access to the badness of others' suffering. Their access would perhaps be similar to a form of indirect access, "tracking" process, or "acquaintance" with the badness of others' suffering.

I do not need to be committed to a specific account of the link between suffering's aversiveness and badness here or between phenomenal and evaluative properties more broadly. I just need to defend the view that emotions can have an epistemic role to play in our moral lives. That general thesis is fairly uncontroversial in the philosophy of emotion and moral philosophy, although philosophers do not agree on the precise epistemic role that emotions play and to what extent they are needed for us to access moral facts.[311] In the same vein, the argument developed in this chapter should not be conflated with animals' possession

[308] See, for instance, Lee (forthcoming).
[309] For a summary of these different options, see Echeverri (2019), p. 551.
[310] Lovett & Riedener (2024). I am grateful to Jonathan Birch for pointing out this option to me.
[311] For an overview, see De Sousa (2001).

of concepts, as already noted in my introduction. It rather concerns animals' access to moral facts.

2.4. Some Additional Thoughts on Emotional Contagion in Infants

Before turning up to my third chapter, I shall say a few things about empathy in infants. As I already noted in my introduction, my account opens the door to including emotional contagion as a form of empathy that is relevant to moral agency. However, some authors have been reluctant to include emotional contagion within the realm of empathy and to recognise that this phenomenon has a role to play in our moral lives. Several philosophers have claimed that emotional contagion, often described as a form of personal distress, occurs when individuals lack a strong sense of self and a self-other distinction. For example, Elisa Aaltola writes that emotional contagion is triggered when we "unreflectively confuse the emotions of others with those of our own."[312] Similarly, Martha Nussbaum writes in *Political Emotions* (2013) that empathy should not be confused with emotional contagion because the former "requires entering into the predicament of another and this, in turn, requires some type of distinction between self and other, and a type of imaginative displacement."[313] Julien Deonna claims that what distinguishes empathy from emotional contagion is that empathy involves a recognition that another individual is going through an emotion that is distinct from the observer's emotional state.[314] Robert Solomon also claims that empathy requires some minimal separation of the self from others and that infants "cannot feel empathy although they can get very upset when their mothers are upset."[315] Finally, Lawrence Blum writes that empathy has "another's

[312] Aaltola (2018), p. 82.
[313] Nussbaum (2013), p. 145.
[314] Deonna (2007), p. 111.
[315] Solomon (2008). p. 68.

state of mind or situation" as an object, unlike emotional contagion. The latter is simply "morally indifferent", according to Blum.[316] To sum up, emotional contagion is self-absorbed. Without a clear distinction between others' emotions and ours, we cannot take another individual in a state of suffering as an intentional object of our empathy and suffer in tune with that individual.

However, the evidence that is often cited in support of these theses has faced serious criticisms over the last decades, which also casts doubt on the robustness of Hoffman's theory of empathy's development I mentioned in my first chapter, §1.3.1. As Maayan Davidov and her colleagues have noted, even though infants do not possess an explicit self-other distinction and a reflective sense of self, they can have a pre-reflective type of self-knowledge.[317] For example, newborns do not react to their own recorded cries but show signs of self-distress when hearing other babies' cries, which indicates that they may already distinguish their own cries from those of others.[318] This is also confirmed by the fact that they cry when hearing other babies' cries but do not react to other loud noises. They seem to be already responsive to others' suffering.[319]

Furthermore, infants' distress may not be caused by the total absence of capacities for self-other distinction but rather by a lack of emotional regulation. Because babies lack control over their own distress, they may feel overwhelmed when witnessing others' pain, though not immediately, as is the case with Liam, the reactive twin.[320] This is confirmed by study results that show that babies start to cry on average after one to three minutes of being exposed to

316 Blum (2011), p. 172.
317 Davidov, Zahn-Waxler, Roth-Hanania & Knafo (2013), p. 127.
318 See Dondi, Simion & Caltran (1999), cited in Davidov, Zahn-Waxler, Roth-Hanania & Knafo (2013), p. 127.
319 See Simner (1971) and Geangu, Benga, Stahl & Striano (2010), cited in Davidov, Zahn-Waxler, Roth-Hanania & Knafo (2013), p. 127.
320 Davidov, Zahn-Waxler, Roth-Hanania & Knafo (2013), p. 127.

others' cries, probably because it takes children many months to develop cortical inhibitory mechanisms that are essential for regulating their own emotions.[321] They also lack the cognitive capacities to understand what can be done to alleviate others' suffering. As children grow, occurrences of personal distress become scarce, and empathetic helping becomes more frequent.

These studies suggest that very young infants may recognise suffering in others and do not merely react to superficial distress cues, although the capacity to differentiate themselves from others starts to be more robustly acquired between four and nine months of age and is further consolidated during the second year of life.[322] My account of empathy is thus minimal and allows me to acknowledge the epistemic role that emotional contagion plays in our moral lives. If infants who "catch" the distress of others possess an implicit or explicit self-other distinction, can attribute suffering to others and can suffer in tune with others, they can have access to the badness of others' suffering.

However, if infants lack these capacities, their suffering is likely to be self-absorbed, and the criticisms that emotional contagion has faced may be justified to some extent. Emotional contagion that does not come with a certain self-other distinction and an awareness that another individual suffers runs the risk of being merely *caused* by another individual in a state of suffering but not *directed at* that individual. The distinction between the cause and intentional object of an emotion that I introduced in §2.2.1 is helpful here.

Indeed, when newborn infants reactively cry, their distress could be *caused* by other individuals in a condition of suffering but is not *directed* at that individual. Their emotional state does not seem to rest on the awareness that another child is suffering. Hence, because infants'

[321] See Simner (1971), Panksepp & Panksepp (2013), p. 492, and Geangu, Benga, Stahl & Striano (2010), cited in Davidov, Zahn-Waxler, Roth-Hanania & Knafo (2013), p. 128.
[322] Lazarus (1991), p. 307.

distress is not part of an emotional state that takes another individual in a condition of suffering as an intentional object, we cannot say they have access to the badness of others' suffering. Furthermore, this type of emotional contagion, which I prefer to describe as personal distress, cannot be said to be a *fitting reaction* to others' suffering if it does not have others in a state of suffering as an intentional object in the first place. That emotional state may be caused by an aversive stimulus that infants do not understand yet. That reaction may be more akin to a form of personal distress than an emotion with an intentional object.

Nonetheless, this causal relationship may play a role in the formation of more sophisticated forms of empathy. It might enable infants to gradually recognise suffering in another individual and to be properly affected by the suffering of others. Yet that type of emotional contagion's link with moral agency may be more developmental than directly epistemic.

2.5. Summary of Chapter 2 and Concluding Remarks

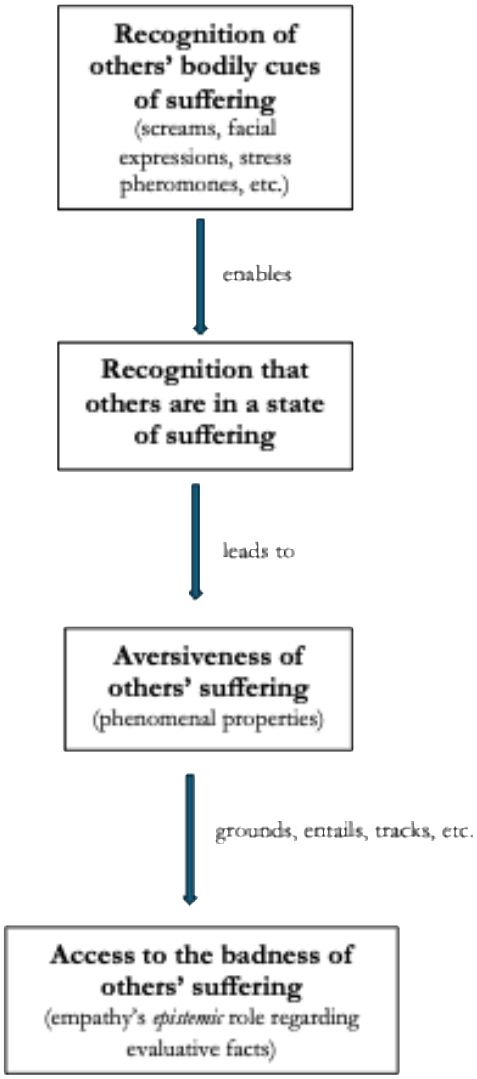
To sum up, I propose that an animal can show a form of empathy that is epistemically relevant if:

(1) The intentional object of her empathy is another individual in a condition of suffering.
(2) She suffers in tune with the sufferer.
(3) This feeling allows her to have access to the badness of others' suffering.

Hence, an animal who suffers at the sight of another individual suffering, as is the case with Bataille, can have access to a moral fact. I have defended this claim by appealing to some

specific components of empathy, namely, its intentional object, the status that suffering has in moral philosophy, and the "other-directedness" of empathy's felt component. The line of argument developed in my second chapter can be illustrated with the following simple diagram:

Summary of Chapter 2



Moreover, I have identified several ways in which empathy admits different degrees of sophistication. I have mentioned some limitations in animal empathy regarding its cognitive base, intentionality, and evaluative dimension. I have noted that animals who do not possess a full-fledged theory of mind have a more superficial understanding of others' emotions, and this renders them vulnerable to detection and discrimination mistakes. I have also highlighted how empathy can entail various evaluations, ranging from a broad negative evaluation to the moral judgement that suffering is bad. These limitations had not been thoroughly examined before in the debate on animal morality.

As we have seen, philosophers like Mark Rowlands and Susana Monsó have been concerned with how empathy can track moral propositions that concern *evaluative* facts – facts about *states of affairs* – but have left out how empathy could partly enable animals to access deontic facts – facts about *actions*. This chapter opens the question of whether empathetic animals can have access to the wrong-making features of causing suffering. Indeed, if animals can have access to the badness of suffering, could they have access to the wrong-making features of intentionally harming another individual? This question will be answered in my next chapter.

Chapter 3

Empathy's Further Epistemic Role: Can Animals Have Access to the Wrong-Making Features of Causing Suffering?

In the previous chapter, I argued that animals can have access to some evaluative facts like the badness of others' suffering. However, most philosophers would agree that moral agency requires not only the capacity to have access to moral facts about states of affairs, but also moral facts about actions. Moral agency is typically associated with the capacity to have access to the wrong-making features of some actions. Furthermore, several philosophers have argued that animals are unaware of what is right or wrong, in part because animals lack moral concepts such as *right* and *wrong*, as seen in Chapter 1.

Could recent findings about animals' capacities challenge this picture? I think so. To see how, let us first look at the following example:

> *The Foxes in Magdalen College:* On November 19th, 2024, England was hit by a snowstorm. Schools were closed, many trains were cancelled or delayed. Largely unaware of the English's collective panic, two foxes were seen delightfully playing together in the main quad of Magdalen College, right in the heart of Oxford. The foxes obviously enjoyed their "rough-and-tumble" play session, biting, jumping on, and chasing each other in the snow and adjusting to each other to continue playing. The scene was filmed by a student and caused a national sensation, capturing hearts.[323]

[323] For the video, see Patrick (2024).

According to threshold views of moral agency, the two foxes playing in Magdalen College could not be said to recognise wrongdoing if their playmate ever bit them too hard.

In this chapter, I cast doubt on this picture of moral agency and argue that many animals (1) can recognise when other individuals intentionally cause suffering and (2) can have access to the badness of others' suffering. These two capacities allow animals to (3) have access to the features that explain why causing suffering is wrong, namely, that another individual causes a bad state of affairs. To defend this thesis, which has not been put forward so far within the animal morality debate[324], I shall first clarify my approach, define concepts such as *wrongdoing* and *responsiveness to norms,* and narrow down the scope of this chapter (§3.1). I will then expand on the three proposed elements (§3.2) and defend my account against three possible objections (§3.3).

It is also worth noting that animals' anger, which could be interpreted as a form of blame, provides additional evidence that animals can have access to the wrong-making features of causing suffering, as many philosophers would agree that access to an action's wrongness often gives rise to moral responsibility practices and reactive attitudes like anger, resentment, and the like.[325] This idea will be further explored in Chapter 4.

3.1. Defining Wrongdoing: Some Clarifications

Before developing my account of animals' access to wrongdoing, three points of clarification are in order.

---

[324] As noted in Chapter 1, Asia Ferrin has argued that empathy enables animals to respond to moral reasons related to others' well-being, although it is unclear what she means by "moral reasons". My third chapter aims to provide an answer to that question. See Ferrin (2019), pp. 143-144.

[325] See, for instance, Adams (1999), p. 238, Baier (1966), Brandt (1979), pp. 163–176, Gibbard (1990), p. 42, Skorupski (1999), p. 142, and Shafer- Landau (2003), all cited in Darwall (2006), p. 27 and p. 92. See also Copp (1997), Smith (1983), Wallace (1994), and Widerker (2000), all cited in McKenna (2012), p. 14. Finally, see Pettigrove & Tanaka (2014), p. 270.

First, I shall concentrate on intentional wrongdoing in this chapter. More precisely, I shall focus my attention on wrongdoing that involves *intentionally* inflicting suffering on another individual. Hence, I shall leave open the question of whether animals could have access to other forms of wrongdoing, which is a plausible hypothesis. For example, one could argue that animals can react to types of wrongdoing akin to promise-breaking when they see that others fail to reciprocate and return favours such as food or grooming.[326] Similarly, many animals, such as most species of primates, elephants, dolphins, crows, ravens, dogs, wolves, cats, mice, rats, and cleaner fish, exhibit a sense of fairness and react strongly when being rewarded with a different treat than other animals for the same effort.[327] Finally, dogs have social preferences for helping individuals, and American crows have been found to "scold" dangerous human beings. Such public mobs enable the birds to share information about harmful individuals with other crows.[328] These fascinating studies raise the question of whether some animals could be said to possess a sense of wrong associated with Kantian ethics or virtue ethics. At first glance, animals seem to ascribe qualities to other individuals based on their actions, which can reflect well or badly on them. These developments are promising avenues for future research. But I shall put them aside here.

Second, although the arguments developed in Chapters 2 and 3 centre on suffering and infliction of suffering, they could also be applied to positively valanced emotional states to account for the goodness of others' joy and the rightness of causing a good state of affairs. For example, the two foxes in Magdalen College playing in delight and sharing each other's

[326] See, for instance, Schino & Aureli (2009).
[327] See, for instance, van Wolkenten, Brosnan & de Waal (2007), Proctor, Williamson, de Waal & Brosnan (2013), Range, Horn, Viranyi & Huber (2009), and Wascher & Bugnyar (2013).
[328] See Chijiwa, Kuroshima, Hori, Anderson & Fujita (2015), Cornell, Marzluff & Pecoraro (2011), and the evidence cited in Zawidski (2013), p. 54.

joy could be said to have access to the goodness of joy and the rightness of maintaining or increasing their playmate's happiness.

Third, I shall avoid the vocabulary of *norms* and *responsiveness to norms* in this chapter and the remainder of this thesis to focus my attention on *wrongdoing* instead. As we have seen in Chapter 1, §1.2.1, some authors who argue that some animals could qualify as moral agents do so by stressing the fact that species like great apes could be aware of norms without possessing a full-fledged theory of mind[329] or by claiming that animals are moral agents only within the context of their communities' norms.[330] In other words, animals could follow social norms, broadly defined as rules of behaviour specifying what actions are allowed or prohibited.[331]

It intuitively makes sense to describe animals' behaviour as responsive to norms, especially when animals play with each other. Play is an activity that is typically bounded by norms. There is also further empirical evidence that animals, especially domesticated animals, can respond to norms. As de Waal points out, dogs will often obey commands in the absence of human beings.[332] The same thing can be said of great apes.[333] Some recent studies suggest that chimpanzees and bonobos can wait for several minutes and repress their immediate desires if it leads to greater future rewards.[334] That said, further research on different species is needed to test animals' capacities to internalise norms, and there may be important variations across animal species. For example, some evidence indicates that once human beings have left the room, rats quickly eat the food they were not supposed to consume in the absence of

[329] See, for instance, Andrews (2009), pp. 433-448, and Fitzpatrick (2020).
[330] Bekoff & Pierce, Jessica (2009), p. 144.
[331] See Bicchieri (2017), p. 35 and Brennan, Eriksson, Goodin & Southwood (2013), pp. 2-3.
[332] de Waal (1996), p. 106.
[333] *Ibid.,* pp. 109-110.
[334] See Rosati, Stevens, Hare & Hauser (2007), pp. 1663-1668.

experimenters. When it comes to following human beings' commands on their own, rats are "a morally bankrupt species," in de Waal's words.[335]

We might have good reasons to think some animals could respond to norms of behaviour. However, notions such as "norms" or "responsiveness to norms" pose a problem, as several philosophers working in the philosophy of social science or legal philosophy would be reluctant to include animals in the realm of agents capable of responding to norms. These philosophers have argued that some more sophisticated cognitive capacities are required for responsiveness to norms. We can think of metacognition or shared intentionality, the latter being the capacity for several individuals to simultaneously hold the same mental representation.[336] According to these authors, simply exhibiting behavioural regularity or merely having the desire that others behave in a certain way is insufficient. Animals could not be said to possess norms or to respond to them.[337]

The argument defended in this third chapter does not get tangled up in heated debates on the definition of norms and the capacities required to respond to them. It rather rests on the simpler but novel notion of *wrongdoing* in animals, which has been largely overlooked by philosophers working on animal morality. The account developed in this third chapter builds on my previous discussion of empathy's epistemic role. To show how empathy could play a role in enabling animals to have access to the wrong-making features of causing suffering, I do not need to defend a thorny notion like responsiveness to norms, or to show how animals' social interactions could be constrained by norms. Being capable of (1) recognising intentional harm and (2) having access to the badness of others' suffering (3) already enables animals to

[335] de Waal (1996), p. 106.
[336] For an overview, see Fitzpatrick (2020), pp. 11-21. See also Birch (2021).
[337] For an overview, see Brennan, Eriksson, Goodin & Southwood (2013), pp. 15-16 and pp. 22-24.

have access to the features that make intentional infliction of suffering wrong. Let us now turn to my argument, which rests on these three simple claims.

3.2. Animals' Access to Wrongdoing: A Minimal Account

3.2.1. Animals can recognise the intentional infliction of suffering.

As previously mentioned, my account is grounded in the premise that animals can recognise intentional actions in others. In this part of my chapter, I am less interested in exploring how the notion of intention should be understood, for instance, as a belief-desire pair that rationalises agents' actions, as a strong desire, or as an evaluative attitude.[338] My goal is rather to explore what animals understand about others' intentions as they manifest in their actions and to defend a minimal account of animals' awareness of others' intentions that also makes room for including infants in the realm of beings capable of having access to the wrong-making features of causing suffering. This section thus partly builds on my discussion of animals' capacity to recognise suffering in other individuals in Chapter 2. I shall follow a similar line of argument and briefly explore the empirical evidence that supports the view that animals can recognise intentional action in others.

Many examples of behaviour indicate that animals can rely on bodily cues to recognise intentional infliction of suffering. These include animals' reactions to others' anger, which suggests that animals could be aware that another individual has the *intention* to attack them if nothing is done to de-escalate tense interactions. I shall come back in greater detail to the communicative role of anger in my fourth chapter. Here, I shall briefly explore two other examples that render plausible the view that animals could recognise intentional harm.

[338] For an overview, see Setiya (2022).

First, several species of animals can distinguish between human beings' intentional and unintentional actions in cases involving food withdrawal. In their experiment carried out on dogs, researchers Britta Schünemann, Judith Keller, Hannes Rakoczy, Tanya Behne, and Juliane Bräuer submitted the dogs to three situations: one in which a treat was intentionally withdrawn from them and two in which the experimenter unintentionally withdrew the treat, either because she was clumsy or because she could not physically give it to them.[339] The researchers found out that dogs behaved differently with the experimenter who had intentionally withdrawn food. Indeed, dogs waited much longer before approaching her, often sat or laid down, and stopped moving their tails, indicating they were attentive to the interaction.[340] According to the researchers, it is plausible that dogs interpreted it as threatening or confusing.[341] The study thus offers evidence that dogs could distinguish between an unwilling and unable agent. Furthermore, as highlighted by the researchers, dogs are not the only animals capable of recognising goal-directed behaviour. Similar experiments have been conducted on chimpanzees, capuchin monkeys, Tonkean macaques, African grey parrots, and horses. These animals have been found capable of making a similar distinction between unwilling and unable agents.[342]

Second, studies on social play in animals suggest that they can use play signals to express the desire to play and encourage other conspecifics to play. These signals, such as the play bows that many canine species use, enable animals to initiate play or to maintain a joyful mood while playing, especially after rough biting or hitting.[343] Dog cognition specialist Alexandra Horowitz notes that almost all play sequences begin with play signals. These can be

339 Schünemann, B., Keller, J., Rakoczy, H. *et al.* (2021).
340 *Ibid.*
341 *Ibid.*
342 *Ibid.*
343 Horowitz (2010), p. 119.

short or long, depending on the playmates' relationship and level of familiarity.[344] Play bows are not merely unnecessary greetings; they are required for animals to engage in social play, Horowitz notes. These signals enable animals to recognise that others' assaults are part of a game and do not constitute serious aggression motivated by anger and a desire to harm others. Horowitz writes:

> Dogs typically play together rambunctiously and at a breakneck pace. Since they are doing all manner of actions that could easily be misinterpreted—biting each other on the face, mounting from behind or fore, tackling the legs out from under another dog—the playfulness of their actions has to be manifest. If you fail to signal before biting, jumping on, hip-slamming, and standing over your playmate, you are not in fact playing; you are assaulting him.[345]

Horowitz describes dogs' most used play signal, the play bow. She writes: "A dog bent on his forelegs, mouth open and relaxed, with his rump in the air and tail high and wagging is pulling out all the stops to induce someone to play."[346] This signal is also used by other canine species, such as wolves, coyotes, and the foxes in Magdalen College, who all use similar play bows to explicitly express their intention to play rather than to engage in a fight.[347] In the same vein, studies on rats show that the behavioural sequence of social play and aggressions are highly different. Rodents, when playing, constantly monitor their partner's emotional states to maintain a playful mood.[348]

Hence, animals who engage in social play may recognise others' playful intentions by relying on play markers and emotional cues. As ethologist Marc Bekoff highlights, animals

---

344 *Ibid.*
345 *Ibid.*
346 *Ibid.*
347 Bekoff (2004), p. 502.
348 Bekoff (2008), pp. 120-121.

almost never use play signals to deceive other individuals or engage in a real fight.[349] They are rather used by animals as a reliable source of information about others' intentions. Social play thus provides strong evidence that animals can distinguish between playful and harmful intentions in a specific context, and many mammalian and avian species engage in social play. These include great apes, monkeys, dogs, wolves, foxes, coyotes, gazelles, elephants, wildebeest, cows, wallabies, and magpies, to name a few.[350]

However, just as empathy could rely on deeper insight into the minds of others and a more sophisticated understanding of others' emotional states, some sceptics could argue that access to wrongdoing could require a full-fledged theory of mind. As seen previously, a more demanding view of this kind faces three important challenges that minimal accounts of animals' moral capacities can easily meet.

First, animals could recognise suffering in others by observing their bodily cues and making inferences about their emotional states. Similarly, animals could recognise intentions by being sensitive to agents' attentional state and behaviour. Indeed, philosopher Tadeusz Zawidski notes that animals and infants seem to recognise goal-directed behaviour even if they do not possess sophisticated mind-reading capacities or the explicit concept of *intention*. Instead, they do so by being attentive to others' facial expressions, gaze, and eye contact, by making associations between bodily cues and agents' mental states, and by using induction to predict others' behaviour.[351] For instance, dogs know when human beings pay attention to them and are often said to "outperform apes" when it comes to following human beings' gaze.[352] Furthermore, researchers Josep Call, Juliane Braüer, Juliane Kaminski, and Michael

[349] Bekoff (2004), p. 502.
[350] *Ibid.,* p. 499.
[351] See evidence cited in Zawidski (2013). p. 16, p. 38 and p. 46.
[352] Benz-Schwarzburg, Monsó & Huber (2020), p. 4.

Tomasello note that it is very unlikely that dogs react only to superficial bodily cues. They rather seem to rely on human beings' eyes to detect their intentions and emotional states.[353] They can also anticipate retaliation and use their past experiences to predict human beings' behaviour by making connections between bodily cues and actions.[354] Hence, possessing a sophisticated theory of mind is likely not required to recognise intentional action, although such capacity could allow them to understand others' mental states more deeply and more accurately.

Second, it is highly plausible that children can recognise intentional action before the age of three. They could possess a self-other distinction and be aware that others have goals without representing intentional action *as* intentional action or intentions *as* intentions. Psychologists Kiley Hamlin, Karen Wynn, and Paul Bloom highlight several studies that suggest that children as young as three months can already ascribe goals to others and even identify unfulfilled goals.[355] Similarly, Zawidski points out that six-month-old babies recognise that others pursue goals and choose the most efficient means available to achieve them[356], a thesis that was also confirmed by studies conducted on nine-month-old babies and adult chimpanzees.[357] Finally, some studies indicate that 12-month-old infants can already represent simple mental states like intentions and that this capacity is more robustly acquired by 18 months.[358] A demanding criterion like a full-fledged theory of mind would exclude pre-verbal children from the realm of agents capable of recognising the intentional actions of others and,

[353] Call, Bräuer, Kaminski & Tomasello (2003), p. 262.
[354] *Ibid.*
[355] See studies cited in Hamlin, Wynn & Bloom (2010), pp. 923-924.
[356] Zawidski (2013), p. 15.
[357] For an overview of the empirical literature, see Scholl & Tremoulet (2000).
[358] Povinelli (1996), p. 297 and p. 299.

by extension, nonhuman animals. However, this consequence is simply incompatible with the empirical evidence pertaining to children's cognitive capacities.

Third, as mentioned in my second chapter, even if we accept the view that a theory of mind and more sophisticated mind-reading capacities are required for an individual to recognise intentional action in others, *some* animals could recognise intentional action in others. Indeed, it is plausible that great apes possess a theory of mind, but this claim is contentious and still debated, as already noted.[359]

3.2.2. Animals can have access to the badness of others' suffering.

We can now move to the second pillar of the thesis I defend in this chapter. As argued in my second chapter, animals capable of even a basic form of empathy like emotional contagion can have access to the badness of suffering not only for themselves, but also for others. The argument I developed previously can be summarised as follows:

> (P1) An empathetic animal takes another individual in a state of suffering as the intentional object of her empathy, which
>
> (P2) Leads her to suffer in tune with the sufferer, which
>
> (C) Allows her to have access to the badness of suffering for that individual, and not just for herself.

Three additional sub-premises should be added between the second premise and the conclusion to fully account for the link between emotional contagion, included in P2, and animals' capacity to have access to the badness of others' suffering. These are the following:

[359] For an overview of the debate, see de Waal (2006a), p. 70 and O' Connell (1995), pp. 397-410.

(P1) An empathetic animal takes another individual in a state of suffering as the intentional object of her empathy, which

(P2) Leads her to suffer in tune with the sufferer.

(P2.1) Suffering in tune with the sufferer is deeply aversive, is negative well-being.

(P2.2) The fact that suffering is deeply aversive, is negative well-being, is a property that explains suffering's badness, either by being identical to it, grounding it, tracking it, etc.

(P2.3) The empathetic animal's emotional contagion is directed at the sufferer.

(C) This allows her to have access to the badness of suffering for that individual, and not just for herself.

The argument I developed in my second chapter underpins the thesis that I intend to defend in this chapter, namely, that basic forms of empathy like emotional contagion, along with animals' capacity to recognise intentional action in others, enable animals to have access to the wrong-making features of causing suffering.

3.2.3. Animals can have access to the wrong-making features of causing suffering.

I have argued previously that emotional contagion can enable animals to have access to the badness of others' suffering, which is felt as deeply aversive. Animals can show some minimal awareness of the features that explain why suffering is *bad*: because it is negative well-being. The experiential badness of suffering is closely linked to its moral badness, and animals can have access to that moral fact.

That said, emotional contagion alone can only provide access to why others' suffering is a bad state of affairs.[360] It cannot explain why the *action* of intentionally causing suffering is wrong. For that, the capacity to recognise intentional action in other individuals is necessary. With capacities for emotional contagion and recognition of intentional action in others, animals can be aware of how some actions can negatively affect others' well-being. Animals thus seem to be capable of associating the action itself, its consequences on others' emotional state, and the badness of others' suffering.

It is precisely the badness of suffering that explains why causing suffering is *wrong*. Causing suffering is tantamount to causing a bad state of affairs, and animals can have access to that moral fact. Thus, animals already possess some elements that explain why causing suffering is wrong: because it causes something bad. The badness of suffering is a wrong-making feature of causing suffering.

The relationship between suffering's badness and the wrongness of inflicting suffering is accepted by proponents of various moral theories. For instance, Thomas Scanlon acknowledges that the fact that suffering makes individuals worse off gives agents a reason to prevent it or alleviate it and that such a meta-ethical framework "has its greatest plausibility with respect to experiential states of this kind."[361] That said, Scanlon does not hold the same view for other things we value, such as friendship or knowledge.[362] Similarly, Thomas Nagel, while being critical of consequentialism, still accepts the idea that suffering is a bad state of affairs that ought to be prevented.[363] Hence, one does not need to defend a specific normative

[360] For a similar claim, see Mallon & Nichols (2010), p. 307.
[361] Scanlon (1998), p. 92.
[362] *Ibid.,* pp. 95-97.
[363] Nagel (1986), pp. 147-163. Also cited in Scanlon (1998), pp. 81-82.

framework to accept the relationship between suffering's badness and the wrongness of causing it.

Yet it is worth noting that my account of animals' access to wrongdoing fits well with the consequentialist tradition, especially hedonistic utilitarianism. According to this version of utilitarianism, an action is wrong if it fails to maximise well-being or happiness, the latter being defined in relation to the experience of pleasure and the absence of suffering. The experience of pleasure is taken to be a good state of affairs that ought to be maximised, whereas suffering is considered a bad state of affairs that ought to be minimised. Animals can recognise the intentional infliction of suffering, can make the association between the intentional action and its immediate consequences, and can have access to the badness of suffering. Hence, their access to the wrong-making features of causing suffering aligns with moral theories that rest on suffering's badness. These include hedonistic utilitarianism as a paradigmatic example.

The parallel between animals' access to wrongdoing and hedonistic utilitarianism is also supported by the view, common to consequentialists, that what we ought to do or refrain from doing depends on evaluative properties like *good* and *bad*. In other words, hedonistic utilitarians would agree that the *badness* of suffering is prior to the *wrongness* of causing suffering and that it explains why causing suffering is wrong, although consequentialists do not agree on the precise nature of the relationship between evaluative and deontic properties.[364] Animals can have access to the wrong-making characteristics of causing suffering by virtue of having access to the badness of suffering, which is prior to the wrongness of causing it.

The general line of argument I propose in this chapter might be compared to the account of good will developed by Nomy Arpaly and Timothy Schroeder. In their book *In*

[364] For an overview of the different options, see Berker (2018), pp. 739-741. As noted by Berker, evaluative properties' priority could be either understood in terms of epistemic priority, supervenience, logical entailment, identity, or grounding. I shall not address this question here.

Praise of Desire (2013), Arpaly and Schroeder define complete good will as intentionally doing what is right, as "an intrinsic desire for the right or good, correctly conceptualised."[365] By that, they mean that an individual with complete good will possesses the conceptual elements associated with the definition of right or good, according to various moral theories. Kantians would define the right as *respecting persons*, whereas utilitarians would define it as *maximising happiness*. An individual with complete good will has the intrinsic desire that persons be respected or that happiness be maximised, depending on which moral theory is the correct one.[366] This question is left open.

More interestingly, Arpaly and Schroeder acknowledge that some individuals can possess partial good will, which they define as "an intrinsic desire for *some part* of the right or good, correctly conceptualised."[367] [emphasis added] An individual with partial good will does not desire the good or right *per se* or does not conceptualise the good or right by referring to notions such as *respecting persons* or *maximising happiness* but still desires some elements associated with *respecting persons* or *maximising happiness*. To take hedonistic utilitarianism as an example, an individual with partial good will would perhaps desire that others' suffering be alleviated or avoided or that others' pleasure be maintained or increased. Concepts such as *maximising* and *happiness* would not be part of her desire for the good, but other closely related notions would be.

I think Arpaly's and Schroeder's account could leave the door open to recognising animals as capable of good will, although they do not explore this implication in their work and could even argue that animals cannot "correctly conceptualise" what is right or good.[368]

---

[365] Arpaly & Schroeder (2013), pp. 161-163.
[366] *Ibid.,* pp. 164-165.
[367] *Ibid.,* pp. 161-163.
[368] In fact, this claim is consistent with Arpaly's views on animals' moral capacities. According to Arpaly, animals cannot respond to moral reasons but can come close. See Arpaly (2003), p. 146.

That said, my interpretation also finds support in Arpaly's and Schroeder's claim that partial good will could take the form of an *aversion* for what is bad or wrong and not merely a *desire* for what is good or right. According to Arpaly and Schroeder, an individual with an aversion for causing suffering or witnessing the infliction of suffering would have partial good will, but not complete good will, as the latter requires the explicit possession of more sophisticated capacities and moral concepts.[369] Hence, maybe animals would qualify as individuals capable of partial good will, for it seems that they already possess some elements associated with the definition of wrongdoing, such as *suffering* and *causing* suffering, and have access to the *badness* of suffering. I shall come back to this thesis in Chapter 4 and its implications for Strawsonian theories of moral responsibility.

Arpaly and Schroeder explicitly claim that they do not need to be committed to a specific meta-ethical framework or moral theory to defend the view that some individuals can have partial ill will or good will. Similarly, to defend my argument, I do not need to specify what I mean by "having access" to the features that explain why causing suffering is wrong. The wrongness of such action could be identical to causing a bad state of affairs or could be grounded in the badness of suffering without being reducible to it.[370] However, I shall not answer these questions here. My argument is compatible with various meta-ethical frameworks pertaining to the relationship between deontic and evaluative facts, and states of affairs.

Before turning to objections, I want to highlight my account's main strengths. The argument I propose here offers several advantages, in addition to avoiding the shortcomings of threshold views of moral agency that I already criticised in my first chapter. I shall stress three strengths here.

---

[369] *Ibid.,* p. 180.
[370] For a summary of the debate, see Tappolet (2022).

First, my argument can be applied to both interactions in which animals are wronged and interactions in which animals are the wrongdoers. For instance, because some animals can recognise intentional actions in others and have access to the badness of their own suffering, animals can recognise the wrong-making features of causing suffering in interactions in which they are the victim of wrongdoing. This may also follow if they are the wrongdoer. If one of the foxes in Magdalen College intentionally bites her playmate too hard, which leads her partner to suffer, and if the playmate's distress is felt as aversive, it seems that the fox already has access to the fact that her rough playing constitutes wrongdoing. This may then lead her to change her behaviour and adjust to her partner's emotional state.[371]

Second, my account is also empirically supported by the literature on what neuroscientist James Blair has called the violence inhibition mechanism (VIM). According to Blair, human and nonhuman animals are endowed with a mechanism that allows them to inhibit violence by being attuned to others' distress cues and refraining from causing further suffering. [372] For example, dominant wolves usually stop fighting when their opponent displays submission cues like throat-baring. The VIM has also been taken to interact with animals' capacity for empathy. Indeed, finding others' distress aversive could lead agents to withdraw from the situation and deescalate aggressive interactions, which are felt as highly unpleasant.[373]

Third, my account is minimal and inclusive of several moral agents but makes room for more sophisticated forms of awareness of wrongdoing, which may involve not only

[371] Moreover, this aversion for causing suffering to her playmate could constitute a precursor of full-fledged guilt, which rests on self-reflective capacities, but also includes a negative felt component. Indeed, the cognitive dimension of guilt involves taking our own actions or thoughts as the intentional object of our emotion. This capacity may not be possessed by animals, and there is no strong empirical evidence of guilt in animals, including in guilty-looking dogs. That said, empathetic animals can be aware of the consequences of their own action on others' well-being and find their suffering aversive. These capacities may serve as a basis on which guilt develops. I shall come back to this possibility in Chapter 4.
[372] Blair (1999), p. 143.
[373] Blair (1995), p. 4.

aversiveness for others' suffering, but also moral understanding and reactions like third-party intervention to stop fights or console distressed individuals, which are frequent in great apes. As Claudia Rudolf von Rohr, Judith M. Burkart, and Carel P. van Schairk note, chimpanzee bystanders often intervene in fights or provide consolation to conspecifics, and these reactions can hardly be explained in terms of stimulus-response mechanisms.[374] Moreover, violent behaviour like infanticide usually triggers strong reactions from other members of a chimpanzee community, who make loud barking sounds in protest.[375] Similarly, children as young as three years intervene in interactions that involve the transgression of moral norms.[376] These findings indicate that young children and some animals may have expectations about how others should behave regarding one another and not merely personal expectations about how others should treat them in the context of dyadic interactions.

Hence, my proposal leaves the door open to acknowledging that children and some nonhuman animals can have access to the wrong-making features of causing suffering not only in the context of dyadic interactions, but also in interactions in which they are not involved. Again, this claim is consistent with the literature on children's moral development. According to Paul Bloom, it is highly plausible that babies react to wrongdoing done to another individual by feeling distressed after witnessing intentional harm.[377] Similarly, William Arsenio and Anthony Lover note that the sufferer's distress provides information to children about the consequences of harming others. Agents' negative emotional states can be used by bystander children to determine the wrongness of others' actions.[378] Later in life, children explicitly mention harm as a reason underlying the wrongness of moral transgressions and

374 Rudolf von Rohr, Burkart & Shaik (2011), pp. 17-18.
375 *Ibid.*
376 See Vaish, Missana & Tomasello (2011).
377 Bloom (2013), p. 55.
378 See Arsenio & Lover (1995). Cited in Aureli & de Waal (2000), p. 358.

usually appeal to the consequences of an action on others' well-being to reflect on the rightness or wrongness of an action.[379] Nevertheless, children as young as five months can already recognise harmful intentions, recognise suffering in others, and "catch" others' emotions. These capacities, I argue, are sufficient to grant them the capacity to have access to the features that explain why causing suffering is wrong.

3.3. Three Objections and Responses

3.3.1. Animals do not explicitly possess the concept of what is *morally* wrong.

Some sceptics could be unconvinced by the arguments I develop in this chapter. There are various ways in which they could resist my main conclusions, and several criticisms can be raised against my account. I shall respond to three of them here.

The first line of objection could run as follows: because animals do not explicitly possess the concept of *moral* wrong, as opposed to *conventional* wrong, they cannot be said to access wrongdoing pertaining to moral matters.[380] Indeed, many philosophers and psychologists would agree that it is not sufficient for individuals to have access to wrongdoing at large to be fully developed moral agents. Agents must also distinguish between different *types* of wrongdoing. As we have seen in Chapter 1, §1.3.2, it is widely accepted that two-year-old children already implicitly grasp the distinction between moral and conventional norms.[381] They can understand that hitting another toddler is *morally* wrong and constitutes a serious norm violation pertaining to the well-being of others, as opposed to wearing pyjamas at school, which is taken to be merely *conventionally* wrong.[382] Animals, on their part, might react to all

[379] Nucci (2001), p. 15 and Nucci & Nucci (1982), p. 411.
[380] See, for instance, Jaquet (2022).
[381] Nucci & Turiel (1978), p. 406.
[382] Smetana & Braeges (1990), p. 342.

sorts of things they can broadly recognise as wrong but not as *morally* wrong. In order to be capable of having access to moral wrongdoing in others, individuals should distinguish between an action that is *morally* wrong and an action that is wrong for other reasons, for instance, because it violates social conventions, like etiquette norms. Philosophers like Mark Rowlands could argue that this distinction grounds some capacities essential to moral agency, such as moral understanding, as we have seen in §1.2.3.

In response, it is worth referring again to the limitations of threshold views of moral agency I have explored in my first chapter, §1.4.2, according to which an individual becomes a moral agent when her moral capacities reach a certain level of development. In Chapter 1, I argued that gradualist views of moral agency can better face the empirical challenges other accounts fail to meet. Two additional things can be noted.

First, children react very early in life to moral wrongdoing while remaining largely unresponsive to conventional wrongdoing. For example, infants already find harm done to another child aversive[383] and have preferences for helping individuals.[384] It is only later in life, starting from the third year of life, that children start grasping some key differences between moral and conventional wrongdoing, and this distinction is more robustly acquired by the age of 34 and 42 months.[385] Toddlers understand that moral violations are more serious, universal in scope, and do not depend on a given social context.[386] Children who can grasp the differences between moral and conventional wrongdoing are definitely more robust moral agents. Nonetheless, it is far from obvious that such *explicit* distinctions are needed to have access to the wrong-making features of causing suffering. As shown by the case of infants,

[383] Bloom (2013), p. 55.
[384] Hamlin, Wynn & Bloom, Paul (2010).
[385] Smetana & Braeges (1990), p. 342.
[386] *Ibid.*

individuals may implicitly recognise moral wrongdoing before they reach the age of three, as moral violations already elicit strong empathetic reactions and give rise to social preferences in very young children.

Although few similar experiments have been carried out on animals, the diversity of animals' emotional reactions may provide evidence that they could react to *moral* wrongdoing and not just any type of wrongdoing. For instance, dogs respond to rules that rely on the presence of an external authority, like not eating sausages, by monitoring humans' attentional state. They were also found to obey orders and refrain from eating the tempting sausages more frequently when being watched by their human guardians.[387] Such behaviour has been interpreted by scientists as fear of punishment. These responses sharply contrast with dogs' automatic aversiveness for others' suffering. Others' plights usually trigger empathetic responses like emotional contagion that are spontaneous and do not rely on the presence of an authority figure.[388] This suggests that dogs react differently to things they have *prudential* reasons to avoid, like not being punished for eating sausages, as opposed to morally relevant features, like others' suffering.

This brings me to my second point, which concerns suffering itself. As we have seen in the previous chapter, animals' empathy can be directed at others' suffering, which is a morally relevant feature. The fact that suffering is a morally salient element is relevant to the distinction between moral wrongness and other types of wrongness. Indeed, according to several authors who work on the distinction between moral and conventional norms and who defend what Nicholas Southwood calls the "Content View", moral norms could differ from conventional norms by virtue of their content. Moral norms typically foster normative

[387] Schwab & Huber (2006), pp. 169-175, cited in Benz-Schwarzburg, Monsó & Huber, (2020), p. 1.
[388] For studies on emotional contagion in dogs, see, for instance, Palagi, Nicotra & Cordoni (2015), Huber, Barber, Faragó, Müller & Huber (2017), and Van Bourg, Patterson & Wynne (2020).

attitudes about others' interests and actions that are other-regarding, unlike conventional norms.[389] Hence, although animals cannot distinguish between moral norms and conventional norms, one could argue that they already possess a sensitivity to features that make up moral norms, such as others' well-being.[390]

Rival accounts of the differences between moral and conventional norms include what Southwood calls the "Form View," according to which moral norms differ from conventional norms in virtue of their formal characteristics, such as their scope or normative force.[391] For instance, proponents of the Form View often argue that moral norms differ from conventional ones by being universally applicable, unlike conventional norms, which are specific to groups. Animals are incapable of such sophisticated distinctions and cannot grasp the formal differences between moral and conventional norms, which even young children can minimally understand. But animals already seem to react differently to situations involving the well-being of others, as opposed to situations that rely on the presence of an external authority and could be akin to conventional norms. They thus seem to be sensitive to some degree to the differences between moral and conventional norms, at least when it comes to their content.

3.3.2. Having access to the wrong-making features of causing suffering is not merely having an aversion for some actions.

Second, one could object that having access to the wrong-making features of causing suffering requires more than having a form of aversion for some actions. To put it differently, having access to the wrong-making features of actions is not tantamount to merely feeling others'

[389] Southwood (2011), p. 764 and p. 770-771. See also Brennan, Eriksson, Goodin & Southwood (2013), pp. 64-67.
[390] For an overview of the debate, see Southwood (2011), pp. 763-764 and Brennan, Eriksson, Goodin & Southwood (2013), pp. 59-67.
[391] *Ibid.*

distress as aversive and recognising intentional harm. More sophisticated cognitive capacities might be required. Ron Mallon and Shaun Nichols claim that access to wrongdoing presupposes an awareness of a transgression. According to the authors, a mechanism like the Violence Inhibition Mechanism (VIM) cannot explain alone why an action is wrong. It can merely explain why others' distress is felt as bad.[392] Simply recognising harmful intentions to others and feeling an aversion for a sufferer's pain would be insufficient.

Perhaps surprisingly, I think this objection is essentially correct. In this chapter, I am not committed to the view that having access to wrongdoing is reducible to having an aversive reaction. As philosopher Michael Slote puts it, a capacity like empathy is not the reason why cruelty is wrong, but it rather *gives us reasons* to think acting cruelly is wrong.[393] It does so by enabling us to have some insight into the emotions of others and making us aware of the consequences of certain actions on others' well-being. The view I am defending here resembles Slote's, for animals' capacity to feel others' suffering and recognise intentional action enables them to have access to the reasons why causing suffering is wrong: because infliction of suffering causes a bad state of affairs. The wrongness of such action cannot be disentangled so easily from the consequences that causing suffering has on the well-being of sentient creatures.[394] It rests on its consequences on individuals' emotional states, which empathy makes vivid to us. I can thus be committed to the view that animals have access to the wrong-making features of intentionally causing suffering while acknowledging that full-fledged awareness of wrongdoing interacts with more sophisticated capacities. The latter is not merely a matter of aversion.

---

[392] Mallon & Nichols (2010), p. 307.
[393] Slote (2009), p. 70.
[394] For a similar line of argument, see Nucci (2001), p. 7.

3.3.3. Having access to wrongdoing requires more sophisticated cognitive capacities, according to hedonistic utilitarians.

The third criticism does not target my argument, which could be accepted by proponents of several moral theories, but the similarities I put forward between my main thesis and hedonistic utilitarianism. The objection can be stated as follows: to have access to wrongdoing in the utilitarian sense, individuals must understand that others can cause suffering not only intentionally, but also out of omission, negligence, ignorance, and so on. In other words, moral agents must possess more sophisticated cognitive capacities than mere intentional action recognition and emotional contagion. This is all the more relevant as utilitarianism rejects the distinction between actions and omissions.[395] Furthermore, to fully understand why causing suffering is wrong in the utilitarian sense, agents must be aware of the consequences of an action on *all* sentient individuals. Animals are incapable of such complex calculations.

I think this objection rightly points to important limitations in animals' responses to wrongdoing, for animals' access to wrongdoing appears to be limited to *intentional* actions and their very *short-term* consequences. Indeed, the empirical evidence suggests that most animals might understand basic causal relationships and can make the link between an intentional action and its immediate consequences.[396] Animals such as dogs and elephants are aware of the consequences of their own actions, possess body awareness, and can see their body as an obstacle to the pursuit of a goal.[397] However, the current studies do not show that animals can have more abstract representations of causal chains.[398] This means they cannot grasp the relationship between intentional actions and their consequences in the long term. Similarly,

[395] On that point, see Williams (1973), p. 95 and Crisp (1997), p. 140.
[396] For a review, see Penn & Povinelli (2007).
[397] Lenkei, Faragó, Zsilák *et al.* (2021), p. 2761.
[398] Penn & Povinelli (2007), pp. 110-111.

animals cannot understand omissions' short-term and long-term consequences on others' well-being.

Nevertheless, I think this objection fails to make the case that animals cannot react to wrongdoing *tout court*, even in a minimal hedonistic utilitarian sense. Animals might have access to the wrong-making features of causing suffering, namely, that suffering is deeply aversive, is negative well-being, but they cannot understand how it could be right in some circumstances, if it is done to prevent greater suffering. Their access to wrongdoing is limited to *intentional* infliction of suffering and its short-term consequences, but that limitation does not mean they cannot have access to it at all.

It is also worth noting that this account of animals' reactions to wrongdoing is consistent with hedonistic utilitarianism and, more precisely, with John Stuart Mill's moral psychology and views on the natural origins of justice. In *Utilitarianism* (1861), Mill aims to show that justice can be consistent with utilitarianism and its core moral principle, the greatest happiness principle. Mill argues that the origin of justice can be found in two elements: the belief that harm has been done to an individual and the desire to punish the wrongdoer.[399] The latter desire stems from two impulses or emotional capacities that we have in common with nonhuman animals: the desire to defend oneself against aggression and sympathy, which he defines as the capacity to take on the emotions of others.[400] Sympathy allows us to feel the pain of other individuals as painful and their pleasure as pleasurable. Sympathy is not a unitary emotion like fear, joy, or anger. It is a power that can involve both a cognitive component, like imaginative projection, and an affective one.[401] As highlighted by Dale Miller,

[399] Mill develops his views in Chapter 5 of *Utilitarianism*. See Robson (Ed.). (1985), p. 248. See also Crisp (1997), pp. 156-161.
[400] See Mill's essay "Bain's Pychology", in Robson, J. M. (Ed.). (1978), p. 362.
[401] *Ibid.*

psychologists and many philosophers would today use the word "empathy" to describe what Mill calls "sympathy".[402] This claim also finds support in Mill's remarks on the possibility of sympathising with others' anger, not just their suffering.[403]

According to Mill, we share these sentiments with nonhuman animals, but our moral capacities are more sophisticated in two ways. First, human beings can sympathise with all sentient beings, not just their offspring. Second, humans possess superior intelligence, which enables them to apprehend "a community of interest" whose security should be defended.[404] Although these powers are grounded in the natural sentiments we have in common with other animals, such sentiments are not moral in themselves. They become moral only when subordinated to social sympathies and *rules* aimed at promoting the general good.[405]

My argument is consistent to some extent with Mill's views on the natural origins of justice, although, unlike Mill, I do not posit a sharp distinction between the moral and nonmoral nature of humans' and animals' capacity for expanding the moral circle. I want to suggest here that Mill's claims on the origins of justice are entirely plausible, including his views on animals' desire for punishment, which will be explored in my final chapter. Mill does not defend his theory at length in *Utilitarianism* but rather posits it. I think it can be defended by appealing to animals' access to the badness of suffering and their capacity to recognise intentional action in others, or what Mill would define as "the belief that harm was done".

---

[402] Miller (2006), p. 163.

[403] In an 1859 letter to William George Ward, Mill notes: "Now it is an indispensable condition of all society, except between master & slave, that each shall pay regard to the other's happiness. On this basis, combined with a human creature's capacity of fellow-feeling, the feelings of morality properly so called seem to me to be grounded, & their main constituent to be the idea of punishment. I feel conscious that if I violate certain laws, other people must necessarily or naturally desire that I sh[ould] be punished for the violation. I also feel that I sh[ould] desire them to be punished if they violated the same laws towards me. From these feelings & from my sociality of nature I place myself in their situation, & sympathize in their desire that I sha be punished; & (even apart from benevolence) the painfulness of not being in union with them makes me shrink from pursuing a line of conduct which would make my ends, wishes, & purposes habitually conflict with theirs." See Mineka, F. E. & Lindley, D. N. (Eds.). (1972), p. 650. I am grateful to a reviewer at *Utilitas* for pointing out this letter to me.

[404] The passage is taken from Mill's *Utilitarianism*, Chapter 5. See Robson (Ed.). (1985), p. 248.

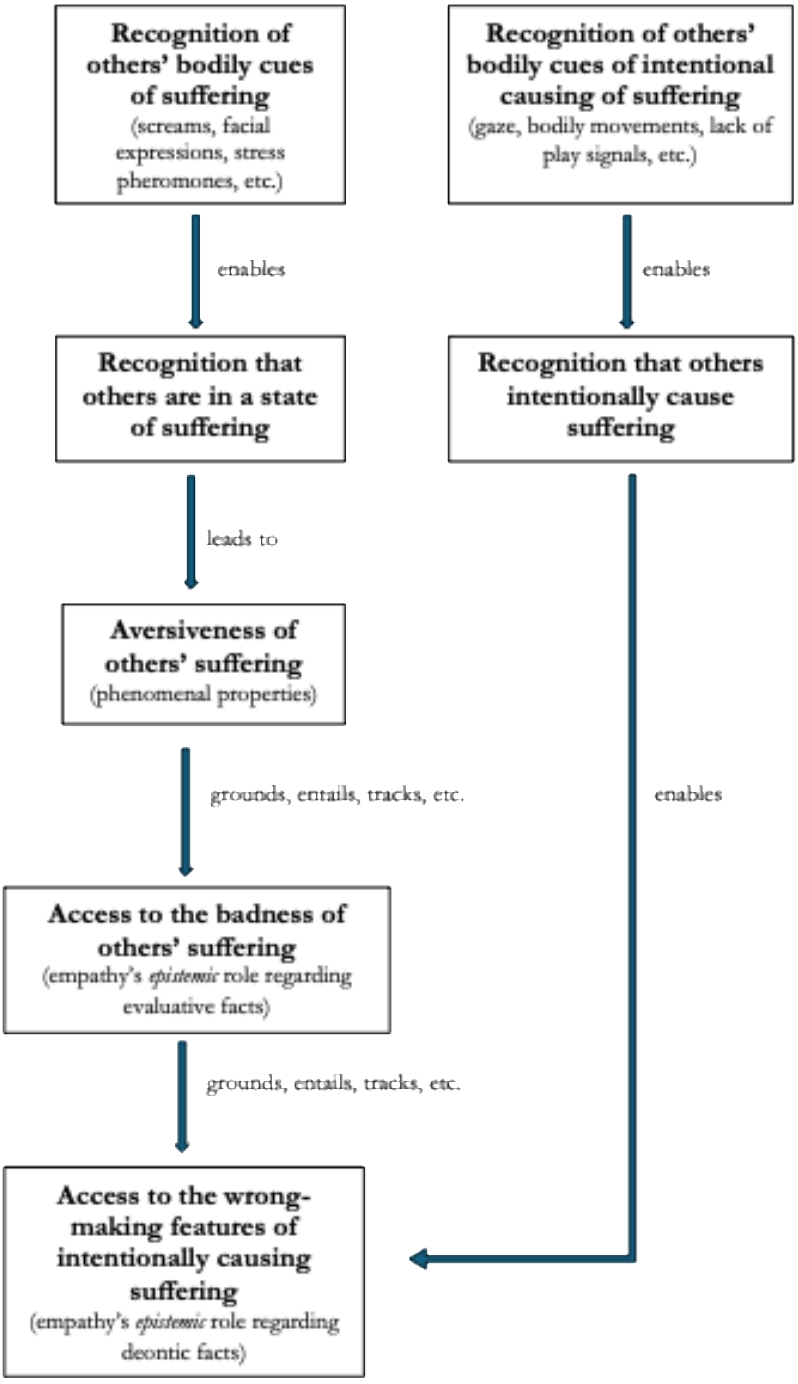
[405] On that point, see Crisp (1997), p. 159.

These two capacities allow them to recognise why some actions are wrong in a way that is akin to a very simple type of act utilitarianism. That said, both *justice* and *full-fledged* awareness of wrongdoing, at least in the hedonistic, utilitarian sense, do not merely consist of witnessing intentional harm and catching others' emotions. Animals can have access to the wrong-making features of causing suffering but lack several crucial capacities required to possess a more developed understanding of wrongdoing and injustice.

3.4. Summary of Chapter 3 and Concluding Remarks

In this chapter, I have argued that minimal empathy like emotional contagion and the capacity to recognise goal-directed behaviour enable animals to access wrongdoing. More precisely, I have argued that animals can have access to the fact that causing suffering is wrong if (1) they can recognise that individuals can intentionally cause suffering and (2) they can have access to the badness of others' suffering. These two capacities allow animals to (3) have access to the wrong-making features of causing suffering. This additional step can be illustrated with the following simple diagram, which builds on the one I proposed at the end of Chapter 2:

Summary of Chapter 3



I have also highlighted several ways in which animals' recognition of wrongdoing suffers from important limitations. Indeed, animals may possess some elements associated with the definition of wrongdoing but lack what Arpaly and Schroeder call "complete" good will or ill will. Furthermore, animals' recognition of wrongdoing is limited to intentional actions and their immediate consequences, since they cannot understand the distinction between actions and omissions and their long-term consequences on others' well-being. In other words, animals have access to the wrong-making features of *a certain subset of actions* and have access to the reasons why *intentionally* causing suffering is generally wrong.

If the arguments proposed in my second and third chapters are successful, they leave the door open to arguing that animals could engage in moral responsibility practices grounded in the recognition of others' ill will. If animals can have access to wrongdoing in others, they might rightly hold other group members responsible for their actions and feel reactive attitudes toward them, such as moral anger, resentment, and the like. I have not addressed the complex relationship between empathy and anger, how animals' empathy for the sufferer could lead them to feel angry toward the offender, nor how anger could be understood as a form of blame. These questions will be explored in my fourth chapter.

Chapter 4

Empathy's Self-Control Role: Can Animals Hold Each Other Morally Responsible?

As Mill writes, "we do not call anything wrong unless we mean to imply that a person ought to be punished in some way or other for doing it; if not by law, by the opinion of his fellow creatures; if not by opinion, by the reproaches of his own conscience."[406] In other words, we cannot say that an agent can have access to wrongdoing without being susceptible to engage in moral responsibility practices that involve blaming or punishing the wrongdoer. A wrong action is one that agents are expected to refrain from doing. Stephen Darwall and Michael McKenna note that this position is fairly uncontroversial in moral philosophy.[407] Many contemporary philosophers defend the same view as Mill's or simply posit it.[408]

As I remarked in Chapters 1 and 3, animals have traditionally been excluded from the realm of individuals who can recognise wrongdoing and, by extension, from those who can be held morally responsible for their actions. This capacity has been denied on rationalist grounds, according to which animals do not meet a certain high threshold of moral development to be held morally responsible. Animals lack moral understanding, the capacity to form moral judgements, the capacity to govern themselves in light of moral reasons, and the capacity to feel self-directed emotions like guilt and shame.[409] One of the foxes playing in

---

406 The passage is from Mill's *Utilitarianism*, Chapter 5. See Robson (Ed.). (1985), p. 246. Also cited in Darwall (2006), p. 27.
407 Darwall (2006), pp. 27-28 and p. 92, and McKenna (2012), p. 14.
408 See, for instance, Adams (1999), p. 238, Baier (1966), Brandt (1979), pp. 163–176, Gibbard (1990), p. 42, Skorupski (1999), p. 142, and Shafer-Landau (2003), all cited in Darwall (2006), p. 27 and p. 92. See also Copp (1997), Smith (1983), Wallace (1994), and Widerker (2000), all cited in McKenna (2012), p. 14. See also Pettigrove & Tanaka (2014), p. 270.
409 Darwall (2006), p. 78, Dixon (2008b), pp. 22-23 and p. 27, Musschenga (2015), p. 55, Wallace (1994), p. 13, Fischer & Ravizza (1993), p. 5. For a summary, see Ferrin (2019), pp. 135-137.

Magdalen College who would be wronged by their aggressive playmate and feel angry at her could not be said to have access to the wrong-making features of causing suffering or to hold her playmate responsible for her actions.

Several philosophers have recently challenged this picture. They have argued that animals engage in moral responsibility practices and can be held morally responsible by other group members.[410] In this chapter, I defend these authors' central thesis by expanding on anger's relevance to the exercise of moral agency, which has been little explored in the debate on animal morality. I highlight how animals' capacity for anger and empathy enables them to engage in moral responsibility practices that presuppose blame, reconciliation, and access to wrongdoing. To defend this thesis, I follow Asia Ferrin (2019) and Dorna Behdadi (2021 and 2024) in positing a Strawsonian approach to responsibility. This account takes our practice of holding others responsible and the reactive attitudes it entails as a starting point for outlining the nature of moral responsibility.[411] Although Behdadi and Ferrin have provided insightful accounts of how animals' anger could be described as blame (Behdadi) or how empathetic animals could be responsive to moral reasons pertaining to others' well-being (Ferrin), they have not considered the important role that empathy might play in enabling animals to understand others' anger and adjust to it.[412] This thesis chapter aims to fill this gap.

More precisely, I argue that animals' capacities (1) to have access to the wrong-making features of causing suffering and (2) to form interpersonal relationships with other animals (3) give rise to expectations about how others ought to treat them. These expectations find their expression in a specific emotion: anger. Because animals can access the wrong-making features

[410] See Shapiro (2006), Bekoff & Pierce (2009), Bailey (2014), Ferrin (2019), and Behdadi (2021).
[411] For an overview of other approaches to moral responsibility, see Stapleton (2018), pp. 29-34.
[412] See Ferrin (2019) and Behdadi (2024). Bailey (2014), on her part, mentions the possibility of extending a Strawsonian approach to moral responsibility to animals but does not expand on that remark. See Bailey (2014), p. 32.

of causing suffering and form interpersonal relationships with other group members, this renders them susceptible to anger, which interacts with the awareness that an expectation was violated.[413]

To defend this thesis, I shall first summarise Strawsonian accounts of moral responsibility (§4.1) and define some key concepts that Peter Strawson and several philosophers have included in their proposals. These include notions such as "reactive attitudes" (§4.1.1) and "quality of will" (§4.1.2). I will then expand on the three claims proposed above by drawing on the empirical and philosophical literature on animals' capacity for forming interpersonal relationships and feeling anger and empathy (§4.2). I shall then defend my account against one possible objection: that animals' anger expresses mere predictive expectations (§4.3). For now, I shall put aside the question of how *we* may hold domesticated animals morally responsible for some of their actions, to which I shall come back in Chapter 5. I shall instead concentrate on how animals form complex relationships with other group members and may hold *each other* responsible.

4.1. Strawsonian Approaches to Moral Responsibility: An Overview

In his influential paper "Freedom and Resentment" (1962), Peter Strawson defends what Victoria McGeer calls "the metaphysical noncommitment thesis", according to which "the concepts and practices of responsibility, as embodied in our reactive exchanges, do not presuppose anything so metaphysically demanding as libertarian (or contra-causal) free will."[414] To put it differently, whether we are free or, on the contrary, entirely determined by external

[413] In this chapter and Chapter 5, I do not need to be committed to a specific account of trust. I simply take for granted that animals can trust each other. For an overview of the philosophical literature on trust, see McLeod (2023).
[414] Strawson (1962), p. 69, McGeer (2012), p. 302 and Shoemaker (2007), p. 73.

causes is irrelevant to account for our practice of *holding* each other responsible. It is rather human beings' relationships and moral capacities that underlie their moral responsibility practices. No more fundamental metaphysical thesis, such as the existence of free will, is required to make sense of our moral responsibility practices. The nature of moral responsibility is rather best understood from the point of view of *holding* each other responsible.[415] In defending this approach, Strawson has proposed several novel philosophical concepts. At the heart of his arguments lie notions such as "reactive attitudes" and "quality of will". Let us define these in turn.

4.1.1. What Are Reactive Attitudes?

Strawson and philosophers after him have emphasised human beings' relationships and their relevance to philosophical inquiries about our moral responsibility practices. Strawson and many others take interpersonal relationships involving love and trust as paradigmatic examples.[416] These include romantic relationships, friendships, and family ties, to name a few. That said, some authors have recognised that we can expect minimal good will from strangers. Others have challenged that claim. For instance, Thomas Scanlon has argued that "the idea that we have a relationship with everyone in the world sounds odd", for the concept of a relationship usually refers to a specific relationship, such as the one between two friends. I shall not expand on these matters here.[417]

---

[415] On that point, see McKenna (2012), p. 22 and p. 32. For a summary of the distinction between holding and being responsible and the debate on the priority of one over the other, see Dwyer (2003), pp. 184-186, Dixon (2008a), pp. 53-55, Smith (2007), Hieronymi (2004), p. 136, and Todd (2016).

[416] Strawson (1962), pp. 63-64, Darwall (2006), p. 73, Darwall (2017), p. 300, Dwyer (2003), p. 184, Shoemaker (2007), p. 74, Hieronymi (2007), pp. 112-113 and p. 122. For an overview of the debate, see McKenna (2012), p. 45

[417] For an overview of the debate, see Shoemaker (2015), p. 102 and Scanlon (2008), p. 139.

Relationships have two essential characteristics. First, they require individuals to be emotionally committed to each other to a certain degree. In discussing the link between love and moral responsibility practices among adult human beings, philosopher Seth Shabo argues that human beings would not hold their partners responsible if they were not emotionally engaged in their relationships. Indeed, Sabo writes that "if one person is immune to taking another's behaviour personally […], the first person does not care about the second's behaviour and attitudes towards him or her in this essentially personal, emotionally engaged way."[418] An individual cannot hold others responsible with specific emotional reactions if she is *entirely* detached from her relationships with others, whether it be with partners, friends, family members, colleagues, or neighbours.

Second, such relationships give rise to different types of expectations about how individuals should treat each other. These may vary from one type of relationship to another. For example, being friends with someone typically involves much love, trust, care, and time spent together. However, other relationships, such as those between neighbours, do not entail the same expectations. Neighbours might expect each other to be polite, cordial, respectful, and perhaps to help each other sometimes, but these expectations do not amount to the level of mutual support friends expect from each other. Still, all interpersonal relationships entail mutual expectations of behaviour.[419]

These two components – being emotionally involved in relationships and expecting a certain treatment from others – are constitutive of what Strawson calls "the participant stance", which is a stance individuals take toward agents who are members of the moral community and can enter interpersonal relationships.[420] Taking the participant stance makes

[418] Shabo (2012), p. 113.
[419] On that point, see also Korsgaard (1992), pp. 309-311.
[420] Strawson (1962), p. 67.

individuals susceptible to reactive attitudes, which, in the case of human beings, include resentment, indignation, anger, hurt feelings, guilt, gratitude, forgiveness, love, etc.[421] These attitudes are constitutive of our moral responsibility practices. Feeling resentment for someone who has wronged us *is* to hold that person morally responsible for her behaviour.[422]

But we need to be more precise about what reactive attitudes are. Four points are in order. First, reactive attitudes are emotions. This view is quite uncontroversial among Strawsonians.[423] As emotions, reactive attitudes have an *intentional object*: they are about someone's behaviour. They are also *felt*, which means they come with various bodily changes. As noted by Elisa Hurley and Coleen Macnamara, reactive attitudes are not mere beliefs. They are "ways of experiencing that person as having performed a morally significant action."[424] As seen in Chapter 2, §2.2.2, emotions, which include reactive attitudes, bear a certain relationship with motivations and subsequent actions. Finally, reactive attitudes entail evaluations. When we feel resentment for someone, that person's behaviour is evaluated in a certain way: as offensive, contrary to our interests, violating an expectation, or even as *morally wrong* in the case of reactive attitudes that entail *explicit* moral judgements.[425]

These remarks bring me to my second point. Because they involve evaluations, reactive attitudes are sensitive to agents' "quality of will". Strawson broadly defines that notion as a form of concern for others[426], and it is often conflated with concepts such as "wrongdoing" or "wronging" in the philosophical literature.[427] When we feel positive or negative reactive attitudes toward someone, we evaluate that person's behaviour as having upheld or violated

---

421 *Ibid.*, pp. 63-64.
422 *Ibid.*, pp. 70-71.
423 Shoemaker (2015), pp. 7-8, Hurley & Macnamara (2010), p. 373, Wallace (1996), p. 76, McKenna (2012), p. 64, Watson (2004), p. 222 and p. 227, and Macnamara (2012), p. 153.
424 Hurley & Macnamara (2010), p. 373.
425 *Ibid.,* p. 385.
426 Strawson (1962), p. 70.
427 See, for instance, Shoemaker (2013), p. 113 and Darwall (2006), pp. 72-73

some expectations of good will, which may vary according to our relationship with that person. As such, reactive attitudes are sensitive to the impact of others' behaviour on relationships and the expectations of good will to which such relationships give rise.[428]

Third, reactive attitudes are constitutive of our moral responsibility practices. They are best understood as a form of praise and blame.[429] When we feel gratitude for others, we implicitly or explicitly praise them for having shown good will toward us. Conversely, when we resent others, we blame them for failing to meet our expectations of good will.[430] In that sense, many reactive attitudes – at least the ones that are expressed and not repressed – are communicative: they signal approval or disapproval of others' behaviour. When an individual who has been wronged feels resentment and expresses it, what that person fundamentally expresses through her emotional state is an expectation of good will.[431] She expresses how she feels about the other's lack of good will and expects that person to understand her emotional state, recognise her expectation of good will as legitimate, and adjust to it by sincerely apologising and vowing to do better in the future.

Fourth, reactive attitudes are appropriate responses to others' quality of will insofar as others possess some *epistemic* and *self-control* capacities. In other words, members of the moral community must be capable of *having access* to others' expectations as expressed in their reactive attitudes and of *changing* their behaviour in light of such expectations. These two capacities have traditionally been defined in highly rationalist terms by Strawsonians and philosophers working on moral responsibility more broadly. For instance, philosopher David Shoemaker,

---

[428] Scanlon (2008), pp. 131-132 and p. 140, Wallace (1994), p. 19, Macnamara (2012), pp. 142-143, and McCormick (2022), pp. 61-66.
[429] Hurley & Macnamara (2010), p. 389. See also Darwall (2017), p. 300, Rosen (2015), p. 68, and McCormick (2022), p. 63.
[430] Coates & Tognazzini (2012), pp. 9-16, Sher (2006), p. 13, and Nussbaum (2016), pp. 257-259.
[431] Shoemaker (2013), pp. 116-119, Watson (2004), pp. 229-230, McKenna (2012), p. 4, Sher (2006), p. 13, Darwall (2006), pp. 75-76, Shoemaker (2007), pp. 71-76, Macnamara (2011), p. 92, and McGeer (2012), p. 303.

in his summary of the literature on Strawsonian accounts, writes that to be held morally responsible, moral agents typically "understand the concepts in moral judgements and appreciate justifications."[432] Similarly, R. Jay Wallace writes that an individual must possess "the power to grasp and apply moral reasons, and the power to control one's behaviour by the light of such reasons"[433] to be appropriately held morally responsible. Regarding self-control, this capacity has often been taken to be closely tied to *reflective* self-control, which is the capacity to scrutinise one's desires and make decisions according to the outcome of a deliberation process, as seen in Chapter 1, §1.2.3.[434]

Hence, individuals who do not possess such sophisticated capacities have been excluded from the realm of agents who can be held morally responsible for their behaviour. These individuals, whom Strawson describes as "psychologically abnormal" or "morally undeveloped", include agents that are "warped or deranged, neurotic or just a child."[435] These may include psychopaths and human beings who are severely cognitively disabled. Animals are also typically included in that category.

When confronted with these cases, we usually adopt what Strawson calls the "objective stance", which includes attitudes like frustration, relief, sadness, or disappointment, as pointed out by Kelly McCormick, but not reactive attitudes.[436] In the case of a morally undeveloped person, taking the objective stance toward her often entails seeing her "as an object of social policy", as an individual who must be "managed or handled or cured or trained" but not as a member of the moral community.[437] That person cannot enter interpersonal relationships and

[432] Shoemaker (2007), p. 75.
[433] Wallace (1994), p. 6.
[434] Shoemaker (2007), p. 76, Shoemaker (2013), p. 99, Rosen (2015), p. 79, Hieronymi (2007), p. 110, and Wallace (1994), p. 13.
[435] Strawson (1962), p 66.
[436] McCormick (2022), p. 62.
[437] Strawson (1962), p 66.

understand the expectations of good will they involve.[438] In the case of children, Strawson acknowledges that the line between objective and reactive attitudes might be blurred. We might sometimes regard them as morally responsible agents, as apt targets of reactive attitudes, and sometimes not.[439]

To sum up, Peter Strawson and several philosophers after him all seem to acknowledge that agents' moral responsibility practices rest on three capacities:

(1) The capacity to recognise good will or ill will in others.
(2) The capacity to form interpersonal relationships with others.
(3) The capacity to expect others not to show ill will and show good will toward oneself.

As already noted, the first and third components are strongly related to agents' *epistemic* and *self-control* capacities. To recognise good will or ill will in others, one must be capable of having access to wrongdoing. Similarly, expecting others to show good will requires recognising others as individuals who can understand expectations and adjust their behaviour accordingly.

4.1.2. What Is Quality of Will?

Reactive attitudes are thus responsive to agents' quality of will. In the case of ill will, that notion often overlaps with other concepts such as "wrongdoing", "wronging", or "desire for what is wrong", as already highlighted in this chapter and Chapter 3. The notion of "quality

[438] *Ibid.*, pp. 64-66.
[439] *Ibid.*, p. 67.

of will" can also have more specific meanings. According to Shoemaker, who provides a comprehensive and helpful summary of the philosophical literature on Strawson's approach, three different definitions of that concept have been put forward since the publication of Strawson's seminal article.

First, agents' quality of will can be grounded in the quality of their judgements. That account, which Shoemaker calls the "Pure Quality of Judgment theory", has been developed by Thomas Scanlon, according to whom we can blame agents if they have the capacity for "critically reflective, rational self-governance" but fail to govern themselves in light of some standards of judgement.[440] In that case, their capacity for reflective self-control would be faulty, and we would blame them for such defects by feeling approval or disapproval toward them.[441]

Second, reactive attitudes can be responsive to the quality of agents' character. Shoemaker calls that approach the "Quality of Character" view. According to that interpretation, the notion of quality of will is closely tied to agents' character traits: their way of acting, feeling, and thinking.[442] Hence, when we hold others morally responsible in that sense, we are likely to feel contempt, disdain or admiration for them, as these emotions take others' character as an intentional object.[443]

Third, agents can show poor quality of will by disregarding others and their perspectives. Shoemaker calls that account the "Quality of Regard" theory, which he associates with Michael McKenna's account of moral responsibility.[444] Regard should not be understood in highly rationalist terms and should not be conflated with judgements about moral reasons. As Shoemaker notes, regard is best understood as "a kind of attitude or stance one takes

[440] Shoemaker (2015), pp. 10-11. See also Scanlon (1998), p. 174.
[441] Shoemaker (2015), p. 26.
[442] Shoemaker (2013), p. 107 and Shoemaker (2015), p. 12. See also McGeer (2012), p. 305.
[443] Shoemaker (2015), p. 13.
[444] Shoemaker (2015), pp. 11-12. See also McKenna (2012), p. 59.

towards others or morality". It involves taking others into consideration when acting.[445] For example, if someone intentionally steps on my foot, that person has disregarded me, has *slighted* me. Reactive attitudes associated with that account include anger-type emotions and, more precisely, "feelings of intense heat and aggression, thoughts about slights, and motivational impulses to revenge or retribution."[446] As a response to good will, agents could be prone to "feelings of warmth, thoughts about beneficence, and motivational readiness to show appreciation, or return kindness, to the benefactor(s)."[447]

It is worth saying more about the important role Shoemaker places on empathy in his summary of the Quality of Regard view. There are three main ways in which we can show a poor quality of regard for others, which all consist of dismissing or overlooking others' perspectives: (1) by failing to see relevant facts about others' perspectives as reasons for acting in a certain way, (2) by judging that these reasons are in fact not reasons, or (3) by judging that these reasons are less important than others. All involve faulty empathy, according to Shoemaker, who writes that disregarding someone "is at bottom a *failure of empathy*."[448] (emphasis added). These defects can take two forms: (1) a failure to take others' perspectives, to see them as reasons for regarding them in a certain way, or (2) a failure to be affected by others' plights or joys, to resonate with them emotionally. Shoemaker calls the former "evaluational empathy" and the latter "emotional empathy".[449] To have a quality of regard and be held morally responsible to some extent, an agent must be capable of one of these two types of empathy.[450]

---

[445] *Ibid.*
[446] Shoemaker (2015), p. 26.
[447] *Ibid.*
[448] Shoemaker (2013), p. 115.
[449] Shoemaker (2015), pp. 99-101.
[450] *Ibid.,* p. 113.

It is easy to see how the Quality of Regard theory is relevant to my argument. Many animals can feel empathy, either basic or cognitively sophisticated. By extension, they could be said to regard or disregard others' emotions. This is a possibility that David Shoemaker himself acknowledges when he writes that this theory could be extended to "humans or (some) nonhuman animals"[45], although he does not specify which nonhuman animals and does not develop that claim further. Shoemaker is mainly concerned, in *Responsibility from the Margins* (2015), with how this account could help us make sense of our ambivalent responses to the behaviour of psychopaths, autistic people, and humans with intellectual impairments.[451]

We can go back to the three main components of Strawsonian accounts of moral responsibility. If we apply them to the Quality of Regard theory and cases of ill will, a moral agent engages in moral responsibility practices when:

(1) She can have access to "wrongdoing", "wronging", or "ill will". Here, such notions are defined as disregard for others' perspectives.

(2) She can form interpersonal relationships with others.

(3) She can expect others not to show ill will. Agents who are expected to show good will are those who are capable of empathy, either evaluational or emotional, or both.

We could be even more specific here and apply this account to the intentional infliction of suffering, which is mentioned by Shoemaker as a paradigmatic example of disregard for

[451] *Ibid.,* pp. 158-162 and pp. 183-186.

others[452] and is central to the arguments developed in my thesis. I suggest that a moral agent engages in moral responsibility practices when:

(1) She can have access to "wrongdoing", "wronging", or "ill will", at least regarding the intentional infliction of suffering.
(2) She can form interpersonal relationships with others.
(3) She can expect others not to intentionally cause suffering.

Let us explore these three components and how they can apply to many animals, including the foxes in Magdalen College.

4.2. Animals' Moral Responsibility Practices: A Minimal Account

4.2.1. Animals can have access to the wrong-making features of causing suffering.

In Chapter 3, I defended the view that animals can have access to the wrong-making features of causing suffering. The argument I proposed previously is straightforward and can be captured by the following syllogism:

(P1) An empathetic animal can recognise the intentional infliction of suffering and
(P2) Can have access to the badness of others' suffering, which
(C) Allows her to have access to the wrong-making features of causing suffering.

[452] *Ibid.,* p. 12.

This argument, which partly rests on my discussion of emotional contagion in my second chapter, serves as an essential building block of the central thesis I aim to defend in this chapter. If an animal possesses some minimal epistemic capacities and has access to the wrong-making features of causing suffering, it makes sense for others to hold her morally responsible.

4.2.2. Animals can form interpersonal relationships with others.

Furthermore, social animals engage in various relationships with other group members and form very strong bonds with some individuals. This fact is a platitude, if not a tautology. These ties can consist of parent-offspring relationships, relationships among siblings, dominance relationships, etc. Moreover, they involve various activities such as grooming, sleeping, playing, food sharing, eating, and hunting. Some of these activities require high cooperation, coordination, and reciprocity in exchanging food and "services" such as grooming.[453] I shall come back to animals' capacity for reciprocity in my thesis conclusion. It is also widely admitted that animals can recognise others very early in life, especially littermates, and are aware of their own relationships with other group members.[454] Great apes like chimpanzees even show a high degree of social awareness. They can understand others' relationships, and not just their own.[455]

It is worth noting here that although most wild animals engage with conspecifics, they can also form close relationships with members of other species. These cases are rarer for wild

[453] See Bekoff & Allen (2002), Bekoff & Pierce (2009), pp. 59-60, Langergraber, Mitani & Vigilant (2007), Kappeler & van Schaik (2006), Wilkinson (1984), Clutton-Brock (2002 and 2009), Cheney (2011), and Ligon (1983).
[454] See Bekoff (1981), p. 314.
[455] Goodall (1986), pp. 570-571.

animals[456], but animals living in sanctuaries for farmed animals offer an interesting example of such an interspecies community. Farmed animals can often be observed eating, playing and sleeping together, among many activities. This type of animal group also raises the question of whether these interactions between unrelated individuals can be described as friendship, which is taken to require mutual caring, intimacy, and shared activity.[457] I shall not enter the debate on the nature of friendship here.

Animals' close relationships set the ground for two components that are essential to the Strawsonian approach to moral responsibility, as previously seen. First, animals are likely to be *emotionally involved* in their relationships and to feel a form of affection for other members of their group. Such emotional commitment is, in turn, necessary for someone to feel reactive attitudes. Second, animals' interpersonal relationships could give rise to specific expectations about how they should treat each other. These may include expectations related to food sharing, grooming, or playing. To go back to this chapter's main example – the intentional infliction of suffering – a fox in Magdalen College who frequently plays with a conspecific could expect her playmate to behave in a certain way. She could expect her playmate not to engage in real aggression and intentionally inflict suffering on her. A playmate could also be expected to signal the intention to play, avoid rough playing, and adjust to the other's mood to continue playing. I shall now expand on these ideas.

[456] For instance, social play has been observed between members of different animal species: between huskies and polar bears, between gorillas and golden Labradors, between bat-eared foxes and gazelles, between elephant calves and wildebeests, between cows and foxes, and between red-necked wallabies and magpies. See Balcombe (2006), p. 83.
[457] See Helm (2021) and Korsgaard (1992), pp. 307-308.

4.2.3. Animals can expect others not to intentionally cause suffering.

So far, I have highlighted how animals can form complex relationships and have only mentioned the possibility that social animals might be aware of others' expectations. I believe animals' capacity for anger (§4.2.3.1) and empathy (§4.2.3.2) allows me to argue that animals engage not only in moral responsibility practices with each other, but also in *appropriate* moral responsibility practices grounded in animals' epistemic and self-control capacities.

4.2.3.1. Anger and Expectations in Animals

As we have seen, Strawson considers anger to be a reactive attitude. But we need to be more precise here about what anger is and how it can be relevant to exercising moral agency. We can do so without defending the stronger and controversial thesis that anger is *necessary* in our moral lives or a constructive emotion.[458] Instead, I shall stress three components of anger that allow us to account for its relevance to moral agency, especially to moral responsibility practices.

First, anger is usually defined in philosophy as an umbrella emotion that covers several emotional phenomena: irritation, intense rage, resentment, indignation, and the like.[459] Second, all these different types of anger share a specific way of appraising their object as not meeting an expectation.[460] Third, not all types of anger are moral. Feeling utter rage directed at an old laser printer does not have a moral dimension, even though that emotional state is aroused by the perception that an expectation was not met: the expectation that the printer will print what I want to print. In that case, my expectation is merely predictive and closely tied to goal

[458] For an overview of the debate, see Neblett (1979), Bommarito (2017), Thomason (2020), Srinivasan (2017), Cherry (2022), and Kauppinen (2018).
[459] Prinz (2009), p. 523.
[460] Macnamara (2011), pp. 85-86, Gibbard (1990), p. 126 and Taylor (1975), pp. 396.

frustration, but it is not *normative* or *moral*.[461] It does not concern how moral agents *ought* to behave toward each other and meet *moral* expectations such as respect for others' well-being and rights. So, what evaluations are involved in moral anger? These are usually evaluations of others' behaviour, which is appraised as violating moral expectations.[462]

Moreover, anger has been well documented in animals through the study of aggression. It is often taken to be one of the basic emotions both human and nonhuman animals have in common, along with joy, fear, disgust, etc.[463] Furthermore, biologists and ethologists agree on two basic theses about the nature of anger. The first is that anger interacts with the perception of a threat to one's interests, whether it be an interest in eating, mating, protecting one's offspring, or protecting one's physical integrity. In other words, animals feel angry at each other when there is a clash of interests.[464] According to Jane Goodall, the causes of anger in great apes are very easy to understand: "A tries to snatch B's fruit; B attacks A. This is a simple case of competition for food."[465] In examples involving intentional infliction of suffering, the aggression can go as follows: A bites B after using a play signal; A bites B a second time but without using a play signal; B attacks A in return. In other words, animals get angry at each other when an expectation is not met: the expectation to eat one's fruit, the expectation to play, etc. These could be predictive in nature, related to the way others *will* behave, or normative, related to the way others *ought* to behave. I shall return to the nature of these expectations in §4.3, as one could object to my argument that animals' expectations are not normative or moral but merely predictive and that their anger is not of the moral type. For

[461] On the distinction between predictive and normative expectations, see Macnamara (2011), pp. 85-86. It is worth noting that our anger at inanimate objects like printers could be moralized if, in our anger, we ascribe human qualities to them. In that case, our anger would be misplaced, for printers do not possess the moral qualities that make it appropriate to blame them.
[462] Gibbard (1990), p. 45, Cogley (2014), p. 202 and Ben-Ze'ev (2000), p. 380.
[463] On that point, see Ekman (1992).
[464] On that point, see Camperio Ciani (2000).
[465] Goodall (1986), p. 319.

now, I just need to stress that animals' anger fits the general definition of anger as an emotion aroused by the violation of an expectation.

The second thesis on which scientists agree is that anger is communicative. Animals express anger to engage in a fight or to prepare for aggression. The expression of anger – growling, flattening one's ears, arching one's back, etc. – constitutes a warning. It is an invitation for others to change their behaviour before a real fight occurs.[466] Anger thus serves a communicative function. It also further supports the claim, put forward in Chapter 3, §3.2.1, that animals can recognise intentional action in others, for an angry animal clearly signals her intention to attack, and others can understand the message anger conveys.

Animals get angry at each other. This fact is undisputed. What is less evident to biologists and ethologists is whether some animals could feel more sophisticated forms of anger, like indignation and resentment, which Strawson includes in the category of reactive attitudes. These constitute "cognitively sharpened" forms of anger, according to most philosophers of emotion.[467] Indignation is defined as anger that concerns a third party, as anger on behalf of a victim.[468] When one feels indignation, it is usually directed at a wrongdoer who violated a moral expectation in its way of treating another individual.

Resentment, like indignation, is often taken to be a reaction to the violation of a moral expectation, but in the context of dyadic interactions. It does not concern a third party. That said, philosophers disagree on whether there can be nonmoral resentment, for instance, resentment grounded in the belief that an etiquette expectation was not met or resentment

[466] Camperio Ciani (2000).
[467] See, for instance, Shoemaker (2015), p. 112, D'Arms & Jacobson (2003), p. 143, Talbert (2014), p. 289, and Wallace (1994), p. 48.
[468] Shoemaker (2015), p. 88, McKenna (2012), p. 66, Hampton & Murphy (1988), p. 18, and Scanlon (1988) p. 161.

that does not rest on propositional attitudes.[469] Yet they generally agree that resentment is different from anger outbursts in many regards. In his discussion of resentment and anger, philosopher Aaron Ben-Ze'ev argues that although both anger and resentment can be emotional reactions to wrongdoing, resentment is "a long-term attitude" that targets "a general pattern of action".[470] It is more diffuse over time. Resentment can also be different in terms of motivational outcome and cognitive structure. It may lead agents to protest wrongdoing perceived as a *moral injustice* instead of punishing the wrongdoer for one occurrence of personal harm. Finally, Ben-Ze'ev describes resentment as less intense and more general than anger.[471]

There is limited evidence that some animals could feel indignation and resentment. Let us look at a case of rough playing and third-party intervention in chimpanzees. The story is summarised by Jane Goodall in her magisterial book *The Chimpanzees of Gombe* (1986):

> In 1975 Freud was often rough during play with other youngsters. When, as frequently happened, he made the younger Prof scream, there was sometimes a series of reprisals that involved all members of both families (Prof's sister threatened Freud, Freud screamed, Fifi threatened Pom, Passion ran to support Pom). Twice the upshot was a fight between Fifi and the higher-ranking Passion (which Fifi lost).[472]

As stressed in Chapter 3, §3.2.3, third-party intervention is frequent in great apes, whether it be to protest the ill-treatment of infants, to intervene in fights, or to console others. In the example described by Goodall, adult chimpanzees' intervention seems to be motivated by a form of anger aroused by the perception of another chimpanzee inflicting suffering on a youngster. This anecdote, along with numerous studies on third-party intervention in great

469 Roughley (2017), pp. 272-273 and Shoemaker (2015), pp. 88-89.
470 Ben-Ze'ev (2000), p. 396.
471 *Ibid.*
472 Goodall (1986), pp. 570-571.

apes[473], renders plausible the hypothesis that these animals could feel a form of indignation. Their anger could be extended to group members who are not closely related to them, who are not their young.

In the same vein, several primatologists argue that great apes could *resent* each other. Chimpanzees often take several hours before reconciling after a fight. Some of them have even been observed engaging in "false reconciliation" that escalated into a second fight, suggesting that chimpanzees might hold grudges.[474] Although Ben-Ze'ev denies we can attribute resentment to animals, great apes seem to satisfy some criteria associated with his definition of resentment, namely, its longer-lasting dimension and lower intensity. I shall leave this possibility open here.

There are three ways in which animals' anger, whether it be simple anger, indignation, or resentment, supports the view that social animals engage in moral responsibility practices. The first is that this behaviour could provide additional evidence that they can have access to the wrong-making features of causing suffering. In Chapter 3, I defended Mill's views on the natural origins of justice by highlighting how empathy can enable animals to have access to the wrong-making features of causing suffering. However, I have left animals' desire for punishment unaddressed. As mentioned in this chapter introduction, Mill and many other philosophers posit the view that to describe an action as wrong involves seeing the wrongdoer as liable to responsibility responses like punishment and blame. Anger could play an important evolutionary role in emotionally underpinning that thesis.

To see how, we need to examine the way Mill and many other philosophers have traditionally defined anger: as an emotion that involves a desire for "payback." Indeed, Mill

[473] For a summary of the evidence, see Aureli & de Waal (2000), p. 263, pp. 266-270, and p. 289. See also Tajima & Kurotori (2010).
[474] Aureli & de Waal (2000), p. 178.

defends the now highly contentious view that anger is *necessarily* retributive in nature and is followed by a desire for retaliation.[475] To take Mill's views and apply them to my account of animals' access to wrongdoing, anger always involves a desire to make the wrongdoer "pay" for acting wrongly. Animals' anger, because it often motivates animals to confront the wrongdoer or warn her about the wrongness of her actions, could be seen as involving a desire for revenge and could further support the thesis I defended in Chapter 3.[476] If a fox in Magdalen College can have access to the wrong-making features of causing suffering and is attacked by an aggressive playmate, it is entirely plausible that her anger serves as additional evidence that foxes could react the wrong-making features of intentional infliction of suffering.

There is a second way in which the study of anger in animals adds weight to my argument: animals' anger could be interpreted as a form of blame. I have previously remarked that it is uncontroversial that anger, either in human or nonhuman animals, serves a communicative function. As noted by philosopher Zac Cogley, bodily cues associated with anger "can be observed, responded to, or ignored" and "provide an opportunity for engagement."[477] When human moral agents express anger, they send others two messages: that they disapprove of how others treat them and that they want others to change their behaviour. These evaluations may be explicitly moral in some cases: a wrongdoer has acted wrongly toward me, and the wrongdoer should understand my anger and the moral evaluations it involves.[478]

---

[475] For an overview of the debate on the nature of anger, see Darwall (2006), Shoemaker (2015), p. 92, Shoemaker (2013), p. 116, and Nussbaum (2016), pp. 5-6.
[476] This view is also supported by some limited studies on chimpanzees' vengefulness. For an example, see Jensen, Call & Tomasello (2007).
[477] Cogley (2014), p. 213.
[478] *Ibid.*

Although animals' anger does not come with *explicit* moral judgements like the anger of most adult human beings, it might still be a response to the wrong-making features of intentional infliction of suffering and share with human anger its communicative function. In other words, anger is both backward-looking and forward-looking: it is a reaction to a past action and an invitation for others to change their behaviour. It serves as a signal of disapproval of others' actions and, as such, might qualify as a basic form of blame. That type of blame could also allow various developmental stages and be a response to different types of harm: harm done to oneself, one's offspring, other in-group members like it is the case with chimpanzees,[479] and finally, out-group members.

The third way in which examining anger in animals supports my central thesis concerns anger's intensity. Indeed, the intensity of animals' emotional reactions in cases of conflict shows that social animals can form strong interpersonal relationships with other group members, as emphasised in §4.2.2. We have also seen in §4.1.1 that Strawsonians ground moral responsibility practices in interpersonal relationships, which provide the emotional engagement required to take a participant stance and feel reactive attitudes like anger.

For several decades, ethologists and psychologists have studied the effects of conflicts on animals' emotional lives. All agree that conflicts and the unpleasant emotions they cause are very stressful to animals, including human beings. Primatologists have observed self-directed behaviour, such as increased scratching, that indicates high stress levels during conflicts.[480] Interestingly, primatologists have observed more intense stress symptoms in primates who had fought with an individual with whom they had a stronger bond. According to Frans de Waal and Filippo Aureli, this suggests that primates are "concerned about

[479] On that point, see also the study conducted by Silk, Brosnan, Vonk, Henrich, Povinelli, Richardson, Lambeth, Mascaro & Schapiro (2005).
[480] Aureli & de Waal (2000), p. 202. See also Silk (2002) and Koski (2015).

disturbance of their longer-term relationships."[481] This also renders their need for reconciliation higher. I shall come back to the importance of reconciliation in animals in §4.3, for it allows us to clarify the nature of animals' expectations.

4.2.3.2. Animals' Epistemic and Self-Control Capacities: The Role of Empathy

Animals' anger could show that they have access to wrongdoing, blame each other, and value their relationships with others. Yet stressing the relevance of animals' anger to moral responsibility practices does not fully capture the appropriateness of such practices. Earlier in this thesis (§4.1.1 in particular), I have highlighted how philosophers who work on moral responsibility practices agree that two types of capacities are needed to make an individual liable to reactive attitudes like anger: epistemic and self-control capacities. Empathy could play that double epistemic and self-control role.

In Chapters 2 and 3, I argued that animals' empathy plays a crucial epistemic role in enabling them to have access to the badness of suffering and the wrong-making features of causing suffering. Hence, it is highly plausible that animals could satisfy Shoemaker's Quality of Regard account, in which empathy plays a fundamental role. Applied to others' suffering, empathy allows animals to gain insight into the emotional states of others by feeling their suffering.[482]

Animals' empathy could be extended to anger because animals can recognise not only when others are in a state of suffering, but also when they are angry. Shoemaker writes that morally responsible moral agents possess "the capacity to understand emotional communication (for example, to understand that the angry face one sees is an expression of

[481] Aureli & de Waal (2000), p. 235.
[482] For a similar point, see Roughley (2017), p. 277.

anger and communicates a demand for acknowledgement)" in addition to the capacity to understand others' perspective and respond emotionally to their plights.[483] Responsible moral agents usually understand others' suffering and anger, either by emotionally resonating with them, by taking their perspective, or both.[484]

Since the publication of Darwin's *The Expression of the Emotions in Man and Animals* (1897), there is much evidence that animals can recognise anger in others by relying on bodily cues typically associated with anger.[485] These may include baring teeth, growling, flattened ears, etc. By observing anger's typical forms of expression, animals can *perceive* or *infer* that another individual is in a state of anger. This thesis is also confirmed by numerous studies carried out on domesticated animals. Evidence pertaining to the emotional capacities of cats, dogs, horses, and goats suggests that these animals may not only respond to bodily cues of human beings they could see as threatening, but also form a cognitive representation of their inner emotional states.[486] In other words, animals do not merely respond to superficial bodily cues of angry human beings – they most likely recognise anger. The same line of argument developed in Chapter 2, §2.2.1, thus applies to the case of anger. Empathy allows animals to recognise various emotions in others, whether it be suffering or anger, and whether it be by directly *perceiving* these emotional states through their bodily expressions or by making *inferences* about others' emotional states.

Empathy could also foster animals' capacity for self-control and enable them to adjust their behaviour to others' anger. The motivational dimension of empathy I have taken to be unnecessary to account for empathy's epistemic role in Chapter 2, §2.2.2, could have an

[483] Shoemaker (2013), p. 119.
[484] *Ibid.,* p. 118.
[485] See Darwin (1872), chapters IV, V and X.
[486] For an overview of the evidence, see Quaranta, d'Ingeo, Amoruso & Siniscalchi (2020).

important *self-control* role to play. For instance, philosopher Stephen Darwall places empathy at the heart of reciprocity because empathy, either basic or cognitively sophisticated, allows individuals to appreciate others' perspectives.[487] Similarly, Daphne Brandenburg has argued that agents must possess some emotional capacities that can foster "the development of moral responsiveness" and behaviour regulation to be held responsible for their actions.[488] In the same vein, Shoemaker emphasised how empathy is needed for individuals to comply with others' expectations of regard.[489] Summarizing empathy's epistemic and self-control role and applying it to the case of promise-keeping, Shoemaker writes:

> I then, crucially, imagine what it must have felt like for you at that moment by stepping into your emotional shoes (emotional empathy), that is, by thinking about how I would feel had this been done to me (where I share your cares); I come to fully appreciate, via this empathic leap, the anger you are now feeling at me, and I now feel it too in sharing both your perspective and your concern about promise-keeping; I then maintain this anger at myself in returning to my own perspective; and this self-directed anger is a crucial constituent of my guilt, *motivating me, as the source of your hurt, to beat myself up, to grovel, and/or to do what I can to restore our relationship*.[490]
> (emphasis added)

This, though, is a summary of moral responsibility practices among full moral agents. Participants in Shoemaker's example are capable of sophisticated perspective-taking, emotional contagion, metacognition, and guilt. Such a picture is unlikely to apply the same way to nonhuman animals. Empathy's motivational role might take a different form in them.

[487] Darwall (2017), pp. 292-293.
[488] Brandenburg (2019), pp. 174-177.
[489] Shoemaker (2007), p. 72, p. 96 and p. 98.
[490] Shoemaker (2015), pp. 110-111.

In Chapter 2, §2.2.2, I argued against philosophers who partly ground empathy's epistemic role in its motivational outcome. That said, I have acknowledged that emotional contagion, and emotions more broadly, bear a relationship with motivation. Because they are felt as pleasant or unpleasant, emotions are "pushy". This means that when feeling empathy, animals could catch others' emotions, and these could be pleasant or unpleasant and even highly aversive in the case of suffering. This could, in turn, lead animals to do something to change the situation.

As we have seen in Chapters 2 and 3, this thesis is confirmed by studies on how others' suffering is deeply aversive for animals and how it leads them to withdraw aggression. Social play in animals supports the view that empathy plays an essential role in fostering animals' capacity for self-control. When playing, animals constantly monitor others' emotions, adjust to their playmate's strength, and use play signals to avoid aggression escalation. Empathy at work in social play not only enables them to have access to the wrong-making features of causing suffering. It is also an important milestone for self-control. As ethologist Elisabetta Palagi and her colleagues write, "during play animals (including humans) can learn how to better manage their competitive and aggressive drives."[491] This is especially true of animals who engage in very competitive play, like coyotes, and whose play sessions can quickly escalate into real fights.[492] Empathy seems to be required for self-control in animals, as it allows them to adjust to others' emotions and interests, express their own emotional states, expect others to understand them, and maintain peace among group members. These capacities also serve animals' selfish interests, such as the interest in playing, which must require empathy, cooperation, and emotional fine-tuning in order to be satisfied.

[491] Palagi, Cordoni, Demuru & Bekoff (2016), p. 1208.
[492] *Ibid.*, p. 1200 and p. 1203.

This thesis finds further support in the empirical literature on children's moral development. As seen in Chapter 2, §2.4, infants can feel others' emotions. More interestingly, their capacity to catch others' suffering and identify ill-intentioned human beings seems to foster moral responsibility practices. Children as young as 21 months recognise harmful intentions in others and even withdraw treats from them in experimental settings[493], and children's refusal to engage in positive interactions with wrongdoers is well-established by the age of three years.[494] This behaviour has been described by several child psychologists as a form of punishment, suggesting that toddlers already expect others to behave in a certain way and engage in responsibility practices aimed at influencing others.[495] This view is also shared by child psychologist Martin Hoffman, for whom empathy is congruent with a sense of justice similar to the one mentioned by Mill. According to Hoffman, empathy underlies moral emotions such as indignation, an emotion he describes as "empathetic anger", and fosters the rectification of violations.[496]

Hence, empathy could constitute the first step toward sophisticated moral understanding, moral judgement, reflective self-control, and self-directed emotions like guilt and shame, which we expect from full-fledged moral agents. By expressing suffering and anger and by being empathetic, both human and nonhuman animals can feel others' emotions, understand that these are emotional reactions to wrongdoing and expectation violation, and influence others' behaviour toward a peaceful equilibrium. Their practices might qualify as moral responsibility practices.

---

493 Hamlin, Wynn & Bloom (2011).
494 Vaish, Missana & Tomasello (2011), p. 129, Zelazo, Helwig, & Lau (1996), p. 2478, and Weiner & Peter (1973), p. 306.
495 *Ibid.*
496 Hoffman (2000), p. 226.

4.3. One Objection and Response

If the arguments developed in Chapters 2, 3, and 4 are successful, it is plausible that animals engage in moral responsibility practices. The objection according to which animals are too limited in their epistemic and self-control capacities to hold others responsible or being held responsible *at all* seems to lose its force, especially regarding the intentional infliction of suffering.

Still, one could object that my argument does not show that animals' anger is aroused by the perception that a *normative* or *moral* expectation was violated. As already mentioned, anger can be moral or non-moral. Although it is usually provoked by the perception that an expectation was not met, expectations can be of various types.[497] For instance, expectations can also be merely *predictive*. In that case, anger would be caused by the perception that expectations about how others *will* behave are not met. Anger would then be best characterised as goal frustration.

Applied to the argument I developed in this chapter, the objection would run as follows: When an animal is being caused to suffer by another individual, her anger could be compatible with other explanations like goal frustration and the violation of predictive expectations about how others *will* behave. Animals' anger cannot be interpreted as a responsibility response that concerns normative expectations, the way others *ought* to behave. Suffering through the actions of another could be seen as an impediment to predictive expectations and specific goals, such as playing. One of the foxes in Magdalen College desires to play and thinks her playmate *will* behave in a certain way but does not have expectations about how her playmate *ought* to treat her. This objection is also compatible with the ethology

[497] See, for instance, Wallace (1994), p. 35.

and psychology literature on animals' anger, according to which it is well established that animals can get angry out of goal frustration and express anger as a way of regaining control over their environment, but not as a way of *blaming* another individual for her behaviour.[498] Philosophers, too, sometimes posit that animals' anger cannot be normative or moral but rather arises from goal frustration.[499] Their angry reaction would be best described as an objective attitude, not a reactive one.

This objection rightly points out the complex link between normative and predictive expectations. If we think others *ought* to behave in a certain way and assume they understand such expectations, then we expect that they *will* act in a certain way. We both have normative and predictive expectations about their behaviour. That said, I think the objection fails in denying that animals' anger expresses normative expectations and even moral expectations related to others' well-being. Two examples allow us to answer that objection: cases of tolerance and reconciliation in animals.

Several species of animals exhibit great tolerance for young or disabled individuals. For example, de Waal notes that young chimpanzees' behaviour goes largely unpunished for the first four years of their lives and that juveniles enjoy a high level of tolerance from other group members. As de Waal explains, "they can do nothing wrong, such as using the back of a dominant male as a trampoline, stealing food out of the hands of others, or hitting an older juvenile as hard as they can."[500] Similarly, dogs, wolves, and coyotes are tolerant of young individuals who do not yet follow the "rules" of play, and adjust their play intensity to the age, size, and physical strength of youngsters.[501] Finally, the case of Knuckles provides an

[498] See Frijda (1986), Shoemaker (2015), p. 90 and Gibbard (1990), p. 130. On goal frustration in chimpanzees, see Goodall (1986), p. 321.
[499] See, for instance, Shoemaker (2015), p. 90.
[500] de Waal (2014), p. 189.
[501] Bekoff (2004), p. 505.

interesting example of social tolerance among animals. Knuckles is a chimpanzee living at the Center for Great Apes in Florida and suffering from cerebral palsy, a rare condition that led him to be physically limited and perhaps mentally impaired as well. Anthropologist Devyn Carter, who has observed Knuckles for years, details in *The Anthropologist's Dilemma* (2020) how other chimpanzee residents seemed to be sensitive to "his lack of awareness" of others' expectations. Knuckles' behaviour went largely unnoticed, and his presence was tolerated.[502]

These examples suggest two things. First, animals could be capable of distinguishing harmful behaviour motivated by genuine ill will from actions done out of curiosity, clumsiness, or ignorance of expectations, thus further supporting the view I developed in Chapter 3, §3.2.1, according to which animals can recognise intentional infliction of suffering. That said, tolerance among animals is also compatible with other hypotheses that do not appeal to animals' capacity to recognise ill will in others. Instead, tolerance could be explained more simply by animals' awareness of others' physical and cognitive characteristics rather than their capacity to recognise intentional action in others. In the same vein, tolerance could be an attitude that adults unreflectively adopt to take care of youngsters and favour their inclusion within the group.

Second, cases of tolerance add weight to the view that animals see others not simply as obstacles to the fulfilment of a goal. Others are likely perceived as individuals who *can* or *cannot* meet some normative expectations, depending on their level of development and the capacities they possess or do not possess. Interestingly, this also raises the question of whether some animals can see young or disabled individuals as targets of objective attitudes but not reactive attitudes.

[502] Carter (2020), pp. 35-38.

Another practice among animals renders plausible the view that animals likely have normative expectations about how others ought to behave: reconciliation. Ethologists define reconciliation as affiliative behaviour between two or more individuals that is longer than usual and occurs after conflicts.[503] Such affiliative behaviour consists of kisses, hugs, mouth licking, nose nudges, coat licks, head rubs, or body rubs, depending on the species.[504] The Chimpanzee Sanctuary Northwest, for instance, features on its website several stories and videos of reconciliation involving its residents. Let us cite the summary of Gordo's and Willy B's peace-making process. The two chimpanzees reconciled after trying to bite each other's fingers off the day before:

> At breakfast the next morning, Gordo – who was uninjured in the fight and whose "side," you could say, came out on top – approached Willy B to reconcile. Breathy panting serves to express friendly intentions, and Gordo offers both his backside and his fingers and toes to Willy B. Isn't it ironic that the way to make up after a fight in which you tried to bite each other's toes off is to place your toes in each other's mouths again?[505]

Although the most specular examples of fights and reconciliation attempts concern great apes, reconciliation has been observed in *all* species of social mammals and even in some species of social birds, like ravens and crows.[506] It is particularly well studied in nonhuman primates, bottlenose dolphins, and dogs.[507] Some animals, such as gorillas, chimpanzees, bonobos, baboons, and pigtail macaques, even console victims, appease aggressors or act as

[503] For a summary, see Walters, King, Scolaro & Shyan-Norwalt (2020).
[504] *Ibid.*
[505] Chimpanzee Sanctuary Northwest (2023).
[506] de Waal (2023), Fraser & Bugnyar (2011), Sima, Matzinger, Bugnyar & Pika (2017).
[507] Walters, King, Scolaro & Shyan-Norwalt (2020).

mediators.[508] They thus use sophisticated "positive peace" strategies, which ethologists define not as the mere absence of conflicts but as active attempts to make peace after a fight.[509] Animals remember conflicts, which are stressful to them, and actively try to reconcile.

Reconciliation allows me to answer the objection discussed in this section. Indeed, studies on reconciliation indicate that many animals, at least social mammals, have normative expectations about how others treat them. These expectations can concern the intentional infliction of suffering and be moral in regard to their content, unlike conventional expectations. If animals got angry out of pure frustration of predictive expectations, they would not actively try to make peace with each other, nor would they feel angry at each other for several minutes, if not several hours. They would simply feel a form of anger akin to the type of frustration that was observed in animals who cannot fulfil their goals because of the conditions of their physical environment.[510] Other group members would be seen merely as individuals who frustrate their goals, just as printers could frustrate human beings' goals, but not as individuals with whom they *can* make peace and *ought* to make peace to restore relationships.[511] Reconciliation in animals can thus be interpreted in truly *normative* terms: as a form of commitment to a relationship, a way for individuals to recognise the expectations that

[508] Goodall (1986), p. 571 and Aureli & de Waal (2000), p. 263, pp. 266-270, and p. 289.
[509] See Palagi, Cordoni, Demuru & Bekoff (2016), p. 1196.
[510] For a summary of the empirical evidence, see Lewis (1999), p. 19 and Pannewitz and Loftus (2023).
[511] One could argue that reconciliation in animals provides evidence not only that animals could have normative expectations toward each other, but also that they may be capable of another reactive attitude, at least in its basic form: forgiveness. In philosophy, forgiveness is often defined as an attempt to move beyond negative reactive attitudes such as resentment. For instance, Jeffrie Murphy writes that forgiveness presupposes a change in feeling: "the overcoming, on moral grounds, of the intense negative reactive attitudes—the vindictive passions of resentment, anger, hatred, and the desire for revenge—that are quite naturally occasioned when one has been wronged by another responsible agent." According to Jean Hampton and Jeffrie Murphy, motives for forgiving a wrongdoer typically involve acknowledging that the other has repented for her wrong actions or "had a change of heart", that she did mean well after all, that she has suffered enough, that she had been sufficiently humiliated, or that she had been a good friend in the past. In the case of animals, we do not know which of these judgements underpins reconciliation. It is also unclear whether we can describe reconciliation among animals as forgiveness, for there can be reconciliation without judgements of forgiveness, and vice versa. I shall not expand on these remarks here. See Murphy (2012), pp. 6-8 and Hampton & Murphy (1988), p. 24.

bound each other and a social practice that entails the signalling of good intentions that could be akin to promise-making. After reconciling, animals thus have reasons to think that others not only *will* follow the expectations to which their relationships give rise, but also that others are aware that they *ought* to follow them.

That said, studies on reconciliation could also be compatible with a purely predictive and non-moral interpretation, according to which animals engage in reconciliation simply because they want to let others know about their future behaviour and foster predictive expectations. According to such a reading, Gordo the chimpanzee, when reconciling with Willy B, expresses something like "I will not bite your toes again", but not a more morally sophisticated statement that would entail the recognition of his relationship's value and of Willy B's normative expectations about how he ought to behave.

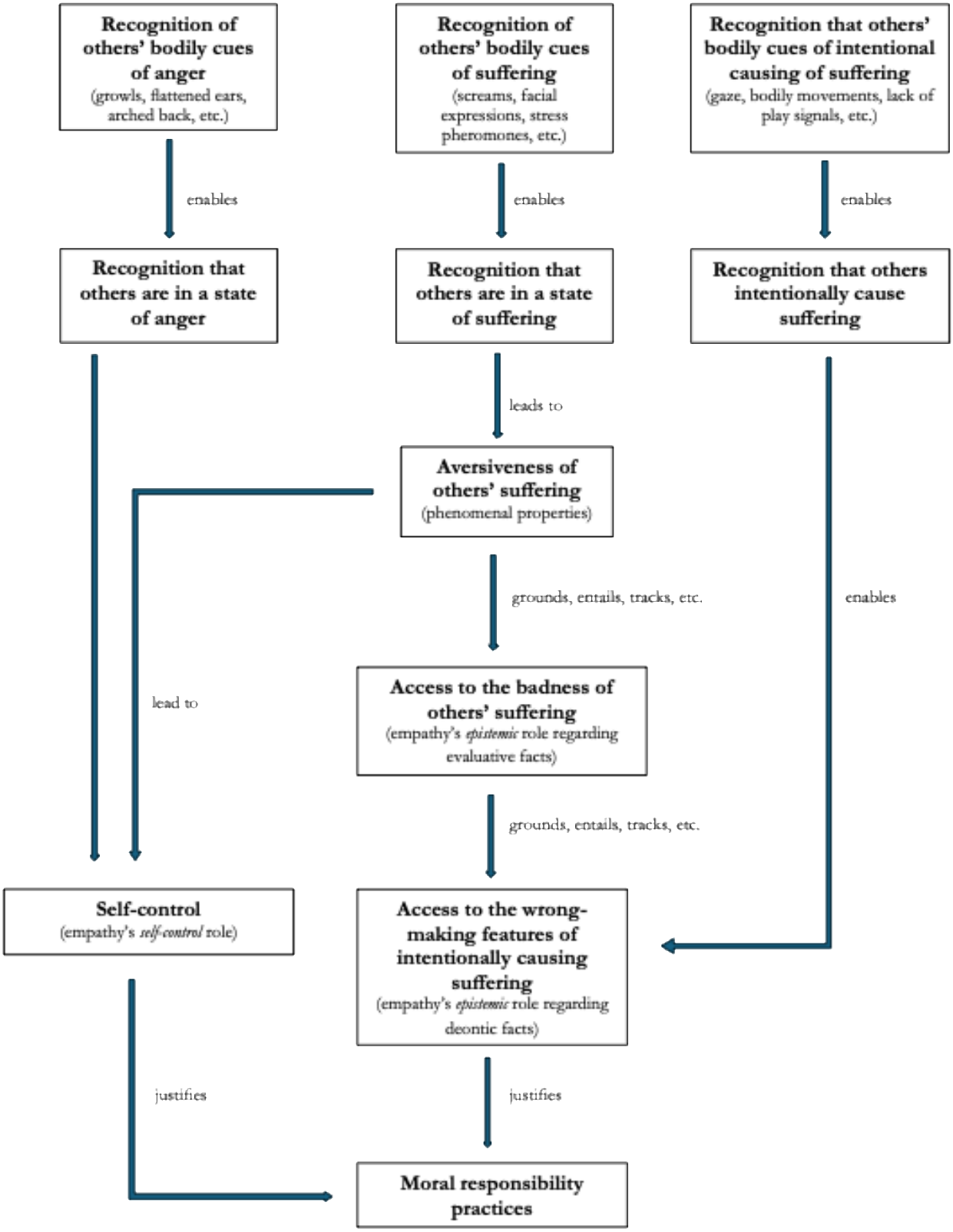
Yet interpreting reconciliation as a practice that entails not only predictive expectations, but also normative expectations, seems to be a more plausible option, for observations of reconciliation in animals provide evidence in favour of two other philosophical theses. First, animals place value on their relationships with others and the benefits they derive from them. They engage in strategies that allow them to restore and maintain relationships that have been damaged by wrongdoing, such as the violent biting of others' toes. This thesis is also strengthened by several studies that suggest that animals' need to reconcile is greater after a conflict with another group member with whom they spend much time and exchange important benefits, as briefly mentioned in §4.2.3.1. Second, reconciliation further supports the claim that animals possess some self-control capacities. These allow them to adjust to others' suffering, anger, and the expectations it expresses. Reconciliation and empathy might serve as milestones for reflective self-control and self-directed emotions, which are important capacities underlying humans' moral responsibility practices.

4.4. Summary of Chapter 4 and Concluding Remarks

In this chapter, I have argued that animals' capacities (1) to have access to the wrong-making features of causing suffering and (2) to form interpersonal relationships with other animals (3) give rise to expectations about how they ought to be treated. Animals can have access to the wrong-making features of causing suffering and form interpersonal relationships with other group members, which come with an array of expectations. This, in turn, makes them susceptible to *moral* anger, which Strawsonians describe as a reactive attitude that is constitutive of moral responsibility practices among moral agents. Hence, it is entirely plausible that social animals' emotional reactions could be described as moral responsibility practices grounded in animals' epistemic and self-control capacities. One of the Magdalen College foxes' angry reaction to her playmate's aggressiveness would likely qualify as blame.

We now have a complete picture of the animal moral agency account I developed in this thesis. My line of argument can be illustrated with the following diagram:

Summary of Chapter 4



In this chapter, I have also highlighted how animals' moral responsibility practices differ from ours and are more minimal. For instance, animals might be aware of others' quality of *regard* but might not satisfy more rationalist accounts of moral responsibility practices grounded in the Quality of Judgement theory. Similarly, animals' capacity to understand others' anger rests on empathy, which is consistent with the literature on the moral development of children and might constitute the first step to the more cognitively sophisticated forms of self-control and self-directed emotions we observe in most adult human beings. It is also unclear whether chimpanzees' anger could qualify as resentment or indignation, which are cognitively more sophisticated. Animals' anger does not necessarily entail explicit moral evaluations but is nevertheless responsive to the wrong-making features of causing suffering, namely, the fact that suffering is bad. The same conclusion applies to animal blame, which is taken to be a form of anger along the lines of Strawsonian approaches to moral responsibility.

So far, I have detailed how animals could engage in moral responsibility practices but have said very little about *our* moral responsibility practices involving domesticated animals. This was intentional. The implications of recognising some animals as moral agents for the way we treat them will be explored in my fifth and last chapter.

Chapter 5

What Are the Implications for Animal Ethics of Recognising Some Animals as Moral Agents?

(§5.2.2 was published in *The Plant-Based and Vegan Handbook*, Springer. See Simoneau-Gilbert 2024).

In the previous chapters, I explored how animals' moral capacities, especially their capacity for empathy, could allow them to have access to the badness of others' suffering and the wrong-making features of causing suffering, and rightfully hold each other responsible. If the arguments proposed in this thesis succeed, they open the door to profoundly changing how we see and treat animals, especially domesticated ones.

As noted by Will Kymlicka and Sue Donaldson (2012) and Christiane Bailey (2014), the moral patiency paradigm has enabled philosophers to argue that animals do not need to be moral agents or highly rational creatures to be included in the moral circle.[512] Many animals are sentient beings. As such, they have interests that should be included in our welfare calculations or should be more strongly protected by the granting of rights.[513] These arguments have been major contributions to animal ethics. However, they also came at the cost of important theoretical and practical limitations in regard to how we understand animal well-being, the way we can harm animals, and the type of relationships we ought to have with them. Recognising some animals as moral agents could solve these remaining issues while being entirely compatible with older arguments about the moral patiency of animals and our duties toward our fellow creatures. As mentioned in my thesis introduction, both categories are not

[512] See Donaldson & Kymlicka (2012) and Bailey (2014), p. 36.
[513] See, for instance, Singer (1975), Francione (2000), and McMahan (2008).

mutually exclusive, and most adult human beings are both moral patients and moral agents. The same conclusion applies to animal moral agents.

This last chapter offers an exploratory overview of the practical implications that come with recognising some animals as moral agents. My goal in this fifth chapter is not to provide a thorough argument in favour of animal moral agency like those in the previous ones. It is instead to highlight how the animal moral agency paradigm could lead us to change our views on how we ought to treat animals. More precisely, I discuss how recognising some animals as moral agents may inform our responsibility practices with them (§5.1). I will also describe how it widens our understanding of how we can harm animals, both subjectively and objectively (§5.2).

5.1. Holding Animals Responsible

5.1.1. Could We Rightfully Hold Some Animals Morally Responsible?

In the previous chapter, I took a Strawsonian approach to moral responsibility and argued that animals could engage in moral responsibility practices grounded in reactive attitudes such as moral anger. Indeed, it seems that many animals can (1) have access to the wrong-making features of causing suffering, (2) form relationships with other animals, and (3) expect them not to cause suffering intentionally. These expectations are typically expressed through anger, which empathetic animals can understand. Animals' capacity for empathy enables them not only to have access to the badness of others' suffering, but also to adjust their behaviour to others' emotions. But I have said nothing yet about how these three criteria could be extended to relationships between human beings and some animals. Let us briefly look at how they might cast light on our responsibility practices with them.

It is evident that human beings (1) can have access to the wrong-making features of causing suffering and can likely do so very early in life. In Chapters 2 and 3, I argued that this capacity could rest on empathy, which allows moral agents to have access to the badness of others' suffering, and the recognition of intentional action in others. Although my approach to moral agency makes room for recognising infants and many nonhuman animals as moral agents, I have highlighted how adult neurotypical human beings can exercise a more sophisticated type of moral agency that entails the capacity to grasp moral concepts and form moral judgements. This means human beings can recognise when animals intentionally inflict suffering on another individual and directly assess others' suffering as *bad*. For example, human beings can realise when dogs attack other animals out of genuine ill will rather than self-defence.

Moreover, human beings (2) can form close relationships with animals. This is especially true of companion animals, whom we can deeply love and trust, such as dogs, cats, and horses. But it is worth noting that human beings can also enter similar relationships with wild animals who became domesticated because they are ill or injured,[514] or with some highly intelligent birds. In *The Book and the Brotherhood*, Iris Murdoch recounts the intense bond between one of the main characters, Gerard, and Grey, an African grey parrot that Gerard's family had adopted when he was a child. Murdoch describes the young boy as loving the bird very profoundly, the animal's presence being "a continual source of trembling joy" and a "touchment". In the novel, Grey returns Gerard's affection with eyes expressing "fearless faith and love" and "tenderness."[515] Such a relationship patently provides the basis of mutual expectations Peter Strawson and other philosophers after him deemed necessary to account

[514] On that point, see Donaldson & Kymlicka (2011), p. 99.
[515] Murdoch (1987), p. 58.

for our reactive attitudes with others, as we have seen in Chapter 4, §4.1.1. Dogs, cats, horses, parrots, and perhaps even domesticated rats might also be the type of animals we can expect to behave in a certain way. This is because they possess the capacities that enable them to catch others' emotions, recognise intentional harm, and adjust to others' emotions and the demands they express.[516] I shall come back to this claim.

Before that, two things need to be noted. First, although companion animals are paradigmatic examples of animals with whom we have close interactions, some human beings can form similar relationships with farmed animals. We can think of people who work or volunteer at sanctuaries for farmed animals. Furthermore, it is plausible that farmed animals can meet the three criteria I proposed in my previous chapter. Chickens, pigs, cows, and goats can feel empathy, which serves as a milestone for more sophisticated epistemic and self-control capacities.[517] Evidence for empathy in farmed animals raises the question of whether these animals could understand our expectations of good will and respond to our expression of anger.

That said, these preliminary ideas concern only a small number of farmed animals, as most of them are not kept in conditions that allow them to develop and exercise their moral agency. Animals kept in factory farms do not form relationships with human beings grounded in the reciprocal recognition of each other's emotional states and the expectations that can be expressed through anger. These elements might explain why some farmers see these animals as targets of objective attitudes, as individuals to be "handled" or "managed", in Strawson's words, rather than participants in the moral community.[518] Because of the nature of such

[516] For instance, kea parrots and cockatiels have been found capable of emotional contagion. See Schwing, Nelson, Wein & Parsons (2017) and Liévin-Bazin, Pineaux, Clerc, Gahr, von Bayern & Bovet (2018).
[517] For reviews of the empirical evidence, see Marino & Colvin (2015), Marino (2017), Marino & Allen (2017), and Nawroth (2017).
[518] Strawson (1962), p. 66.

relationships and the limitations the industry imposes on animals, farm employees are not always encouraged to be empathetic toward animals under their care and see them as agents with beliefs, desires, emotions, and preferences.[519] By extension, they do not see them as moral agents who could be the targets of reactive attitudes. In fact, it might not be advantageous to do so. If farmers profoundly changed their outlook on farmed animals, this would perhaps lead them to seriously question the morality of animal farming and ditch that industry.[520]

Second, we typically do not form close relationships with wild animals and liminal animals. The latter are animals that are neither wild nor domesticated but live close to human buildings, such as the foxes in Magdalen College. Wild and liminal animals generally avoid human beings, and we do not form with them the strong relationships we have with companion animals.[521] This point was rightly highlighted by Asia Ferrin in her work on animal moral responsibility, as mentioned in Chapter 1, §1.2.1.[522]

But this lack of closeness does not mean we should see wild and liminal animals' behaviour as *amoral*. For instance, dolphins are often seen killing porpoises "for sport," and chimpanzees have been observed killing and eating chimpanzee infants or making war with chimpanzees from other groups.[523] During the chimpanzee wars detailed by Jane Goodall in *The Chimpanzees of Gombe*, some individuals were often left in agony for days or would have

[519] A recent study in social psychology suggests that categorizing animals as laboratory animals foster their mind denial. On that topic, see Vezirian, Bègue-Shankland & Bastian (2024).
[520] In fact, several farmers have turned to vegan farming after facing moral crises. Stories of farmers who came to see farmed animals as individuals and became vegetable farmers are featured in documentaries such as *The Last Pig* (2017) and *73 Cows* (2018).
[521] The concept was first proposed by philosophers Sue Donaldson and Will Kymlicka in *Zoopolis* (2011). See Donaldson & Kymlicka (2011), especially Chapter 7.
[522] Ferrin (2019), p. 146.
[523] See Aitken (2023) and Goodall (1986), pp. 520-522.

met harrowing deaths if some researchers had not euthanised them.[524] Such behaviour was described by Goodall as "cruelty" and an evolutionary precursor to sadism.[525]

In the same vein, Susana Monsó argues in her article "Is Predation Necessarily Amoral?" (2021) that although predation is amoral most of the time, it is not always. Many wild animals are empathetic and could be aware that the animals they prey upon suffer. They qualify as "moral subjects", according to Monsó, and their moral capacities raise important questions pertaining to the morality of predation. For example, the behaviour of predators like killer whales, hyenas, lions, and chimpanzees that Monsó puts forward in her work could involve one or several of these four moral elements: (1) intense enjoyment caused or directed at the hunt, (2) killing methods that are painful and driven by the desire to prolong the prey's distress, (3) the use of various killing methods, which suggests animals could intentionally choose a particular method of killing, and (4) playful predation interactions.[526] These components support Monsó's view that some predators could be partly motivated by the pleasure they take in causing suffering and killing other animals. This, in turn, puts pressure on the view that predation is solely motivated by hunger and always done out of pure necessity, for animals' survival. If predators can feel empathy and be sensitive to the emotions of their prey but nevertheless enjoy the suffering they inflict on them, predation is not devoid of a moral dimension.

Does this mean human beings could be justified in holding wild and liminal animals responsible for their behaviour? It is unlikely to be sufficient according to Strawsonian theories, since these approaches do not *solely* rely on agents' moral capacities. It also depends on the nature of our relationship with other moral agents, the specific expectations to which

[524] Goodall (1986), pp. 507-513.
[525] *Ibid.,* p. 533.
[526] Monsó (2021), p. 371.

these relationships give rise, and whether individuals can understand the expectations expressed in anger. In the absence of a certain degree of mutual expectations and understanding of each other's emotions, we can perfectly describe the behaviour of predators as having a moral dimension but without being committed to the view that we should *hold* them responsible for it. Wild and liminal animals could instead be seen as moral agents who engage in moral responsibility practices that make sense in the context of their own communities, as already proposed by Marc Bekoff and Jessica Pierce.[527]

This thesis can be defended by appealing to the communicative function of anger I put forward previously, according to which anger could express a demand: that others change their behaviour. As already highlighted in Chapter 4, §4.2.3.2, wild animals can recognise anger in their conspecifics, act to de-escalate tense interactions and adjust to others' anger to preserve valuable relationships. Could they understand human anger and the demands it expresses in a way that would make sense for us to blame them? This thesis is implausible in the absence of a certain level of proximity and interactions with us. Most wolves could recognise *wolf* anger and the expectations it expresses, but not *human* anger, unless they are raised by human beings. In fact, wolves and coyotes that have grown up among human beings can follow human social cues in a very sophisticated way, as highlighted by philosopher Kristin Andrews.[528] But without such a close bond, it is unlikely they can.

Domesticated animals, especially companion animals, could meet my account's third criterion, namely, that (3) we can expect them not to intentionally cause suffering. Indeed, basic forms of empathy like emotional contagion have been observed in all species of birds and mammals, and several domesticated animals also show a sophisticated understanding of

[527] Bekoff & Pierce (2009), p. 144.
[528] Udell, Dorey & Wynne, (2008) and Udell, Spencer, Dorey & Wynne (2012). Cited in Andrews (2014), p. 162.

humans' emotional states and expectations. Animals like horses and rats are capable of empathy, and African grey parrots can show prosocial behaviour, although it remains unclear which capacity or mechanism underlies it.[529] Parrots can make associations between words and context, suggesting they could be sensitive to humans' verbal commands and expectations.[530]

The same conclusion applies to dogs, but to an even greater extent. Indeed, studies indicate that dogs can understand 30 words on average, including words referring to actions, prohibitions ("No!"), and praise and blame ("Good/bad dog!"). Some dogs who were more extensively trained can understand between 200 and 1,000 words.[531] Dogs were also found to rely on their understanding of emotional cues, like human beings' facial expressions or voice tone, to derive the general meaning of words and act to fulfil the demands they express.[532] Moreover, in their study assessing dogs' preference between food and praise, Peter Cook and his colleagues found that dogs were largely praise-motivated. Between being verbally praised and eating treats, 13 dogs out of 15 derived more pleasure in the former or enjoyed them both equally.[533] Finally, dogs have been observed making reconciliation attempts with human beings after being scolded by their caregivers. They usually tried to reconcile with humans by getting closer, gazing at them, wagging their tails, and showing appeasement behaviour.[534] These findings suggest that dogs can value relationships with human beings, understand basic expectations, and be motivated to please.

Cats, on their part, offer a different story. Indeed, some studies indicate they show important motivational defects when conforming to humans' demands. In their study of cats'

---

529 On African grey parrots' prosocial behaviour, see Krasheninnikova, Brucks, Blanc & von Bayern (2019). On empathy in mammals, see de Waal & Preston (2017).
530 For an overview, see Pepperberg (1999).
531 See Pongrácz, Miklósi & Csányi (2001), Kaminski, Call & Fischer (2004) and Pilley & Reid (2011).
532 Ratcliffe & Reby (2014), Ratcliffe, Taylor & Reby (2015), Prichard, Cook, Spivak, Chhibber & Berns (2018), Magyari, Huszár, Turzó, & Andics (2020) and Reeve & Jacques (2022).
533 See Cook, Prichard, Spivak & Berns (2016).
534 See Cavalli, Dzik, Carballo & Bentosela (2016). Also cited in Behdadi (2024), p. 6.

capacity to distinguish between words, Atsuko Saito and his colleagues found that cats can recognise their own names but are not always motivated to obey human beings' calls.[535] According to researchers, these results can be explained by cats' evolutionary history. Unlike dogs, cats' ancestors were not social pack animals and were not domesticated to follow human beings' commands. The domestication process was more akin to a form of cohabitation.[536] This means that cats simply did not form relationships with us or other felines grounded in a strong desire to please others.

However, these results are contested, given that we still know very little about cats' social and cognitive capacities compared to dogs.[537] The results of Saito's study do not mean that cats lack moral capacities. On the contrary, empirical studies suggest they can form strong bonds with their caregivers[538] and recognise human emotions.[539] But it seems that cats' emotional capacities do not always foster a more robust motivation to adjust to human beings' anger, hence further supporting the view I defended in Chapter 2, illustrated by *The Reactive Twin* case, that the capacity to feel the emotions of others might not always lead to a desire to help others or adjust to others' emotions. Cats' motivational shortcomings might come not from a total lack of empathy, but from their initial social organisation as solitary felines. It might also reflect the lesser value they place on following the expectations to which relationships give rise.

Several companion animals thus seem to possess the basic epistemic and self-control capacities that allow us to make sense of our moral responsibility practices with them. Various

[535] Saito, Shinozuka, Ito *et al.* (2019). Cats have also been found capable of learning the names of other cats. See Takagi, Saito, Arahori, *et al.* (2022).
[536] *Ibid.*
[537] For an overview of the current empirical literature on cats and avenues for future research, see Turner (2021).
[538] Vitale, Behnke & Udell (2019).
[539] Quaranta, d'Ingeo, Amoruso & Siniscalchi (2020).

companion animals can feel empathy, which, I have argued, grounds their capacity to access the badness of others' suffering and the wrong-making features of causing suffering, understand human beings' anger, and adjust to our emotional states. As noted by philosopher Paul Shapiro, a dog who arbitrarily attacks her family members would be liable to blame, and this can be explained by the fact that dogs can understand that attacking another individual causes suffering.[540] Although Shapiro's general thesis is plausible, it could be more refined by appealing to two elements: the epistemic and self-control role that empathy plays, and the communicative function of anger. It is insufficient that dogs and other companion animals are aware of the immediate consequences of their actions. I argue they must also be capable of understanding others' anger and the expectations it expresses. As highlighted in Chapter 4, §4.2.3.2, it is likely they can. This allows us to make sense of Shapiro's example fully: blaming a dog for biting her family members could be justified because she can both have access to the wrong-making features of causing suffering and understand the content of blame that human anger expresses.

5.1.2. Children and Companion Animals: Similarities and Differences

To clarify the nature of our responsibility practices with companion animals, it might be helpful to compare them to parents', caregivers', and teachers' moral responsibility practices with young children, including toddlers. Young children and companion animals are alike in two important ways.

First, they can be blamed for some actions but not for others. This thesis is uncontroversial among philosophers who agree that children could be moral agents to some degree. To apply this thesis to a Strawsonian account of moral responsibility, young children

[540] Shapiro (2006), p. 368.

and companion animals can likely *disregard* others' well-being and fail to empathise with others' emotions. They could thus be blamed for showing a poor *quality of regard* for others, which, according to Shoemaker, involves a "failure of empathy."[541] But it is unlikely they could satisfy other definitions of quality of will, such as the Quality of Judgement view, and be held responsible according to more demanding definitions of quality of will. As noted by Shoemaker, a moral agent can show a poor quality of judgement when she fails to govern herself according to certain standards of judgement.[542] This capacity rests, in turn, on the type of reflective self-control Kantian philosophers put forward, which animals and toddlers lack. For example, it would not make sense to hold companion animals and very young children responsible for acts of negligence, as understanding the consequences of one's negligence requires being capable of foreseeing the possible consequences and side effects of one's action and governing oneself in light of such information.

Hence, animals' and young children's limited moral capacities could explain why we might be justified in holding them responsible for some actions but not others. Clarifying the various definitions of quality of will and the required moral capacities that ground them also allows us to make sense of Strawson's remarks on the moral responsibility of children. According to Strawson, our responsibility practices with children are best described as involving both reactive and objective attitudes, as we cannot adopt pure reactive attitudes or, on the contrary, pure objective attitudes to make sense of our interactions with them. Strawson writes: "They [parents] are dealing with creatures who are potentially and increasingly capable both of holding, and being objects of, the full range of human and moral attitudes, but are not yet truly capable of either."[543] Our responsibility practices with children thus "represent a kind

[541] Shoemaker (2013), p. 115.
[542] Shoemaker (2015), p. 10-11. See also Scanlon (1998), p. 174.
[543] Strawson (1962), p. 75.

of compromise", in Strawson's words. If we appeal to Shoemaker's proposed typology to make sense of Strawson's views, we could be justified in blaming very young children for acts of *ill regard* but not for other forms of poor quality of will. Although Strawson excludes companion animals from the realm of individuals we could hold responsible for some actions, the arguments developed in this thesis suggest companion animals could find themselves in a similar position. They could be the targets of both reactive and objective attitudes, just like young children.

Second, the relationship most adult human beings have with children is similar in some regards to the one they have with companion animals. In both cases, the relationship is characterised by love and trust, but also inequality in participants' moral capacities.[544] Hence, blame directed at toddlers and companion animals could be mostly forward-looking in the sense that its underlying purpose is to foster moral agency in beings who are not full-fledged moral agents. This thesis on the forward-looking role of reactive attitudes is relatively uncontroversial among Strawsonians, although philosophers disagree on whether or not children are *really* responsible for their actions. For instance, Gary Watson writes that children lack the deep understanding of demands that we expect from full-fledged moral agents but that we might be justified in *holding* children morally responsible to foster the development of their moral capacities.[545] In the same vein, Susan Dwyer highlights how children could be responsive to blame and how their moral capacities, like pro-social behaviour and sensitivity to norms, could ground our moral responsibility practices with them.[546] Similarly, Cristina Traina argues that one does not need to possess sophisticated moral capacities like reflective

[544] For a similar point, see Scanlon (2008), p. 156.
[545] Watson (2004), p. 229-230. On a similar point, see also Wallace (1994), p. 13.
[546] Dwyer (2003), p. 188.

self-control to be liable to blame.[547] Finally, Daphne Brandenburg argues that because children fall between moral patients and full-fledged moral agents, they might be held morally responsible for their behaviour, but through reproach rather than blame. According to Brandenburg, blame is a response to full moral agents who can consider the morally relevant features of situations but intentionally choose to discard them.[548] By contrast, the purpose of reproach is to make children responsive to the morally relevant features that they would not have seen otherwise.[549]

The same forward-looking direction of blame also underlies our responsibility practices with companion animals. This is especially true of *young* companion animals like puppies, who are unaware of how human beings expect them to behave and are gradually held to certain standards of conduct. In *Adam's Task* (1986), animal trainer Vicki Hearne details the process of training dogs, cats, and horses to meet humans' expectations. According to Hearne, the training process can be grounded in mutual love and respect, and dogs who follow their trainers' commands are committed to mutual expectations.[550] Our use of praise and blame when training companion animals thus rests on consequentialist considerations: bringing them to respond to morally relevant features. For example, we might praise dogs for comforting distressed individuals or blame them for growling at children out of ill will. Such responsibility practices aim to make them attentive to the emotional states of other sentient beings and encourage them to exhibit a good quality of regard toward others. As Hearne notes, the training process creates a "society" between human beings and companion animals.[551] In the training process of dogs, Hearne writes, "a moral transformation" occurs as animals become

[547] Traina (2009), p. 23.
[548] Brandenburg (2019), p. 180.
[549] *Ibid.,* p. 174.
[550] Hearne (1986), p. 30.
[551] *Ibid.*, p. 67.

"trustworthy" and autonomous. They are gradually expected to behave "responsibly".[552] Our close relationship with companion animals and the training process to which they are submitted give rise to specific expectations. Hence, there is a sense in which companion animals can do otherwise and can be held morally responsible in a more retributive sense. Once they reach adulthood, dogs could be blamed not just for forward-looking or instrumental reasons, but also for backward-looking reasons. This is also the case with children who gradually learn to meet moral expectations and can be held responsible both for backward and forward-looking reasons.

However, there is an important way in which our moral responsibility practices with children and companion animals differ. Notice the vocabulary that Hearne uses to describe our relationship with dogs, cats, and horses: the one of *training*. This contrasts with the notion of *moral education* philosophers typically use to describe children's moral development process. To morally educate their children, parents rely on the use of moral concepts to explain why children have reasons to do or refrain from doing certain things.[553] Moral education is not mere training. The aim of caregivers is not to foster children's empathy or affect children's emotional reactions to make them respond to morally relevant features as we could do with dogs and other domesticated animals. Rather, the goal of moral education is to bring children to see how the circumstances call for certain reactions, how their emotions can be assessed in light of moral principles, and why certain actions are right or wrong. This requires sophisticated use and understanding of human language, among several capacities.

[552] *Ibid.* pp. 24-25 and p. 21.
[553] See, for instance, Halstead & McLaughlin (Eds.) (1999) and Halstead (2010), among many. For an overview of the philosophical literature on moral education, see Chazan (2022). On the moral training of domesticated animals, see also Garthoff (2019).

Companion animals like dogs can understand basic words, including praise and blame, which makes it appropriate for human beings to train them and hold them morally responsible. But they are limited in their moral agency in a way that children are not. Companion animals can adjust their behaviour by being attuned to others' emotions, but they will never reach the same level of moral development as children, who can early in life use concepts such as "good", "right", "bad", and "wrong", feel self-directed emotions like guilt and shame, and reflect on their motivations for action. Animals might be outperformed very quickly by children and even by children as young as two or three years. As such, they might tentatively be compared to people with important cognitive disabilities who, like children and animals, learn to respond to others' expectations but who have permanent limitations in their moral capacities.

Shoemaker explores the case of intellectually impaired people in *Responsibility from the Margins* and, more specifically, the case of human beings who are disabled in their capacity to "engage in abstract thought or to apply principles or information from one situation to another very well."[554] These people, Shoemaker argues, can still show regard for other people because they are capable of empathy. By extension, they could be held responsible for showing a poor quality of regard.[555] Because they can be empathetic, cognitively disabled people can form a strong emotional bond with their caregivers and respond to their praise and blame reactions. In Shoemaker's words, cognitively disabled people can be "empathy trained".[556] Such interactions also provide the basis for expanding people with intellectual impairments' relational and moral capacities so that they can appropriately respond to other people's

[554] Shoemaker (2015), p. 183.
[555] *Ibid.,* pp. 185-186.
[556] *Ibid.,* p. 188.

expectations and emotions, not just their caregivers'.[557] I want to suggest that the same line of argument could apply to domesticated animals, although Shoemaker does not explore that possibility in his work.

5.1.3. Could Companion Animals Hold Us Responsible?

Before examining the relationship between animal moral agency and well-being, it is worth highlighting how my account of our moral responsibility practices toward animals leaves the door open to acknowledging *mutual* expectations between us and them. If animals are moral agents and can expect a certain treatment from us, it is entirely plausible that their emotional reactions could arise from the perception that their expectations were unmet. This is a possibility that some Strawsonians have acknowledged, but not Strawson himself. As mentioned in my previous chapter, David Shoemaker states that his Quality of Regard account could be extended to some nonhuman animals but does not expand on that remark.[558] In *Moral Dimensions* (2008), Thomas Scanlon writes that our relationship with companion animals is one of "mutual trust and affection" and can be impaired by animals' conduct. If a dog's actions show a lack of care for human beings, she could be liable to blame.[559] Finally, Stephen Darwall writes that we could see an infant's or animal's cry as a form of complaint, although young children and nonhuman animals lack the rational and second-personal capacities to make demands as members of the moral community.[560]

In Chapter 4, I detailed how anger is an emotion that is widespread among animals, how it can interact with the perception that others have not met a certain expectation of good

[557] *Ibid.*
[558] Shoemaker (2015), pp. 158-162 and pp. 183-186.
[559] Scanlon (2008), p. 166.
[560] Darwall (2006), p. 29, pp. 39-40, and p. 43.

will, and how animals' expectations are likely normative. These observations also raise the following questions: If companion animals are moral agents, can they expect a certain quality of will from their human guardians? Can they get angry at us and blame us for failing to show good will? Very recently, philosopher Dorna Behdadi (2024) has argued they can and has highlighted how infants' and dogs' anger toward us could be interpreted as a form of blame.

The current evidence also supports the view that animals can get upset if threatened or ill-treated by others, as mentioned in Chapter 4. This means that companion animals could recognise the intentional infliction of suffering from human beings and react strongly to it. As noted by Behdadi, animals could see human beings' behaviour as a threat, get angry at us, and ask us to adjust to their desires.[561] For instance, a dog who does not want to be petted usually flattens her ears and growls in warning. According to Behdadi, this could be interpreted as blame, as an emotional state expressing a demand that the person change her behaviour before a conflict occurs.[562] This thesis is also entirely compatible with the literature on the evolutionary function of anger I emphasised in Chapter 4, §4.2.3.1.

However, we do not have evidence supporting the view that dogs, cats, and other companion animals could feel more cognitively sharpened forms of anger like resentment, nor that they could "sulk", at least to my knowledge. The question of companion animals' expectations and anger is largely empirical, so I shall leave these observations as mere hypotheses here. Nevertheless, they raise important questions pertaining to the link between animals' moral capacities and well-being and how disappointing animals' expectations could constitute a form of subjective harm. Let us now explore these new issues.

[561] Behdadi (2024), pp. 6-10.
[562] *Ibid.,* pp. 9-10.

5.2. Harming Animal Moral Agents

Recognising animals as moral agents affects the way we define animal well-being and assess how it can be impacted by human action. Here, I am concerned with how the animal moral agency paradigm may lead us to widen our understanding of how we can harm animal moral agents. Harm can be understood in a more subjective sense for proponents of hedonism or desire satisfaction theory (§5.2.1), but it can also be defined in more objective terms if one wants to defend an objective approach to animal well-being (§5.2.2.). Recognising some animals as moral agents likely has important implications for *all* theories of animal well-being.[563]

5.2.1. Subjective Harm

Taking into account the moral capacities of animals has important implications for subjective theories of animal well-being. This is especially true of hedonism, which defines well-being in terms of the experience of pleasure and the absence of suffering. Recognising animals as moral agents may bolster our understanding of how we can make animals suffer, which is not limited to mere physical suffering or psychological suffering caused by extreme boredom, for example.

Traditionally, animal ethicists have criticised our current treatment of domesticated animals by appealing to these forms of suffering.[564] But as pointed out by Susana Monsó, Judith Benz-Schwarzburg, and Annika Bremhorst, philosophers who underestimate the moral capacities of animals run the risk of overlooking some specific ethical problems raised by some industries: those in which animals often see other individuals suffer, catch others' distress, but

[563] Similarly, there are several related questions I shall not answer here, such as whether desire satisfaction and objective list theories of animal well-being could collapse into hedonism, or to what extent pleasure contributes to animal well-being, especially compared to non-hedonic goods. For a summary of these issues, see Rice (2015) and Fischer (2022).
[564] See, for instance, Singer (1975), Regan (1983), Francione (2000), and McMahan (2008).

cannot act to help others. For example, sows witnessing the castration of piglets or the ill-treatment of other sows might feel distressed themselves and, without being given the possibility of alleviating others' suffering, could increasingly suffer through emotional contagion.[565] The same observation applies to animals kept in tiny enclosures for experimentation or for their meat, eggs, milk, skin, and fur. We can think of chickens, pigs, sheep, cows, foxes, minks, mice, rats, rabbits, and primates. It is widely admitted that such animals are empathetic creatures, either toward their young, in the case of chickens, or other group members.[566] Several industries thus harm animals when forcing them to witness the suffering of their conspecifics.

Acknowledging empathetic animals' suffering has three significant consequences for animal ethics. First, although Monsó, Benz-Schwarzburg, and Bremhorst are concerned with hedonism, one could argue that such practices are problematic because they frustrate animals' desire to help others in distress, which likely stems from their capacity for empathy. The same conclusion could thus apply to desire satisfaction theories of animal well-being. Second, considering animals' capacity for emotional contagion could strengthen utilitarian arguments in favour of veganism, in particular those grounded in hedonism as a theory of well-being. It seems that farmed animals' suffering could have been underestimated by philosophers so far, rendering even less plausible the thesis that human beings' pleasure in eating meat could outweigh farmed animals' suffering. Third, the general line of argument could be applied to animal rights theories if one wants to argue that animals' interest in avoiding suffering should be more strongly protected. Indeed, one could make the case that industries that force animals to see others in distress further violate their right not to be harmed, whether we define harm

[565] Monsó, Benz-Schwarzburg & Bremhorst (2018), pp. 291-293.
[566] For a summary of the empirical evidence, see Marino (2017) and de Waal & Preston (2017).

as suffering or desire frustration. Philosopher Tom Regan, for instance, places animals' right not to be harmed at the heart of his theory of animal rights, along with animals' right not to be treated as mere means.[567]

But there is another way, not mentioned by Monsó, Benz-Schwarzburg, and Bremhorst, in which the moral capacities of domesticated animals could be relevant to our understanding of their suffering, especially suffering caused by disappointed normative expectations. This omission could be explained by the fact that the authors endorse the view that animals are moral subjects – individuals who can have emotions that track moral propositions – but not moral agents who can have expectations toward others and hold them responsible for their behaviour.[568]

To see how positing the moral agency paradigm is of relevance here, we need to look again at the nature of our relationship with domesticated animals and the expectations to which it gives rise. In his article "Betraying Animals" (2019), philosopher Steve Cooke argues that some animals *trust* human beings and do not merely *rely* on them for food or care and that animals or young children do not need to entertain explicit judgements of trust to be said to trust their caregivers.[569] Some practices, like animal experimentation, can even be rendered possible by animals' cooperation, since it often involves training animals before conducting studies. As noted by Cooke, experimenters have acknowledged that animals need to *trust* human beings to cooperate with them in experimental settings.[570]

The trust domesticated animals place in us, along with their dependence, raises important new questions about the wrongness of breeding and killing them for profit or to

[567] Regan (1983), especially sections 7.6, 7.8 and 7.9.
[568] Monsó, Benz-Schwarzburg & Bremhorst (2018), pp. 291-293.
[569] Cooke (2019), pp. 187-191.
[570] *Ibid.*, p. 184.

use them for experimentation. These are not fully captured by notions of harm, rights, or obligations, according to Cooke.[571] They have to do with the systemic betrayal of animals' trust, which exemplifies important moral defects. According to Cooke, we can rightfully criticise individuals who consistently betray animals' trust on the basis that, in doing so, they exhibit untrustworthiness.[572] More recently, Judith Benz-Schwarzburg, Susana Monsó and Ludwig Huber (2020) have adopted Cooke's main argument to account for the morality of our relationship with dogs.

Yet there is a certain sense in which the notion of harm, especially harm done to moral agents, is relevant to account for the wrongness of betraying animals. To see how, let us examine an extract of George Eliot's first novel, *Adam Bede* (1859), which features several scenes between Adam, a carpenter, and his dog Gyp:

> Hitherto Gyp had kept his comfortable bed, only lifting up his head and *watching Adam more closely* as he noticed the other workmen departing. But no sooner did Adam put his ruler in his pocket, and begin to twist his apron round his waist, than Gyp ran forward and looked up in his master's face with *patient expectation*.[573]
> (emphasis added)

In the extract, Gyp expects Adam to feed him, and Adam later satisfies Gyp's expectation.[574] Indeed, Eliot describes Adam as having a soft heart for his dog and always taking great care of him. But suppose it was not the case, and Adam was instead portrayed as a sadistic human being. Suppose that instead of feeding Gyp, Adam had turned to him and hit him with his ruler out of sheer cruelty. Such action would have been morally reprehensible for

[571] *Ibid.,* p. 183.
[572] *Ibid.*, p. 184.
[573] Eliot (1859), p. 10.
[574] *Ibid.*, p. 11.

several reasons. Cooke would perhaps argue that it would have been wrong because it would have constituted a betrayal of Gyp's trust. It would have badly reflected on Adam's character, thus revealing his untrustworthiness, among numerous vices.

Another reason why this action would have been wrong is because it causes unnecessary suffering to a sentient creature. Undoubtedly, Adams' blow would have been painful and distressing to Gyp. But if we assume that domesticated animals can trust us, and if we defend the view that they can be moral agents, expect a specific treatment from us, and blame us for failing to exhibit good will toward them, Gyp's suffering would likely have been caused by *two things* here: Adam's physical abuse and his failure to uphold Gyp's "patient expectation", which partly stems from the nature of their relationship. Adam's treatment would have been *doubly* distressing to Gyp. Perhaps, if Eliot's novel had taken a different direction, Gyp would have even reacted angrily when confronted with Adam's cruelty. The same line of argument applies to children who can have expectations about how their caregivers will and ought to treat them.

Moreover, my remarks are compatible with the empirical evidence on reconciliation in dogs I mentioned earlier, which suggests that conflicts with human beings are stressful to them and that dogs use peace-making strategies. Hence, arguing that some domesticated animals can be moral agents and engage in responsibility practices with us allows us to cast light on the various ways in which we can harm them. Cooke, Benz-Schwarzburg, Monsó and Huber only briefly mention that type of harm, which the animal moral agency paradigm makes patent.[575] Hence, the types of subjective harm we inflict on animals expand beyond direct physical abuse. It can be caused by various forms of ill-treatment: exposure to others'

[575] See Benz-Schwarzburg, Monsó and Huber (2020), p. 13.

suffering, failure to allow animals to fulfil their desire to help others, and betrayal of animals' expectations.

5.2.2. Objective Harm

There is another sense, more objective, in which human beings could harm domesticated animals who are moral agents. The notion of objective harm is often closely connected to objective theories of animal well-being that link well-being to Aristotelian notions such as "the good life" or "flourishing". These concepts are often described in terms of animals' natural capacities, which vary across species.[576] For instance, flying might be part of what it means for some species of birds to live a good life, but not for species of land mammals, who do not possess that capacity.

The most influential objective account of animal well-being is Nussbaum's capability approach, which rests on the following core idea: for each animal species, there exists a set of basic capabilities that are available to the members of such species and that are essential for animals to flourish and live a good life. The list of basic capabilities includes life, bodily health, bodily integrity, senses, imagination, thought, emotions, practical reason, affiliation, relationships with members of other species, play, and control over one's environment.[577] According to Nussbaum, animals are entitled to develop and exercise these capabilities. Animals' capabilities should be protected more strongly by rights.[578]

Following Nussbaum's approach, Susana Monsó, Judith Benz-Schwarzburg, and Annika Bremhorst (2018) have argued that Nussbaum's theory could be extended to the

[576] For an overview of theories of animal well-being, including objective theories, see Rice (2015), especially pp. 381-384 for objective theories.
[577] Nussbaum (2006), Chapter 6.
[578] *Ibid.*

development of moral capacities. Indeed, we could understand animals' moral capacities as basic capabilities if we define them as character traits, as "dispositions to feel and behave in certain ways" rather than mere contingent motivations or cognitive-affective mechanisms that give rise to moral behaviour.[579] An empathetic pig is one who is more consistently *disposed* to suffer when others suffer. As noted by the authors, this interpretation is compatible with Nussbaum's list of capabilities, which includes emotions and affiliations.[580]

According to Monsó, Benz-Schwarzburg, and Bremhorst, it is plausible that moral capacities possess both instrumental and intrinsic value in regard to animals' well-being. For example, empathy is instrumentally valuable "because of the good it brings to the world".[581] It can motivate animals to help others in need and increase their well-being. It could also be intrinsically valuable if one sees it as a character trait, as a moral virtue that should be cultivated for its own sake and be part of the definition of the "good life". If we adopt the view that empathy has value in itself, regardless of its good consequences on sentient creatures, this allows us to argue that animal well-being should not be understood merely as depending on subjective mental states such as pleasure. A complete account of animal well-being would also include capabilities like moral ones and would comprise both hedonic and non-hedonic goods. This means that a pig who is empathetic would be seen as having a better life than a pig who is aggressive. The empathetic pig possesses a higher number of goods that are part of what makes a life good for a pig.

An objective list theory of animal well-being like Nussbaum's applied to animal moral agents would have important implications for our understanding of how we can harm animals. In their article, Monsó, Benz-Schwarzburg, and Bremhorst argue that the harm we can cause

[579] Monsó, Benz-Schwarzburg & Bremhorst (2018), p. 295.
[580] *Ibid.,* p. 296.
[581] *Ibid.*

to animals may not be entirely captured by hedonistic and desire satisfaction theories, which define harm in terms of subjective states such as suffering or desire frustration. According to the authors, we could harm animals by hindering the exercise or development of their moral capacities.[582] Such harm can be caused by practices that involve denying animals the possibility to exercise their moral agency.[583] We can think of industries that require keeping animals in small enclosures, such as the crates used in industrial farms, zoos, and laboratories. Some industries also entirely preclude animals from developing their moral capacities, as is the case with the training of dogs and bulls used for animal fights, which encourages aggressiveness in animals rather than empathy. Some experiments involve keeping laboratory animals in total social isolation to study the psychological effects of confinement.[584] In such settings, animals cannot develop their moral capacities if they are not given the opportunity to interact with others in the first place.

I think Monsó, Benz-Schwarzburg, and Bremhorst develop an important and insightful argument that accounts for the wrongness of certain practices. I wish to briefly mention two ways in which their argument could be developed further. First, it could be applied to our positive duties toward animals, especially domesticated animals. If we take moral agency to have value, either instrumental or intrinsic, in the lives of animals, it seems that human beings could have the positive duty to foster its development. For instance, training dogs to be empathetic could be the morally right thing to do for several reasons: because it allows dogs to help others in need, because it enables them to bond with other animals, or because empathy has intrinsic value.

---

582 *Ibid.,* p. 297.
583 *Ibid.,* p. 302.
584 *Ibid.,* p. 303.

In the same vein, philosophers Sue Donaldson and Will Kymlicka (2011) argue that human beings should act in a way that encourages domesticated animals' agency. This could be done by giving them more access to public spaces like off-leash parks and public transportation or to sanctuaries for farmed animals.[585] In *Zoopolis* (2011), Donaldson and Kymlicka broadly construe the notion of animal agency as being related to animals' capacity to make choices.[586] That said, their arguments on including domesticated animals in the political community could be applied to account for animals' *moral* agency more specifically, for off-leash parks and sanctuaries are precisely the type of spaces that allow animals to develop their moral capacities. In a dog park or sanctuary, animals can play with other animals, form meaningful relationships, and learn to adjust to others' emotions and expectations. This could give us reasons to support their implementation.

Second, the argument developed by Monsó, Benz-Schwarzburg, and Bremhorst could account for some limitations in the way animal ethics has been developed so far. It is common in some strands of animal ethics, especially in works that address the wrongness of killing animals, to reflect on how death is bad for some animals, for example, for pigs, by appealing to the capacities possessed by a neurotypical pig, the average lifespan of a pig, and the goods that death forecloses for members of that species.[587] That said, authors have often overlooked how animals' capacities, including moral ones, partly form through social interactions, either positive or negative. To put it differently, it seems that animals' capacities are not just a matter of being a member of a certain species and not being intellectually impaired at birth or after an accident. Animals' moral development very closely interacts with the relational opportunities they are given. Indeed, it is widely agreed in biology and psychology that social

[585] Donaldson and Kymlicka (2011), p. 113 and p. 121.
[586] *Ibid.,* pp. 65-66.
[587] See, for instance, Singer (1993), Chapter 5, McMahan (2002), Chapter 3, and McMahan (2008).

isolation has tremendous negative effects on animals, especially in regard to their overall health, resistance to stress, and social competence.[588] For instance, rats and mice who were kept in social isolation for several weeks or months generally show lower capacities for social recognition, spatial memory, willingness to learn, and stress resistance.[589] These rats and mice showed a lower level of cognitive development than other rats and mice despite being members of the same species and not showing any sign of permanent intellectual impairment prior to the studies. This raises, in turn, new questions about how animals' sociality and its influence on the development of their capacities should be considered in philosophical works that address the wrongness of killing animals.

To sum up, human beings could objectively harm animals by either precluding them from exercising their moral capacities or thwarting their development, and these could be necessary for animals to flourish. A good life for a pig could be one in which she can develop her empathy, share others' joys and plights, and engage in meaningful relationships that rest on recognising mutual expectations. This ideal is far from the current reality of pigs raised for their meat in industrial farms.

5.3. Summary of Chapter 5 and Concluding Remarks

Taking animals' moral capacities seriously can have significant consequences on the way we think about the well-being of animals. In this exploratory chapter, I have offered an overview of the main practical implications of recognising some animals as moral agents. I have concentrated on our moral responsibility practices with companion animals (§5.1) and how considering animals as moral agents can widen our understanding of how we can harm them,

[588] On that topic, see Bailey & Moore (2018) and Mumtaz, Khan, Zubair & Dehpour (2018).
[589] See Begni, Sanson, Pfeiffer, Brandwein, Inta, Talbot, Riva, Gass & Mallien (2020). Krupina, Shirenova & Khlebnikova (2020), and Magalhães, Mampay, Sebastião, Sheridan & Valente (2024).

subjectively and objectively (§5.2). These fall outside the scope of physical harm or suffering caused by extreme boredom, which has caught the attention of most animal ethicists since the 1970s.

The implications of the animal moral agency paradigm extend beyond issues related to our moral responsibility practices with domesticated animals and the main theories of animal well-being. They include questions in political philosophy that touch upon the political rights of animals and their inclusion in our political community. I shall briefly mention these avenues for future research in my thesis conclusion.

Conclusion

Summary

Did the altruistic rhesus monkeys studied by Wechkin, Masserman, and Terris do the right thing when refusing to inflict an electric shock on their conspecifics? Can Bataille the horse, Kuni the bonobo, the two foxes playing in the snow in Magdalen College, the chimpanzees observed by Goodall, Gyp the dog, and all the other animals mentioned throughout this thesis be moral agents? I think so.

The key takeaways from my five chapters are the following:

Chapter 1: Moral agency is best understood as a gradual and multi-faceted phenomenon that is not always closely tied to moral responsibility. This account of moral agency leaves the door open to recognising many animals as moral agents.

Chapter 2: Basic forms of empathy like emotional contagion, of which many animals are capable, allow them to have access to the badness of others' suffering.

Chapter 3: The capacity for emotional contagion and recognition of intentional action enables animals to access the wrong-making features of causing suffering to others. Empathy thus has an *epistemic* role in allowing them to have access to both evaluative (Chapter 2) and deontic facts (Chapter 3).

Chapter 4: Empathy also plays a *self-control* role in the moral lives of animals, as it enables them to adjust to others' anger and the normative expectations it expresses. If we describe anger as a form of blame, it follows that many social animals engage in moral responsibility practices.

Chapter 5: Recognising some animals as moral agents helps us make sense of our moral responsibility practices with domesticated animals and widens our understanding of the ways in which we can harm them, either subjectively or objectively.

Avenues for Future Research

I wish to close this thesis by highlighting two possible avenues for future research, which touch upon how the animal morality debate could allow us to account for the evolutionary origins of virtue ethics and contractarian theories of moral agency. Shedding light on animals' capacities for reciprocity could also have important implications for political philosophy, in particular for theories of justice in which reciprocity is a central component.

Virtue Ethics Theories of Moral Agency

The argument I developed in this thesis is limited to animals' epistemic access to the badness of others' suffering and the wrong-making features of causing suffering. In other words, it concerns what animals can understand certain states of affairs and actions. I did not explore how causing suffering could reflect well or badly on animals and how they could see each other more broadly. Could animals ascribe characteristics to other individuals?

The cases of tolerance and social play I described in Chapters 3 and 4 provide evidence for the claim that animals could be minimally aware of others' age, physical strength, and overall development. I also mentioned in §3.1 that dogs have been found to prefer individuals who help others and that American crows have been observed "scolding" harmful human

beings.[590] Capuchin monkeys also accept food less often from human beings who have previously refused to help others in need.[591] Moreover, in a field study on coyotes' dispersal, Marc Bekoff found that young individuals who displayed more aggressive behaviour and were less playful with their littermates were more likely to be avoided by others. Individuals who use play markers before engaging in a real attack ended up never being chosen as playmates, and this led them to be feared by other members of their group and, more dramatically, to be entirely excluded from the community. Aggressive coyotes formed less strong social bonds than their playful conspecifics and were more likely to be rejected by other group members.[592]

These findings raise the following question: Can animals have access to moral virtues and vices? If animals see another individual intentionally harm others, could their access to the wrong-making features of causing suffering also give them insight into the character traits of that other individual? Could repeated wrong actions track vices such as cruelty and untrustworthiness? Answering that question would require a detailed analysis that falls outside the scope of this thesis. A further step would need to be added to outline the relationship between bad states of affairs, wrong actions, and vices.

This question is, in turn, important for virtue ethics theories of moral agency, which are concerned with the definition of virtues and vices, how actions can reflect well or badly on agents, and how a virtuous moral agent would act.[593] It could constitute a promising avenue for future research, especially given that philosophers working on animal morality have been concerned with how human beings ascribe virtues or vices to animals but not how animals could be aware of others' character traits. As mentioned in Chapter 1, §1.2.1, philosophers

[590] See Chijiwa, Kuroshima, Hori, Anderson & Fujita (2015), Cornell, Marzluff & Pecoraro (2011), and the evidence cited in Zawidski (2013), p. 54.
[591] Anderson, Kuroshima, Takimoto *et al.* (2013).
[592] See Bekoff (1977), especially pp. 720-724.
[593] See Annas (2006) and Foot (1978).

Stephen Clark and Steve Sapontzis have argued that animals can be virtuous because they can act from stable character traits such as kindness and courage.[594] A dog we would call "a good boy" is empathetic, whereas a "bad" or "vicious" one is aggressive and does not show any concern for others' well-being. But little has been said so far in philosophy on how animals could see each other and how human capacities associated with virtue ethics theories of moral agency could be placed on the same continuum as animals' capacities.

Contractarian Theories of Moral Agency and Justice

Describing some animals as moral agents could also have consequences for contractarian theories of moral agency and, by extension, political philosophy. In fact, the implications of that debate might be more significant for traditional theories of justice that include reciprocity as an essential component than for animal ethics. Proponents of theories of justice as reciprocity generally agree on two fundamental theses: (1) only individuals who can participate in a cooperative scheme can advance claims of justice, such as claims to rights and resources, and (2) rationality is needed to understand the terms and conditions of such social bargain.[595] Only individuals who can understand the demands of reciprocity and choose to reciprocate can be the subject of duties of justice.[596]

These theories of justice have faced substantial criticisms over the last decades. Sceptics usually explore two lines of response. First, several authors have argued that because they conflate *by* and *for* whom principles of justice are designed, proponents of justice as reciprocity find themselves in a difficult position when facing the challenge of explaining why

[594] See Clark (1984), p. 107 and Sapontzis (1987), pp. 32-34 and p. 147.
[595] For a summary, see Nussbaum (2006), Chapter 1. See also Rawls (1963).
[596] Gauthier (1986), Chapter 1 and Rawls (1971), Chapter 1.

we do not have duties of justice toward individuals who cannot reciprocate.[597] Such individuals include animals, to whom we might have "duties of compassion and humanity" but not duties of *justice*, according to Rawls.[598] Second, some authors have claimed that animals could engage in relationships grounded in reciprocity with human beings and other animals.[599] Some animals could understand how others expect them to behave, comply with these expectations, and, in return, expect others to behave in specific ways.

The second response has yet to be developed in the philosophical literature, although the question of reciprocity in animals has been of great interest to ethologists and biologists since the 1970s.[600] Over the last decades, several social animals, mainly birds and mammals, have been observed exchanging goods and services, like food and grooming.[601] Moreover, animals typically feed or groom more frequently individuals who have reciprocated in the past.[602] This suggests that animals could have a sophisticated "book-keeping" system of goods and services. This also means they can form meaningful relationships grounded in reciprocity, as explored in Chapters 4 and 5.

My thesis does not address how animals react to failures to exchange goods and services, although reciprocity plays an important role in my discussion of animals' moral responsibility practices. It is plausible that animals' empathy could enable them to recognise anger in another individual, understand how it may be aroused by the perception of a failure to meet some reciprocity demands and lead them to adjust their behaviour to others' expectations. For instance, in a now-famous study, primatologists Frans de Waal and Sarah

[597] Buchanan (1990) and Nussbaum (2006), Chapter 1.
[598] Rawls (1999), p. 448.
[599] See Donaldson & Kymlicka (2011) and Valentini (2014).
[600] See, for instance, Trivers (1971).
[601] Clutton-Brock (2009), Ligon (1983), and MacDonald & Moehlman (1982).
[602] de Waal (1996), Hauser, Chen, Chen & Chuang (2003), and Wilkinson (1984).

Brosnan submitted some capuchin monkeys to an experiment in which the animals had to exchange tokens for two types of reward: a piece of cucumber or a grape. The latter is a reward that monkeys much more favour. After seeing a cage mate receive a grape for the same task, exchanging the token, but receiving a piece of cucumber instead of a grape, monkeys usually had a strong, angry reaction. This response could imply that they are sensitive to inequity and have expectations about how they should be treated for the same effort. Megan Van Wolkenten, Sarah Brosnan and Frans de Waal also found that effort had a major effect on monkeys' reactions and that their responses could not be explained merely by greed or goal frustration.[603] The same experiment was later carried out on dogs, ravens, crows, and various primate species and has yielded similar results.[604] In these experiments, the animals' anger was likely caused by the perception of the experiment's failure to reciprocate in the same way as for the other animals. Could have access to the wrong-making features of "deal-breaking" and, hence, possess some moral capacities relevant to contractarian moral theories and theories of justice that are partly built on reciprocity?

This possibility also raises an array of questions about the capacities that an individual must possess to be recognised as capable of having a sense of justice. Traditionally, justice theorists have argued that individuals with a sense of justice are those who can understand what a contract is and its terms and conditions. Individuals must be *rational* and reach a certain *threshold* of rationality to be said to have a sense of justice.[605] Proponents of such theories have denied that less sophisticated abilities, such as the capacity to recognise emotions and intentional action in others or the capacity to react with anger when faced with others' failures

[603] van Wolkenten, Brosnan, and de Waal (2007).
[604] Brosnan & de Waal (2014).
[605] For an example of such accounts, see Rawls (1963). For a summary of traditional accounts of justice as reciprocity, see Nussbaum (2006), p. 53.

to reciprocate, are sufficient. Challenging such assumptions would require a more thorough defence of a gradualist understanding of the sense of justice. These new arguments could be similar to the ones I developed in Chapter 1.

If animals can have access to wrongdoing in a more contractarian sense, this could bolster arguments in favour of recognising domesticated animals as citizens of our political communities, since citizenship has been traditionally defined as involving a certain degree of reciprocity and participation in the *polis*.[606] Similarly, it could give additional support to the view that wild animals should be seen as members of sovereign states and that intervention in nature should be limited to assistance that does not threaten animals' sovereignty. In *Zoopolis*, Donaldson and Kymlicka claim that, unlike human beings, wild animals live outside the circumstances of justice but argue that they should be granted sovereignty rights.[607] To do so, the authors draw a parallel with sovereign states.[608] However, in the case of human beings, the state's purpose consists of giving rise to circumstances of justice that allow its members to reciprocate and have access to some primary goods, such as basic liberties, which in turn enable them to live their lives according to their conception of the good life.[609]

Philosopher Per-Anders Svärd highlights how this places Donaldson and Kymlicka in front of a paradox: "Why do we have a duty to respect the sovereignty of societies where the circumstances of justice do not exist?"[610] In other words, why should we grant wild animals sovereignty rights if they entirely live outside the conditions of justice on which sovereignty rights rest? Recognising that wild animals can engage in reciprocity relationships could solve

606 Citizenship for domesticated animals was first put forward by Donaldson and Kymlicka in *Zoopolis* (2011). See Donaldson & Kymlicka (2011), Chapters 2 and 3.
607 *Ibid.,* p. 182.
608 *Ibid.,* Chapter 6.
609 Rawls (1971).
610 Svärd (2013), p. 198.

this paradox. If animals are aware of reciprocity expectations and if such reciprocity "deals" allow them to obtain primary goods such as food and grooming, these primary goods could, in turn, enable them to live their lives according to their preferences. The argument in favour of wild animal sovereignty might share several similarities with human cases.

In sum, recognising some animals as moral agents has implications that extend beyond moral theory and even animal ethics. My thesis is a modest theoretical contribution to these broader practical debates.

Bibliography

Aaltola, E. (2014). Affective Empathy as Core Moral Agency: Psychopathy, Autism, and Reason Revisited. *Philosophical Explorations, 17*(1), 76–92.

——— (2018). *Varieties of Empathy: Moral Psychology and Animal Ethics*. Rowman & Littlefield.

Adams, R. (1999). *Finite and Infinite Goods: A Framework for Ethics*. Oxford University Press.

Aitken, C. (2023, 24 August). Dolphin spotters shaken by Ceredigion porpoise killing. *BBC*. https://www.bbc.com/news/uk-wales-66589355.

Allen, C. (1998). Animal Concepts. *Behavioral and Brain Sciences, 21*(1), 66–66.

Anderson, J., Kuroshima, H., Takimoto, A. *et al.* (2013). Third-party social evaluation of humans by monkeys. *Nature Communications*, *4*, 1561.

Annas, J. (2006) Virtue Ethics. In D. Copp (Ed.), *The Oxford Handbook of Ethical Theory* (pp. 515–536). Oxford University Press.

Andrew, J. P. (2021). *Freedom, Mattering, and the Intrinsic Badness of Suffering*. The University of Texas at Austin.

Andrews, K. (2009). Understanding Norms Without a Theory of Mind. *Inquiry, 52*(5), 433–448.

——— (2013). Great Ape Mindreading: What's at Stake? In R. Corbey & A. Lanjouw (Eds.) 2013, *The Politics of Species: Reshaping our Relationships with Other Animals* (pp. 115–125). Cambridge University Press.

——— (2014). *The Animal Mind: An Introduction to the Philosophy of Animal Cognition*. Routledge.

——— (2020a) *How to Study Animal Minds*. Cambridge University Press.

——— (2020b). Naïve Normativity: The Social Foundation of Moral Cognition. *Journal of the American Philosophical Association, 6*(1), 36–56.

Andrews, K. & Gruen, L. (2014). Empathy in Other Apes. In H. L. Maibom (Ed.), *Empathy and Morality* (pp. 193–209). Oxford University Press.

Andrews, K., & Monsó, S. (2022). Animal Moral Psychologies. In M. Vargas & J. Doris (Eds.), *The Oxford Handbook of Moral Psychology* (pp. 388–420). Oxford University Press.

Aristotle, *Nicomachean Ethics*. Roger Crisp (Ed./Trans.). Cambridge University Press.

Arpaly, N. (2002). *Unprincipled Virtue: An Inquiry Into Moral Agency*. Oxford University Press.

Arpaly, N., & Schroeder, T. (2013). *In Praise of Desire*. Oxford University Press.

Arsenio, W., & Lover, A. (1995). Children's Conceptions of Sociomoral Affect: Happy Victimizers, Mixed Emotions, and Other Expectancies. In M. Killen & D. Hart (Eds.), *Morality in Everyday Life: Developmental Perspectives* (pp. 87–128). Cambridge University Press.

Aureli, F., & de Waal, F. B. (Eds.) (2000). *Natural Conflict Resolution*. University of California Press.

Baier, K. (1966). Moral Obligation. *American Philosophical Quarterly, 3*(3), 210–226.

Bailey, C. (2014). Le double sens de la communauté morale. *Les ateliers de l'éthique / The Ethics Forum, 9*(3), 31–67.

Bailey, N. W., & Moore, A. J. (2018). Evolutionary Consequences of Social Isolation. *Trends in Ecology & Evolution, 33*(8), 595–607.

Bain, D. (2013). What Makes Pains Unpleasant? *Philosophical Studies, 166*(1), S69–S89.

Bagnoli, C. (2024). Constructivism in Metaethics. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*. Winter 2024 Edition. https://plato.stanford.edu/entries/constructivism-metaethics/.

Balcombe, J. (2006). *Pleasurable Kingdom: Animals and the Nature of Feeling Good*. St. Martin's Griffin.

Batson, D. C. (1991). *The Altruism Question: Toward a Social Psychological Answer*. Lawerence Erlbaum.

——— (2009). These Things Called Empathy: Eight Related but Distinct Phenomena. In J. Decety & W. Ickes (Eds.), *The Social Neuroscience of Empathy* (pp. 3–15). MIT Press.

——— (2010). Empathy-Induced Altruistic Motivation. In M. Mikulincer & P. R. Shaver (Eds.), *Prosocial Motives, Emotions, and Behavior: The Better Angels of Our Nature* (pp. 15–34). American Philosophical Association.

Beauchamp, T. L. (1999). Hume on the Nonhuman Animal. *Journal of Medicine and Philosophy, 24*(4), 322–335.

Begni, V., Sanson, A., Pfeiffer, N., Brandwein, C., Inta, D., Talbot, S. R., Riva, M. A., Gass, P., & Mallien, A. S. (2020). Social Isolation in Rats: Effects on Animal Welfare and Molecular Markers for Neuroplasticity. *PloS one*, *15*(10), e0240439.

Behdadi, D. (2021). A Practice-Focused Case for Animal Moral Agency. *Journal of Applied Philosophy, 38*(3), 226–243.

——— (2024). Blame as Participant Anger: Extending Moral Claimant Competence to Young Children and Nonhuman Animals. *Philosophical Psychology*, 1–24.

Bekoff, M. (1974). Social Play in Coyotes, Wolves, and Dogs. *Bioscience, 24*(4), 225–230.

——— (1977) Mammalian Dispersal and the Ontogeny of Individual Behavioral Phenotypes. *American Naturalist*, 715–732.

——— (1981). Mammalian Sibling Interactions: Genes, Facilitative Environments, and the Coefficient of Familiarity. In D. Gubernick, P. H. Klopfer (Eds.), *Parental behavior in mammals* (pp. 307–333). Plenum.

——— (2000) Animal Emotions: Exploring Passionate Natures: Current interdisciplinary research provides compelling evidence that many animals experience such emotions as joy, fear, love, despair, and grief—we are not alone. *BioScience, 50*(10), 861–870.

——— (2004). Wild Justice and Fair Play: Cooperation, Forgiveness, and Morality in Animals. *Biology and Philosophy, 19*(4), 489–520.

——— (2008). *The Emotional Lives of Animals: A Leading Scientist Explores Animal Joy, Sorrow, and Empathy and Why They Matter*. New World Library.

Bekoff, M., & Allen, C. (2002). The Evolution of Social Play: Interdisciplinary Analyses of Cognitive Processes. In M. Bekoff, C. Allen, & G. M. Burghardt (Eds.), *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition* (pp. 429–435). The MIT Press.

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

Belger, J., & Bräuer, J. (2018). Metacognition in Dogs: Do Dogs Know They Could Be Wrong? *Learning & Behavior, 46*(4), 398–413.

Ben-Ze'ev, A. (2000). *The Subtlety of Emotions*. The MIT Press.

Benz-Schwarzburg, J., Monsó, S., & Huber, L. (2020). How Dogs Perceive Humans and How Humans Should Treat Their Pet Dogs: Linking Cognition with Ethics. *Frontiers in Psychology, 11*, 584037.

Berker, S. (2009). The Normative Insignificance of Neuroscience. *Philosophy & Public Affairs,* 37(4), 293–329.

——— (2018). The Unity of Grounding. *Mind, 127*(507), 729–777.

Bicchieri, C. (2017). *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford University Press.

Birch, J. (2021). Toolmaking and the Evolution of Normative Cognition. *Biology and Philosophy*, *36*(1), 1–26.

Bischof-Köhler, D. (1991). The Development of Empathy in Infants. In M. E. Lamb & H. Keller (Eds.), *Infant Development: Perspectives from German-speaking Countries* (pp. 245–273). Lawrence Erlbaum Associates.

Blair, R. J. R. (1995). A Cognitive Developmental Approach to Morality: Investigating the Psychopath. *Cognition*, *57*(1), 1–29.

Blair, R. J. R. (1999). Responsiveness to Distress Cues in the Child with Psychopathic Tendencies. *Personality and Individual Differences, 27*(1), 135–145.

Bloom, P. (2013). *Just Babies: The Origins of Good and Evil.* Penguin.

——— (2016). *Against Empathy: The Case for Rational Compassion*. Ecco.

Blum, L. (1994). *Moral Perception and Particularity*. Cambridge University Press.

——— (2011). Empathy and Empirical Psychology: A Critique of Shaun Nichols's Neo-Sentimentalism. In C. Bagnoli (Ed.) (2011), *Morality and the Emotions* (pp. 170–193). Oxford University Press.

Bommarito, N. (2017). Virtuous and Vicious Anger. *Journal of Ethics & Social Philosophy, 11*(3), 1–28.

Brady, M. (2018). *Suffering and Virtue*. Oxford University Press.

Brandenburg, D. (2019). Inadequate Agency and Appropriate Anger. *Ethic Theory Moral Practice, 22*, 169–185.

Brandt, R. (1979). *A Theory of the Good and the Right.* Oxford University Press.

Brennan, G., Eriksson, L., Goodin, R., & Southwood, N. (2013). *Explaining Norms.* Oxford University Press.

Brosnan, S., & de Waal, F. B. (2014). Evolution of Responses to (Un)fairness. *Science*, *346*(6207), 1251776.

Bruck, J. N. (2013). Decades-Long Social Memory in Bottlenose Dolphins. *Proceedings. Biological sciences, 280*(1768), 20131726.

Buchanan, A. (1990). Justice as Reciprocity versus Subject-Centered Justice. *Philosophy & Public Affairs, 19*(3), 227–252.

Burroughs, M. D. (2020), Navigating the Penumbra: Children and Moral Responsibility. *The Southern Journal of Philosophy, 58*(1), 77–101.

Call, J., Bräuer, J. K., & Tomasello, M. (2003). Domestic dogs (*Canis familiaris*) are sensitive to the attentional state of humans. *Journal of Comparative Psychology, 117*(3), 257–263.

Call, J., & Tomasello, M. (2008). Does the Chimpanzee Have a Theory of Mind? 30 years Later. *Trends in Cognitive Sciences, 12*(5), 187–92.

Campbell, E. Q. (1964). The Internalization of Moral Norms. *Sociometry, 27*(4), 391–412.

Camperio Ciani, A. (2000). When to Get Mad: Adaptive Significance of Rage in Animals. *Psychopathology, 33*(4), 191–197.

Carruthers, P. (2013). Evolution of Working Memory. *Proceedings of the National Academy of Sciences of the United States of America, 110*(2), 10371–10378.

Carter, D. (2020). *The Anthropologist's Dilemma: Studying Chimpanzees, Teaching Evolution, and the Intersections of Faith and Science*. Pygmaeus Press.

Cassam, Q. (2017). *The Possibility of Knowledge.* Oxford University Press.

Cavalli, C., Dzik, V., Carballo, F., & Bentosela, M. (2016). Post-Conflict Affiliative Behaviors Towards Humans in Domestic Dogs (Canis familiaris). *International Journal of Comparative Psychology, 29*, 31021.

Chazan, B. (2022). *Principles and Pedagogies in Jewish Education*. Palgrave Macmillan.

Cheney D. L. (2011). Extent and Limits of Cooperation in Animals. *Proceedings of the National Academy of Sciences of the United States of America, 108*(Suppl 2), 10902–10909.

Cherry, M. (2022). Political Anger. *Philosophy Compass, 17*(2), e12811.

Chijiwa, H., Kuroshima, H., Hori, Y. Anderson, J. R., & Fujita, K. (2015). Dogs Avoid People Who Behave Negatively to their Owner: Third-Party Affective Evaluation. *Animal Behaviour*, *106*, 123–127.

Chimpanzee Sanctuary Northwest (2023, February). Conflict and Reconciliation. *Chimpanzee Sanctuary Northwest*. https://chimpsnw.org/2023/02/conflict-and-reconciliation-2/.

Clark, S. R. L. (1984). *The Nature of The Beast: Are Animals Moral?* University of Chicago Press.

Clutton-Brock T. (2002). Breeding Together: Kin Selection and Mutualism in Cooperative Vertebrates. *Science, 296*(5565), 69–72.

——— (2009). Cooperation Between Non-Kin in Animal Societies. *Nature, 462*, 51–57.

Coates, D. J., & Tognazzini, N. A. (2012). The Contours of Blame. In D. J. Coates and N. A. Tognazzini (Eds.), *Blame: Its Nature and Norms* (pp. 3–26). Oxford University Press.

Coeckelbergh, M. (2009). Virtual Moral Agency, Virtual Moral Responsibility: on the Moral Significance of the Appearance, Perception, and Performance of Artificial Agents. *AI and Society, 24*(2), 181–189.

Cogley, T. (2014). A Study of Virtuous and Vicious Anger. In T. Kevin, & B. Craig (Eds.), *Virtues and Their Vices* (pp. 199–224), Oxford University Press.

Cook, P., Prichard, A., Spivak M., & Berns, G. (2016). Awake canine fMRI predicts dogs' preference for praise *vs* food. *Social Cognitive and Affective Neuroscience, 11*(12), 1853–1862.

Cooke, S. (2019). Betraying Animals. *Journal of Ethics, 23*, 183–200.

Coplan, Amy (2011). Understanding Empathy: Its Features and Effects. In A. Coplan & P. Goldie (Eds.), *Empathy: Philosophical and Psychological Perspectives* (pp. 2–18). Oxford University Press.

Copp, D. (1997). Defending the Principle of Alternative Possibilities: Blameworthiness and Moral Responsibility. *Noûs, 31*, 441–456.

Cornell, H., Marzluff, J., & Pecoraro, S. (2011). Social Learning Spreads Knowledge about Dangerous Humans among American crows. *Proceedings of the Royal Society, B 279*(1728), 499–508.

Cova, F., Deonna, J., & Sander, D. (2015). Introduction: Moral Emotions. *Topoi, 34*, 397–400.

Crisp, R. (1997). The *Routledge Philosophy Guidebook to Mill on Utilitarianism*. Routledge.

——— (2008). Compassion and Beyond. *Ethical Theory and Moral Practice, 11*(3), 233–246.

D'Arms, J., & Jacobson, D. (2003). The Significance of Recalcitrant Emotion (or, Anti-Quasijudgmentalism). In A. Hatzimoysis, (Ed.), *Philosophy and the Emotions* (pp. 127–146). Cambridge University Press.

D'Arms, J., & Jacobson, D. (2000). The Moralistic Fallacy: On the 'Appropriateness' of Emotions. *Philosophy and Phenomenological Research, 61*(1), 65–90.

Darwall, S. (2017). The Empathy in Moral Obligation. In N. Roughley & T. Schramme (Eds.), *Forms of Fellow Feeling: Empathy, Sympathy, Concern and Moral Agency* (pp. 265–291). Cambridge University Press.

——— (2006). *The Second-Person Standpoint: Morality, Respect, and Accountability*. Harvard University Press.

Darwin, C. (1871). *The Descent of Man*. John Murray. Volume I. First edition.

——— (1872). *The Expression of the Emotions in Man and Animals*. John Murray. First edition.

Davidov, M., Zahn-Waxler, C., Roth-Hanania, R., & Knafo, A. (2013). Concern for Others in the First Year of Life: Theory, Evidence, and Avenues for Research. *Child Development Perspectives, 7*(2), 126–131.

Davidson, T. J., Kloosterman, F., & Wilson, M. A. (2009). Hippocampal Replay of Extended Experience. *Neuron*, *63*(4), 497–507.

De Sousa, R. (2001). Moral Emotions. *Ethical Theory and Moral Practice, 4*(2), 109–126.

De Vignemont, F., & Singer, T. (2006). The Empathic Brain: How, When, and Why? *Trends in Cognitive Sciences, 10*(10), 435–441.

de Waal, F. B. (1996). *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*. Harvard University Press.

——— (2006a). Morally Evolved: Primate Social Instincts, Human Morality, and the Rise and Fall of the 'Veneer Theory'. In S. Macedo & J. Ober (Eds.), *Primates and Philosophers: How Morality Evolved* (pp. 1–58). Princeton University Press.

——— (2006b). The Tower of Morality. In S. Macedo & J. Ober (Eds.), *Primates and Philosophers: How Morality Evolved* (pp. 161–182). Princeton University Press.

——— (2008). Putting the Altruism Back into Altruism: The Evolution of Empathy. *Annual Review of Psychology, 59*, 279–300.

——— (2009). *The Age of Empathy: Nature's Lessons for a Kinder Society*. Harmony Books.

——— (2014). Natural Normativity: The 'Is' and 'Ought' of Animal Behavior. *Behavior, 151*, 185–204.

——— (2023, March). *Animal Emotions and Animal Rights Law: First Animal Rights Lecture 2023* [Conference presentation]. Cambridge Centre for Animal Rights Law, Cambridge, United Kingdom. https://www.youtube.com/watch?v=6lIr775m4-c.

de Waal, F. B., & Preston, S. (2017). Mammalian Empathy: Behavioural Manifestations and Neural Basis. *Nature Reviews Neuroscience* 18: 498–509.

Debes, R. (2009). Neither Here nor There: The Cognitive Nature of Emotion. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 146*(1), 1–27.

Decety, J., & Cowell, J. M. (2014). The Complex Relation Between Morality and Empathy. *Trends in Cognitive Sciences, 10*(7), 337–339.

DeGrazia, D. (1996). *Taking Animals Seriously: Mental Life and Moral Status*. Cambridge University Press.

——— (2009). Self-Awareness in Animals. In R.W. Lurz (Ed.), *The Philosophy of Animal Minds* (pp. 201–217). Cambridge University Press.

——— (2014). What Is Suffering and What Sorts of Beings Can Suffer? In R. M. Green & N. J. Palpant (Eds.), *Suffering and Bioethics* (pp. 134–153). Oxford University Press.

Delon, Nicolas (2022). Strangers to Ourselves: A Nietzschean Challenge to the Badness of Suffering. *Inquiry*, 1–30 (ahead of print).

Deonna, J. (2007). The Structure of Empathy. *Journal of Moral Philosophy, 4*(1), 99–116.

Deonna, J., & Teroni, F. (2014). In What Sense Are Emotions Evaluations? In S. Roeser & C. Todd (Eds.), *Emotion and Value* (pp. 15–31). Oxford University Press.

Dixon, B. (2008a). *Animals, Emotion and Morality: Marking the Boundary*. Prometheus Books.

——— (2008b). The Moral Responsibility of Children and Animals. *Thinking: The Journal of Philosophy for Children, 19*(1), 20–30.

Donaldson, S., & Kymlicka, W. (2011). *Zoopolis: A Political Theory of Animal Rights*. Oxford University Press.

——— (2012, July). *Do We Need a Political Theory of Animal Rights?* [Conference presentation], *Minding Animals International Conference*, Utrecht, Netherlands.

Dondi, M., Simion, F., & Caltran, G. (1999). Can Newborns Discriminate Between their Own Cry and the Cry of Another Newborn Infant? *Developmental Psychology*, *35*, 418–426.

Dwyer, S. (2003). Moral Development and Moral Responsibility. *The Monist, 86*(2), 181–199.

Echeverri, S. (2019). Emotional Justification. *Philosophy and Phenomenological Research, 98*(3), 541–566.

Edwards, S. (2003). Three Concepts of Suffering. *Medicine, Health Care, and Philosophy, 6*, 59–66.

Eisenberg, N., & Miller, P. A. (1987). The Relation of Empathy to Prosocial and Related Behaviors. *Psychological Bulletin, 101*(1), 91–119.

Eisenberg, N., Fabes, R. A., & Spinrad, T. L. (2006). Prosocial Development. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of Child Psychology: Social, Emotional, and Personality Development* (pp. 646–718). John Wiley & Sons. Sixth edition.

Ekman, P. (1992). An Argument for Basic Emotions. *Cognition and Emotion, 6*(3/4), 169–200.

Eliot, G. (1859). *Adam Bede*. Icon Classics. Edition of 2005.

Escolano-Prérez, E., Herrero-Nivela, M. L., & Anguera, M. T. (2019). Preschool Metacognitive Skill Assessment in Order to Promote Educational Sensitive Response from Mixed-methods Approach: Complementarity of Data Analysis. *Frontiers in Psychology, 10*, 1–22.

FeldmanHall, O., Son, J., & Heffner, J. (2018). Norms and the Flexibility of Moral Action. *Personality Neuroscience, 1*(e15).

Ferrigno, S., Kornell, N., & Cantlon, J. F. (2017). A Metacognitive Illusion in Monkeys. *Proceedings of the Royal Society B: Biological Sciences, 284*(1862), 20171541.

Ferrin, A. (2019). Nonhuman Animals Are Morally Responsible. *American Philosophical Quarterly*, *56*(2), 135–154.

Fischer, B. (2022, 14 November). Theories of Welfare and Welfare Range Estimates. *Rethink Priorities*. https://rethinkpriorities.org/publications/theories-of-welfare-and-welfare-range-estimates.

Fischer, J. M., & Ravizza, M. (1993). Introduction. In J. M. Fischer & M. Ravizza (Eds.). *Perspectives on Moral Responsibility* (pp. 1–42). Cornell University Press.

——— (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge University Press.

Fitzpatrick, S. (2017). Animal Morality: What Is the Debate About? *Biology and Philosophy, 32*(6), 1151–1183.

——— (2020). Chimpanzee Normativity: Evidence and Objections. *Biology & Philosophy, 35*, 45.

Floridi L., & Sanders, J. W. (2004). On the Morality of Artificial Agents. *Minds and Machines, 14*, 349–379.

Foot, P. (1978). *Virtues and Vices and Other Essays in Moral Philosophy*. Basil Blackwell.

Foster, D. J., & Wilson, M. A. (2006). Reverse Replay of Behavioural Sequences in Hippocampal Place Cells During the Awake State. *Nature, 440*(7084), 680–683.

Francione, G. (2000). *Introduction to Animal Rights: Your Child or the Dog?* Temple University Press.

Fraser, O., & Bugnyar, T. (2011) Ravens reconcile after aggressive conflicts with valuable partners. *PloS One, 6*(3), e18118.

Friend, S. (2016). Fiction and Emotion. In A. Kind (Ed.), *The Routledge Handbook of Philosophy of Imagination* (pp. 217–229). Routledge.

Frijda, N. H. (1986). *The Emotions*. Cambridge University Press.

Fu, G., Evans, A. D., Xu, F., & Lee, K. (2012). Young children can tell strategic lies after committing a transgression. *Journal of Experimental Child Psychology, 113*, 147–158.

Gallagher, S. (2012). Empathy, Simulation, and Narrative. *Science in Context, 25*(3), 355–381.

Garthoff, J. (2019). Animal Punishment and the Conditions of Responsibility. *Philosophical Papers, 49*(1), 69–105.

Gaut, Berrys (1999). Identification and Emotion in Narrative Film. In C. Plantinga & G. Smith (Eds.), *Passionate Views: Film, Cognition and Emotion* (pp. 200–216). Johns Hopkins University Press.

Gauthier, D. (1986). *Morals By Agreement*. Oxford University Press.

Geangu, E., Benga, O., Stahl, D., & Striano, T. (2010). Contagious Crying Beyond the First days of Life. *Infant Behavior & Development, 33*, 279–288.

Gibbard, A. (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Harvard University Press.

Goldie, P. (2004). Emotion, Feeling, and Knowledge of the World. In R. Solomon (Ed.), *Thinking About Feeling: Contemporary Philosophers on Emotions* (pp. 91–106). Oxford University Press.

——— (2007). Emotion. *Philosophy Compass, 2*(6), 928–938.

Goldman, A. (1992). Empathy, Mind, and Morals. *Proceedings and Addresses of the American Philosophical Association, 66*(3), 17–41.

Goodall, J. (1986). *The Chimpanzees of Gombe: Patterns of Behavior*. Harvard University Press.

Goupil, L., & Kouider, S. (2016). Behavioral and Neural Indices of Metacognitive Sensitivity in Preverbal Infants. *Current Biology, 26*(22), 3038–3045.

Green, M. (2007*). Self-Expression*. Oxford University Press.

Greene, J. (2008). The Secret Joke of Kant's Soul. In W. Sinnott-Armstrong (Ed.), *Moral Psychology* (pp. 35–79). MIT Press. Volume 3.

——— (2009). Dual-Process Morality and the Personal/Impersonal Distinction: A Reply to McGuire, Langdon, Coltheart, and Mackenzie. *Journal of Experimental Social Psychology, 45*, 581–584.

Greene, J., & Haidt, J. (2002). How (and Where) Does Moral Judgment Work? *Trends in Cognitive Sciences, 6*(12), 517–523.

Gruen, L. (2019). Empathy in Mind. In K. Andrews & J. Beck (Eds.), *The Routledge Handbook of Philosophy of Animal Minds* (pp. 485–490). Routledge.

Gupta, A. S., van der Meer, M. A., Touretzky, D. S., & Redish, A. D. (2010). Hippocampal replay is not a simple function of experience. *Neuron, 65*(5), 695–705.

Haksar, V. (1998). Moral Agents. In *The Routledge Encyclopedia of Philosophy*. Taylor and Francis. https://www.rep.routledge.com/articles/thematic/moral-agents/v-1.

Halstead, J. M. (2010). Moral Education. In C. S. Clauss-Ehlers (Ed.), *Encyclopedia of Cross-Cultural School Psychology* (pp. 630–631). Springer.

Halstead, J. M., & McLaughlin, T. H. (Eds.) (1999). *Education in Morality*. Routledge.

Hamlin, J., Wynn, K., & Bloom, P. (2007). Social Evaluation by Preverbal Infants. *Nature, 450*, 557–559.

Hamlin, J. K., Wynn, K., & Bloom, P. (2010). Three-month-olds show a negativity bias in their social evaluations. *Developmental Science, 13*(6), 923–929.

Hamlin, J. K., Wynn, K., & Bloom, P. (2011). How Infants and Toddlers React to Antisocial Others. *Proceedings of the National Academy of Sciences of the United States of America, 108*(50), 19931–19936.

Hampton, J., & Murphy, J. (1988). *Forgiveness and Mercy*. Cambridge University Press.

Harris, P., & Saarni, C. (1989). Children's Understanding of Emotion: An Introduction. In P. Harris & C. Saarni (Eds.), *Children's Understanding of Emotion* (pp. 3–24). Cambridge University Press.

Hart, H. L. A. (2008). *Punishment and Responsibility: Essays in the Philosophy of Law*. Oxford University Press.

Hauser, M. D. (2000). *Wild Minds: What Animals Really Think*. Henry Holt and Company.

Hauser, M., Chen, K., Chen, F., & Chuang, E. (2003). Give Unto Others: genetically unrelated cotton-top tamarin monkeys preferentially give food to those who altruistically give food back. *Proceedings of the Royal Society of London Series B*, 2363–2370.

Hay, D., Nash, A., & Pedersen, J. (1981). Responses of Six-Month-Olds to the Distress of their Peers. *Child Development, 52*, 1071–1075.

Hearne, V. (1986). *Adams' Task: Calling Animals by Name*. Skyhorse Publishing. Edition of 2007.

Helm, B. (2021). Friendship. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*. Fall 2023 Edition. https://plato.stanford.edu/entries/friendship/.

Hieronymi, P. (2004). The Force and Fairness of Blame. *Philosophical Perspectives, 18*, 115–148.

——— (2007). Rational Capacity as a Condition on Blame. *Philosophical Books, 48*(2), 109–123.

Hoffman, M. L. (2000). *Empathy and Moral Development: Implications for Caring and Justice*. Cambridge University Press.

Horowitz, A. (2010). *Inside of a Dog: What Dogs See, Smell, and Know*. Scribner Book Company.

Huber, A., Barber, A. L. A., Faragó, T., Müller, C. A., & Huber, L. (2017). Investigating Emotional Contagion in Dogs (*Canis Familiaris*) to Emotional Sounds of Humans and Conspecifics. *Animal cognition*, *20*(4), 703–715.

Hume, D. (1739). *A Treatise of Human Nature*. Clarendon Press. Edition of 1978.

——— (1748). *An Enquiry Concerning the Principles of Morals*. Oxford University Press, Edition of 1998.

Hurley, E. A., & Macnamara, C. (2010). Beyond Belief: Toward a Theory of the Reactive Attitudes. *Philosophical Papers, 39*(3), 373–399.

Ichikawa, J. J., & Steup, M. (2024) The Analysis of Knowledge. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*. Fall 2024 Edition. https://plato.stanford.edu/entries/knowledge-analysis/.

Jaquet, François (2022, May 2024). Non, les animaux n'ont pas de devoirs moraux. *La Conversation*. https://theconversation.com/non-les-animaux-nont-pas-de-devoirs-moraux-177782.

Jensen, K., Call, J., & Tomasello, M. (2007). Chimpanzees are vengeful but not spiteful. *Proceedings of the National Academy of Sciences of the United States of America, 104*(32), 13046–13050.

Jo, H., McCune, K. B., Jablonski, P. G., & Lee, S. I. (2023). Long-term memory of experienced jays facilitates problem-solving by naïve group members in the wild. *Scientific Reports*, *13*(1), 21593.

Jordan, M. R., Amir, D., & Bloom, P. (2016). Are Empathy and Concern Psychologically Distinct? *Emotion, 16*(8), 1107–1116.

Joyce, R. (2005). *The Evolution of Morality*. MIT Press.

Kahane, G. (2009). Pain, Dislike, and Experience. *Utilitas, 21*(3), 327–336.

——— (2016). Pain, Experience, and Well-Being. In G. Fletcher (Ed.), *The Routledge Handbook of Philosophy of Well-Being* (pp. 209–2020). Routledge.

Kaminski, J., Call, J., & Fischer, J. (2004). Word Learning in a Domestic Dog: Evidence for "Fast Mapping". *Science, 304*(5677), 1682–1683.

Kappeler, P. M., & van Schaik, C. P. (2006). *Cooperation in Primates and Humans: Mechanisms and Evolution*. Springer.

Karlsson, M. P., & Frank, L. M. (2009). Awake Replay of Remote Experiences in the Hippocampus. *Nature Neuroscience, 12*(7), 913–918.

Kauppinen, A. (2018). Valuing Anger. In M. Cherry & O. Flanagan (Eds.), *The Moral Psychology of Anger* (pp. 31–48). Rowman & Littlefield.

——— (2020). The World According to Suffering. In D. Bain, M. Brady, & J. Corns (Eds.), *Philosophy of Suffering: Metaphysics, Value, and Normativity* (pp. 19–36). Routledge.

——— (2022). *Moral Sentimentalism*. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*. Spring 2022 Edition. https://plato.stanford.edu/entries/moral-sentimentalism/.

Keeling, L. J., Jonare, L., & Lanneborn, L. (2009). Investigating Horse–Human Interactions: The Effect of a Nervous Human. *Veterinary Journal, 181*(1), 70–71.

Kelly, D. (2020). Internalized Norms and Intrinsic Motivations: Are Normative Motivations Psychologically Primitive? *Emotion Researcher, 1*, 36–45.

——— (2022). Two Ways to Adopt a Norm: The (Moral?) Psychology of Internalization and Avowal. In M. Vargas & J. Doris (Eds.), *The Oxford Handbook of Moral Psychology* (pp. 285–309). Oxford University Press.

Kelly, D., Stich, S., Haley, K. J., Eng, S. J., & Fessler, D. M. T. (2007). Harm, Affect and the Moral/Conventional Distinction. *Mind & Language, 22*(2), 117–131.

Kious, B. (2022). Three Kinds of Suffering and Their Relative Moral Significance. *Bioethics, 36*, 621–627.

Kitcher, P. (2006). Ethics and Evolution. How to Get Here from There. In S. Macedo & J. Ober (Eds.), *Primates and Philosophers: How Morality Evolved* (pp. 120–139). Princeton University Press.

Kochanska G., & Aksan, N. (2006). Children's Conscience and Self-Regulation. *Journal of Personality, 74*(6), 1587–1618.

Kögler, H. H., & Stueber, K. (2000). Introduction. In H. H. Kögler & K. Stueber (Eds.), *Empathy and Agency: The Problem of Understanding in the Human Sciences* (pp. 1–61). Westview Press.

Korsgaard, C. (1992). Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations. *Philosophical Perspectives, 6*, 305–332.

——— (2006). Morality and the Distinctiveness of Human Action. In S. Macedo & J. Ober (Eds.), *Primates and Philosophers: How Morality Evolved* (pp. 98–119). Princeton University Press.

——— (2009). *Self-Constitution: Agency, Identity, and Integrity*. Oxford University Press.

——— (2008). *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford University Press.

——— (2010). Reflections on the Evolution of Morality. *The Amherst Lecture in Philosophy* 5, 1–29.

Koski, S. E. (2015). Reconciliation and Peace-Making: Insights from Studies on Nonhuman Animals. In R. Scott & S. Kosslyn (Eds.), *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*. John Wiley & Sons Ltd.

Krasheninnikova, A., Brucks, D., Blanc, S., & von Bayern, A. M. P. (2019). Assessing African Grey Parrots' Prosocial Tendencies in a Token Choice Paradigm. *Royal Society Open Science, 6*(12), 190696.

Krueger, J. (2012). Seeing Mind in Action. *Phenomenology and the Cognitive Sciences, 11*, 149–173.

Krupina, N. A., Shirenova, S. D., & Khlebnikova, N. N. (2020). Prolonged social isolation, started early in life, impairs cognitive abilities in rats depending on sex. *Brain sciences, 10*(11), 799.

Lage, C. A., Wolmarans, D. W., & Mograbi, D. C. (2022). An Evolutionary View of Self-Awareness. *Behavioural Processes, 194*, 1–11.

Lambert, H., Carder, G., & D'Cruze, N. (2019). Given the Cold Shoulder: A Review of the Scientific Literature for Evidence of Reptile Sentience. *Animals: An Open Access Journal from MDPI*, *9*(10), 821.

Langergraber, K. E., Mitani, J. C., & Vigilant, L. (2007). The Limited Impact of Kinship on Cooperation in Wild Chimpanzees. *Proceedings of the National Academy of Sciences of the United States of America, 104*(19), 7786–7790.

Lazarus, R. (1991). *Emotion and Adaptation*. Oxford University Press.

Lee, A. (forthcoming). Metaethical Experientialism. In G. Lee & A. Pautz (Eds.), *The Importance of Being Conscious*. Oxford University Press.

Lei, Y. (2023). Sociality and Self-Awareness in Animals. *Frontiers in Psychology, 13*, 1065638.

Lenkei, R., Faragó, T., Zsilák, B. *et al.* (2021). Dogs (*Canis familiaris*) recognize their own body as a physical obstacle. *Scientific Reports, 11*, 2761.

Lewis, N. J. (1999), Frustration of Goal-Directed Behaviour in Swine. *Applied Animal Behaviour Science, 64*(1), 19–29.

Liévin-Bazin A., Pineaux, M., Clerc, O., Gahr, M., von Bayern, A. M. P., Bovet, D. (2018). Emotional responses to conspecific distress calls are modulated by affiliation in cockatiels (*Nymphicus hollandicus*). *PLoS ONE, 13*(10), e0205314.

Ligon, J. D. (1983). Cooperation and Reciprocity in Avian Social Systems. *The American Naturalist, 121*(3), 366–384.

Lovett, A., & Riedener, S. (2024). Commonsense Morality and Contact with Value. *Philosophy and Phenomenological Research, 109*, 410–430.

Lucca, K., Yuen, F., Wang, Y. *et al.* (2025). Infants' Social Evaluation of Helpers and Hinderers: A Large-Scale, Multi-Lab, Coordinated Replication Study. *Developmental Science, 28,* e13581.

Lyons, W. (1985). *Emotion*. Cambridge University Press.

Macdonald, D., & Moehlman, P. (1982). Cooperation, Altruism, and Restraint in the Reproduction of Carnivores. In Bateson, P.P.G. & Klopfer, P.H. (Eds.) (1982), *Ontogeny, Perspectives in Ethology* (pp. 433–467). Springer. Volume 5.

Macnamara, C. (2011). Holding Others Responsible. *Philosophical Studies, 152*(1), 81–102.

——— (2012). Taking Demands Out of Blame. In D. J. Coates & N. A. Tognazzini. (Eds.) (2012). *Blame: Its Nature and Norms* (pp. 141–161). Oxford University Press.

Magalhães, D. M., Mampay, M., Sebastião, A. M., Sheridan, G. K., & Valente, C. A. (2024). Age-Related Impact of Social Isolation in Mice: Young *vs* Middle-Aged. *Neurochemistry international*, *174*, 105678.

Magyari, L., Huszár, Z., Turzó, A., & Andics, A. (2020). Event-related potentials reveal limited readiness to access phonetic details during word processing in dogs. *Royal Society Open Science, 7*(12), 200851.

Mallon, R., & Nichols, S. (2010). Rules. In J. M. Doris, and the Moral Psychology Research Group. (Eds.), *The Moral Psychology Handbook* (pp. 297–320). Oxford University Press.

Malti, T., & Ongley, S. (2014). The Development of Moral Emotions and Moral Reasoning. In M. Killen & J. Smetana (Eds.), *The Handbook of Moral Development* (pp. 581–609). Psychology Press.

Marino, L. (2017). Thinking Chickens: A Review of Cognition, Emotion, and Behavior in the domestic chicken. *Animal Cognition, 20*, 127–147.

Marino, L., & Allen, K. (2017). The Psychology of Cows. *Animal Behavior and Cognition, 4*(4), 474–498.

Marino, L., & Colvin, C. (2015). Thinking Pigs: A Comparative Review of Cognition, Emotion, and Personality in *Sus Domesticus*. *International Journal of Comparative Psychology, 28*, 23859.

Marsh, A. A. (2014). Empathy and Moral Deficits in Psychopathy. In Maibom, H. L. (Ed.), *Empathy and Morality* (pp. 138–154). Oxford University Press.

Masserman, J. H., Wechkin, S., & Terris, W. (1964a). "Altruistic" Behavior in Rhesus Monkeys. *The American Journal of Psychiatry, 121*(6), 584–585.

Wechkin, S., Masserman, J. H., & Terris, W., Jr. (1964b). Shock to a Conspecific as an Aversive Stimulus. *Psychonomic Science, 1*(2), 47–48.

McCormick, K. (2022). *The Problem of Blame: Making Sense of Moral Anger*. Cambridge University Press.

McGeer, V. (2012). Co-Reactive Attitudes and the Making of Moral Community. In R. Langdon, & C. Mackenzie (Eds.), *Emotions, Imagination, and Moral Reasoning* (pp. 299–325). Psychology Press.

McLeod, C. (2023). Trust. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*. Fall 2023 Edition. https://plato.stanford.edu/entries/trust/.

McKenna, M. (2012). *Conversation and Responsibility*. Oxford University Press.

McMahan, J. (2002). *The Ethics of Killing: Problems at the Margins of Life*. Oxford University Press.

——— (2008). Eating Animals the Nice Way. *Daedalus, 137*(1), 66–76.

McNeil, W. (2012). On Seeing That Someone Is Angry. *European Journal of Philosophy, 20*(4), 575–597.

Mealey, L. (1995). The Sociobiology of Sociopathy: An Integrated Evolutionary Model. *Behavioral and Brain Sciences*, *18*(3), 523–599.

Mele, A. R. (2001). *Autonomous Agents: From Self-Control to Autonomy*. Oxford University Press.

Michalson, L., & Lewis, M. (1985). What Do Children Know about Emotions and When Do They Know It? In M. Lewis & C. Saarni (1985), *The Socialization of Emotions* (pp. 117–139). Plenum Press.

Miller, D. (2006). Mill's Theory of Sanctions. In H. R. West (Ed.), *The Blackwell Guide to Mill's Utilitarianism* (pp. 159–173). Blackwell.

Mineka, F. E., & Lindley, D. N. (Eds.). (1972). *Collected Works of John Stuart Mill.* University of Toronto Press, Routledge and Kegan Paul. Volume XV – The Later Letters 1849–1873 Part II.

Mischel, W., Shoda Y., & Rodriguez M. L. (1989). Delay of Gratification in Children. *Science, 244*, 933–938.

Misselhorn, C. (2022). Artificial Moral Agents: Conceptual Issues and Ethical Controversy. In S. Voeneky, P. Kellmeyer, O. Mueller, & W. Burgard (Eds.), *The Cambridge Handbook of Responsible Artificial Intelligence: Interdisciplinary Perspectives* (pp. 31–49). Cambridge University Press.

Mitchell, R. W. (Ed.). (2002). *Pretending and Imagination in Animals and Children*. Cambridge University Press.

Monsó, S. (2015). Empathy and Morality in Behaviour Readers. *Biology and Philosophy, 30*(5), 671–690.

——— (2017). Morality Without Mindreading. *Mind & Language, 32*(3), 338–357.

——— (2021). Is Predation Necessarily Amoral? In A. Siegetsleitner, A. Oberprantacher, M. Frick & U. Metschl (Ed.), *Crisis and Critique: Philosophical Analysis and Current Events: Proceedings of the 42nd International Ludwig Wittgenstein Symposium* (pp. 367–382). De Gruyter.

Monsó, S., Benz-Schwarzburg, J., & Bremhorst, A. (2018). Animal Morality: What It Means and Why It Matters. *The Journal of Ethics, 22*(3), 283–310.

Monsó, S., & Wrage, B. (2020). Tactful Animals: How the Study of Touch Can Inform the Animal Morality Debate. *Philosophical Psychology, 34*(1), 1–27.

Montreuil, M., Noronha, C., Floriani, N., & Carnevale, F. A. (2018). Children's Moral Agency: An Interdisciplinary Scoping Review. *Journal of Childhood Studies, 43*(2), 17–30.

Moore, G. E. (1903). *Principia Ethica*. Cambridge University Press.

Moore, M. S. (2009). *Causation and Responsibility: An Essay in Law, Morals, and Metaphysics*. Oxford University Press.

Mulligan, K. (2009a). Moral Emotions. In D. Sander & K. Scheter (Eds.), *Oxford Companion to Emotion and the Affective Sciences*. Oxford University Press.

——— (2009b). Emotions and Values. In P. Goldie (Ed.), *The Oxford Handbook of Philosophy of Emotion*. (pp. 475–500). Oxford University Press.

Mulligan, K., & Scherer, K. (2012). Toward a Working Definition of Emotion. *Emotion Review, 4*(4), 345–357.

Mumtaz, F., Khan, M. I., Zubair, M., & Dehpour, A. R. (2018). Neurobiology and Consequences of Social Isolation Stress in Animal Model: A Comprehensive Review. *Biomedicine & Pharmacotherapy, 105*, 1205–1222.

Murdoch, I. (1987). *The Book and the Brotherhood*, Vintage Classics. Edition of 2003.

Murphy, J. G. (2012). *Punishment and the Moral Emotions: Essays in Law, Morality, and Religion*. Oxford University Press.

Musschenga, A. (2015). Moral Animals and Moral Responsibility. *Les ateliers de l'éthique / The Ethics Forum, 10*(2), 38–59.

Musschenga, B. (2013). Animal Morality and Human Morality. In B. Musschenga & A. van Harskamp (Eds.), *What Makes Us Moral? On the Capacities and Conditions for Being Moral* (pp. 99–116). Springer Science + Business Media.

Naar, Hichem (2019a). Emotion: Animal and Reflective. *The Southern Journal of Philosophy, 57*(4), 561–588.

——— (2019b). Are Emotions Events, Processes, States or Dispositions? In A. Scarantino (Ed.), *The Routledge Handbook of Emotion Theory* (pp. 469–487). Routledge.

——— (2021). The Fittingness of Emotions. *Synthese*, *199*(5–6), 13601–13619.

Nagel, T. (1986). *The View from Nowhere*. Oxford University Press.

Nailer, T. (2022). *Moral Agency* [Master's thesis, University of Adelaide, School of Humanities].

Nawroth, C. (2017). Invited Review: Socio-Cognitive Capacities of Goats and their Impact on Human-Animal Interactions. *Small Ruminant Research, 150*, 7–75.

Neblett, W. (1979). Indignation: A Case Study in the Role of Feelings in Morals. *Metaphilosophy*, *10*(2), 139–152.

Neill, A. (1996). Empathy and (Film) Fiction. In D. Bordwell & N. Carroll (Eds.), *Post-Theory: Reconstructing Film Studies* (pp. 175–194). University of Wisconsin Press.

Schwing, R., Nelson, X. J., Wein, A., & Parsons, S. (2017). Positive Emotional Contagion in a New Zealand Parrot. *Current Biology, 27*(6), R213–R214.

Nichols, S. (2001). Mindreading and the Cognitive Architecture Underlying Altruistic Motivation. *Mind & Language, 16*(4), 425–455.

Nucci, L. P. (2001). *Education in the Moral Domain.* Cambridge University Press.

Nucci, L. P., & Nucci, M. (1982). Children's Social Interactions in the Context of Moral and Conventional Transgressions. *Child Development*, *53*(2), 403–412.

Nucci, L. P., & Turiel, E. (1978). Social Interactions and the Development of Social Concepts in Preschool Children. *Child Development, 49*(2), 400–407.

Nussbaum, M. (2001). *Upheavals of Thought: The Intelligence of Emotions*, Cambridge University Press.

——— (2006). *Frontiers of Justice: Disability, Nationality, Species Membership.* Harvard University Press.

——— (2013). *Political Emotions: Why Love Matters for Justice.* Harvard University Press.

——— (2016). *Anger and Forgiveness: Resentment, Generosity, Justice*. Oxford University Press.

O' Connell, S. (1995). Empathy in Chimpanzees: Evidence for a Theory of Mind? *Primates, 36*, 397–410.

O'Neil, E. (2017). Kinds of Norms. *Philosophy Compass, 12*(5), e12416.

Ortony, A., Clore, G., & Collins, A. (1988). *The Cognitive Structure of Emotions*. Cambridge University Press.

Palagi, E., Cordoni, G., Demuru, E., & Bekoff, M. (2016). Fair Play and Its Connection with Social Tolerance, Reciprocity and the Ethology of Peace. *Behaviour, 153*(9/11), 1195–1216.

Palagi, E., Nicotra, V., & Cordoni, G. (2015). Rapid Mimicry and Emotional Contagion in Domestic Dogs. *Royal Society Open Science, 2*(12), 150505.

Panksepp, J., & Panksepp, J. (2013). Toward a Cross-Species Understanding of Empathy. *Trends in Neurosciences, 36*(8), 489–496.

Pannewitz, L., & Loftus, L. (2023). Frustration in Horses: Investigating Expert Opinion on Behavioural Indicators and Causes Using a Delphi Consultation. *Applied Animal Behaviour Science, 258*, 105818.

Patrick, H. (2024). Adorable foxes frolic in snow outside University of Oxford's historic Magdalen College. *The Independent*. https://www.independent.co.uk/tv/lifestyle/weather-snow-oxford-foxes-magdalen-b2649816.html

Peña-Guzmán, D. (2022). *When Animals Dream: The Hidden World of Animal Consciousness*. Princeton University Press.

Penn, D., & Povinelli, D. (2007). Causal Cognition in Human and Nonhuman Animals: A Comparative, Critical Review. *Annual Review of Psychology, 58*(1), 97–118.

Pepperberg, I. M. (1999). *The Alex Studies: Cognitive and Communicative Abilities of Grey Parrots*. Harvard University Press.

Pérez-Manrique, A., & Gomila, A. (2022). Emotional Contagion in Nonhuman Animals: A Review. *Wiley Interdisciplinary Reviews. Cognitive Science, 13*(1), e1560.

Pettigrove, G., & Tanaka, K. (2014). Anger and Moral Judgment. *Australasian Journal of Philosophy*, *92*(2), 269–286.

Pilley, J. W., & Reid, A. K. (2011). Border collie comprehends object names as verbal referents. *Behavioural Processes, 86*(2), 184–195.

Pluhar, E. B. (1995). *Beyond Prejudice: The Moral Significance of Human and Nonhuman Animals*. Duke University Press.

Pongrácz, P., Miklósi, Á., & Csányi, V. (2001). Owner's Beliefs on the Ability of their Pet Dogs to Understand Human Verbal Communication: A Case of Social Understanding. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 20*(1–2), 87–107.

Povinelli, D. (1996). Chimpanzee Theory of Mind? The Long Road to Strong Inference. In P. Carruthers & P. Smith (Eds.). *Theories of Theories of Mind* (pp. 293–329). Cambridge University Press.

Preston, S. D., & de Waal, F. B. (2002). Empathy: Its Ultimate and Proximate Bases. *Behavioral and Brain Sciences, 25*(1), 1–72.

Prichard, A., Cook, P. F., Spivak, M., Chhibber, R., & Berns, G. S. (2018). Awake fMRI Reveals Brain Regions for Novel Word Detection in Dogs. *Frontiers in Neuroscience, 12*, 737.

Prinz, J. (2009). The Moral Emotions. In P. Goldie (Ed.), *The Oxford Handbook of Philosophy of Emotion* (pp. 519–538). Oxford University Press.

——— (2011). Is Empathy Necessary for Morality? In A. Coplan & P. Goldie (Eds.), *Empathy: Philosophical and Psychological Perspectives* (pp. 211–229). Oxford University Press.

Proctor, D., Williamson R. A., de Waal, F. B., & Brosnan, S. (2013). Chimpanzees Play the Ultimatum Game. *PNAS, 110*(6), 2070–2075.

Quaranta, A., d'Ingeo, S., Amoruso, R., & Siniscalchi, M. (2020). Emotion Recognition in Cats. *Animals, 10*, 1107.

Railton, P. (2006). Normative Guidance. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics* (pp. 3–33). Oxford University Press. Volume 1.

Range, F., Horn, L., Viranyi, Z., & Huber, L. (2009). The Absence of Reward Induces Inequity Aversion in Dogs. *PNAS, 106*(1), 340–345.

Ratcliffe, V. F., & Reby, D. (2014). Orienting Asymmetries in Dogs' Responses to Different Communicatory Components of Human Speech. *Current Biology, 24*(24), 2908–2912.

Ratcliffe, V. F., Taylor, A. M., & Reby, D. (2015). Cross-Modal Correspondences in Non-Human Mammal Communication. *Multisensory Research, 29*(1–3), 49–91.

Rawls, J. (1963). The Sense of Justice. *The Philosophical Review, 72*(3), 281–305.

——— (1971). *A Theory of Justice*. Harvard University Press.

——— (1999). *A Theory of Justice: Revised Edition*. Harvard University Press.

Reeve, C., & Jacques, S. (2022). Responses to Spoken Words by Domestic Dogs: A New Instrument for Use with Dog Owners. *Applied Animal Behaviour Science, 246*, 105513.

Regan, T. (1983). *The Case for Animal Rights*. University of California Press.

Rhodes, M., & Chalik, L. (2013). Social Categories as Markers of Intrinsic Interpersonal Obligations. *Psychological Science, 24*(6), 999–1006.

Rice, C. (2015). Well-Being and Animals. In G. Fletcher (Ed.), *The Routledge Handbook of Philosophy of Well-Being* (pp. 378–388). Routledge.

Ripstein, A. (1999). Equality, Responsibility, and the Law. *Law and Philosophy, 20*(6), 617–635.

Roberts, R. C. (1996). Propositions and Animal Emotion. *Philosophy, 71*, 147–156.

——— (2009). Emotions and the Canons of Evaluation. In P. Goldie (Ed.), *The Oxford Handbook of Philosophy of Emotion* (pp. 561–584). Oxford University Press.

——— (2013). *Emotions in the Moral Life*. Cambridge University Press.

Robson, J. M. (Ed.). (1978). *Collected Works of John Stuart Mill*. University of Toronto Press, Routledge and Kegan Paul. Vol. XI – Essays on Philosophy and the Classics.

Robson, J. M. (Ed.). (1985). *Collected Works of John Stuart Mill.* University of Toronto Press, Routledge and Kegan Paul. Volume X – Essays on Ethics, Religion, and Society.

Rosati, A. G., Stevens, J. R., Hare, B., & Hauser, M. D. (2007). The Evolutionary Origins of Human Patience: Temporal Preferences in Chimpanzees, Bonobos, and Human Adults. *Current Biology* 17(19): 1663–1668.

Rosen, G. (2015). The Alethic Conception of Moral Responsibility. In R. Clarke, M. McKenna, & A. Smith. (Eds.), *The Nature of Moral Responsibility: New Essays* (pp. 65–88). Oxford University Press.

Ross, A. (1983). The Status of Altruism. *Mind*, *92*(366), 204–218.

Roughley N. (2017). The Empathy in Moral Obligation: An Exercise in Creature Construction. In N. Roughley & T. Schramme (Eds.), *Forms of Fellow Feeling: Empathy, Sympathy, Concern and Moral Agency* (pp. 265–291). Cambridge University Press.

Rowlands, M. (2012). *Can Animals Be Moral?* Oxford University Press.

——— (2013). A Response to Clement. *Journal of Animal Ethics, 3*(1), 15–24.

Royka, A., & Santos, L. R. (2022). Theory of Mind in the Wild. *Current Opinion in Behavioral Sciences, 45*, 101137.

Rudolf von Rohr, C., Burkart, J., & Shaik, C. (2011). Evolutionary Precursors of Social Norms in Chimpanzees: A New Approach. *Philosophy and Biology, 26*(1), 1–30.

Saito, A., Shinozuka, K., Ito, Y. *et al.* (2019). Domestic cats (*Felis catus*) discriminate their names from other words. *Scientific Reports, 9*, 5394.

Sapontzis, S. F. (1987). *Moral, Reason and Animals*. Temple University Press.

Sato, N., Tan, L., Tate, K., & Okada, M. (2015). Rats demonstrate helping behavior toward a soaked conspecific. *Animal Cognition, 18*, 1039–1047.

Sayre-McCord, G. (2023). Moral Realism. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*. Winter 2023 Edition. https://plato.stanford.edu/entries/moral-realism/.

Scanlon, T. (1988). The Significance of Choice. In S. M. McMurrin (Ed.), *The Tanner Lectures on Human Values 8* (pp. 149–216). The University of Utah Press.

——— (1998). *What We Owe to Each Other*. Belknap Press of Harvard University Press.

——— (2008). *Moral Dimensions: Permissibility, Meaning, Blame*. Belknap Press of Harvard University Press.

Schapiro, T. (1999). What Is a Child. *Ethics, 109*(4), 715–738.

Schino, G., & Aureli, F. (2009). Reciprocal Altruism in Primates: Partner Choice, Cognition, and Emotions. *Advances in the Study of Behavior, 39*, 45–69.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual Causality and Animacy. *Trends in Cognitive Sciences, 4*(8), 299–309.

Schünemann, B., Keller, J., Rakoczy, H. *et al.* (2021). Dogs distinguish human intentional and unintentional action. *Scientific Reports, 11,* 14967.

Schwab, C., & Huber, L. (2006). Obey or Not Obey? Dogs (Canis familiaris) behave differently in response to attentional states of their owners. *Journal of Comparative Psychology, 120*, 169–175.

Searle, J. (1994). Animal Minds. *Midwest Studies in Philosophy, 19*, 206–219.

Setiya, K. (2022). Intention. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*. Winter 2022 Edition. https://plato.stanford.edu/entries/intention/.

Shabo, S. (2012). Where Love and Resentment Meet: Strawson's Intrapersonal Defense of Compatibilism. *Philosophical Review, 121*(1), 95–124.

Shafer-Landau, R. (2003). *Moral Realism: A Defence*. Oxford University Press.

Shapiro, P. (2006). Moral Agency in Other Animals. *Theoretical Medicine and Bioethics, 27*(4), 357–373.

Shaw, R.C., & Harvey, A. 2020. Long-Term Memory for a Learned Behaviour in a Wild Bird. *Biology Letters, 16*.

Sher, G. (2006). *In Praise of Blame*. Oxford University Press.

Shoemaker, D. (2007). Moral Address, Moral Responsibility, and the Boundaries of the Moral Community. *Ethics, 118*(1), 70–108.

——— (2013). Qualities of Will. *Social Philosophy and Policy, 30*(1–2), 95–120.

——— (2015). *Responsibility from the Margins*. Oxford University Press.

Silk, J. B. (2002). The Form and Function of Reconciliation in Primates. *Annual Review of Anthropology, 31*, 21–44.

Silk, J. B., Brosnan, S. F., Vonk, J., Henrich, J., Povinelli, D. J., Richardson, A. S., Lambeth, S. P., Mascaro, J., & Schapiro, S. J. (2005). Chimpanzees are indifferent to the welfare of unrelated group members. *Nature, 437*(7063), 1357–1359.

Sima, M. J., Matzinger, T., Bugnyar, T., & Pika, S. (2017). Reconciliation and Third-Party Affiliation in Carrion Crows. *Ethology, 124*(1), 33–44.

Simmons, A. (2013). In Defense of the Moral Significance of Empathy. *Ethical Theory and Moral Practice, 17*, 97–111.

Simner, M. L. (1971). Newborn's Response to the Cry of Another Infant. *Developmental Psychology, 5*, 136–150.

Simoneau-Gilbert, V. (2024). Can Animals Be Moral Agents? Why the Debate Matters for Animal Ethics. In Y. Athanassakis, R. Larue & W. O'Donohue, (Eds.), *The Plant-Based and Vegan Handbook* (pp. 51–67). Springer.

Singer, P. (1975). *Animal Liberation*. Avon Books.

——— (1993). *Practical Ethics*. Cambridge University Press.

——— (2006). Morality, Reason, and the Rights of Animals. In S. Macedo & J. Ober (Eds.), *Primates and Philosophers: How Morality Evolved* (pp. 140–158). Princeton University Press.

Singer, T., & Klimecki, O. (2014). Empathy and Compassion. *Current Biology, 24*(18), R875–R878.

Sinhababu, N. (2024). The Epistemic Argument for Hedonism. In S. Chakraborty (Ed.). *Human Minds and Cultures* (pp. 137–158). Springer.

Skorupski, J. (1999). Irrealist Cognitivism. *Ratio, 12*(4), 436–459.

Slote, M. (2009). *Moral Sentimentalism*. Oxford University Press.

Smetana, J., & Braeges, J. (1990). The Development of Toddlers' Moral and Conventional Judgments. *Merrill-Palmer Quarterly, 36*(3), 329–346.

Smith, A. (2007). On Being Responsible and Holding Responsible. *The Journal of Ethics, 11*(4), 465–484.

Smith, A. (1759). *The Theory of Moral Sentiments*, D. D. Raphael and A. L. Macfie (Eds.). Oxford University Press. Edition of 1976.

Smith, H. (1983). Culpable Ignorance. *Philosophical Review, 85*, 449–487.

Smith, J. (2018). The Perceptibility of Emotions. In H. Naar & F. Teroni (Eds.), *The Ontology of Emotions* (pp. 130–148). Cambridge University Press.

Solomon, R. (1973). Emotions and Choice. *The Review of Metaphysics, 27*(1), 20–41.

——— (1999). The Philosophy of Emotions. In M. Lewis, J. Haviland-Jones & L. Barrett, (Eds.), *Handbook of Emotions* (pp. 3–16). The Guilford Press.

——— (2008). *True to Our Feelings: What Our Emotions Are Really Telling Us*, Oxford University Press.

Southwood, N. (2011). The Moral/Conventional Distinction. *Mind, 120*(479), 761–802.

Srinivasan, A. (2017). The Aptness of Anger. *The Journal of Political Philosophy, 26*(2), 123–144.

Stapleton, L. (2018). Animals, Machines, and Moral Responsibility in a Built Environment. *Philosophy Honors Projects*, *12*.

Stich, S. P. (1979). Do Animals Have Beliefs? *Australasian Journal of Philosophy, 57*(1), 15–28.

Stout, R. (2010). Seeing the Anger in Someone's Face. *Proceedings of the Aristotelian Society, Supplementary Volumes, 84*, 29–43.

Strawson, S. (1962). Freedom and Resentment. Reprinted in Watson, G. (Ed.) (1982), *Free Will* (pp. 59–80). Oxford University Press.

Suddendorf, T., & Whiten, A. (2007). Great Ape Cognition and the Evolutionary Roots of Human Imagination. In I. Roth (Ed.), *Imaginative Minds* (pp. 31–59). The British Academy.

Svärd, P. (2013). Animal National Liberation? *Journal of Animal Ethics, 3*(2), 188–200.

Svenaeus, F. (2014). The Phenomenology of Suffering in Medicine and Bioethics. *Theoretical Medicine and Bioethics, 35*, 407–420.

Tajima, T., & Kurotori, H. (2010). Nonaggressive Interventions by Third Parties in Conflicts Among Captive Bornean Orangutans (Pongo Pygmaeus). *Primates, 51*(2), 179–182.

Takagi, S., Saito, A., Arahori, M. *et al.* (2022). Cats learn the names of their friend cats in their daily lives. *Scientific Reports, 12*, 6155.

Talbert, M. (2014). The Significance of Psychopathic Wrongdoing. In T. Schramme, (Ed.), *Psychopathy and Moral Incapacity* (pp. 275–300). The MIT Press.

Tangney, J. P., Stueig, J., & Mashek, D. J. (2007). Moral Emotions and Moral Behavior. *Annual Review of Psychology, 58*(1), 345–372.

Tappolet, C. (2017). Self-Control and Akrasia. In M. Griffith, K. Timply & N. Levy (Eds.), *Routledge Companion to Free Will* (pp. 565–576). Routledge.

——— (2022). Evaluative *vs.* Deontic Concepts. In H. LaFollette (Ed.), *International Encyclopedia of Ethics* (pp. 1791–1799). Wiley.

Taylor, G. (1975). Justifying the Emotions. *Mind, 84*(335), 390–402.

Teroni, F. (2007). Emotions and Formal Objects. *Dialectica*, *61*(3), 395–415.

Thomason, K. (2020). The Moral Necessity of Anger. In C. D. Lewis & G. L. Bock (Eds.), *The Ethics of Anger* (pp. 83–101). Lexington Books.

Tiboris, M. (2014). Blaming the Kids: Children's Agency and Diminished Responsibility. *Journal of Applied Philosophy, 31*(1), 77–90.

Todd, P. (2016). Strawson, Moral Responsibility, and the "Order of Explanation": An Intervention. *Ethics, 127*(1), 208–240.

Traina, C. L. H. (2009). Children and Moral Agency. *Journal of the Society of Christian Ethics*, *29*(2), 19–37.

Trivers, R. (1971). The Evolution of Reciprocal Altruism. *The Quarterly Review of Biology, 46*(1), 35–57.

Turiel, E. (1982). *The Development of Social Knowledge: Morality and Convention*. Cambridge University Press.

Turner D. C. (2021). Unanswered Questions and Hypotheses about Domestic Cat Behavior, Ecology, and the Cat–Human Relationship. *Animals, 11*(10), 2823.

Tye, M. (2008). The Experience of Emotion: An Intentionalist Theory. *Revue internationale de philosophie, 62*(242), 25–47.

Vaish, A., Missana, M., & Tomasello, M. (2011). Three-year-old children intervene in third-party moral transgressions. *The British journal of Developmental Psychology*, *29*(Pt 1), 124–130.

Udell, M. A. R., Dorey, N. R., & Wynne, C. D. L. (2008). Wolves outperform dogs in following human social cues. *Animal Behaviour, 76*(6), 1767–1773.

Udell, M. A. R., Spencer, J. M., Dorey, N. R., & Wynne, C. D. L. (2012). Human-socialized wolves follow diverse human gestures… and they may not be alone. *International Journal of Comparative Psychology, 25*(2), 97–117.

Valentini, L. (2014). Canine Justice: An Associative Account. *Political Studies*, *62*(1), 37–52.

Van Bourg J., Patterson J. E., Wynne C. D. L. (2020). Pet dogs *(Canis lupus familiaris)* release their trapped and distressed owners: Individual variation and evidence of emotional contagion. *PLoS ONE, 15*(4), e0231742.

Van Hooft, S. (1998). Suffering and the Goals of Medicine. *Medicine, Health Care, and Philosophy, 1*, 125–131.

van Roojen, M. (2023). Moral Cognitivism *vs.* Non-Cognitivism. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*. Summer 2024 Edition. https://plato.stanford.edu/entries/moral-cognitivism/.

Van Schoelandt, C. (2018). Moral Accountability and Social Norms. *Social Philosophy and Policy*, *35*(1), 217–236.

Van Wolkenten, M., Brosnan, S. F., & de Waal, F. B. M. (2007). Inequity Responses of Monkeys Modified by Effort. *PNAS, 104*(47), 18854–18859.

Vezirian, K., Bègue, L., & Bastian, B. (2024). Mindless Furry Test-Tubes: Categorizing animals as lab-subjects leads to their mind denial. *Journal of Experimental Social Psychology, 114*, 1–11.

Vitale, K. R., Behnke, A. C., & Udell, M. A. R. (2019). Attachment Bonds Between Domestic Cats and Humans. *Current Biology, 29*(18), R864–R865.

Wallace, R. J. (1994). *Responsibility and the Moral Sentiments.* Harvard University Press.

Walters, K., King, C., Scolaro, C. L. C., & Shyan-Norwalt, M. (2020). Reconciliation in Domestic Dogs (Canis Familiaris): Evidence for the Uncertainty Reduction Hypothesis. *Applied Animal Behaviour Science, 226*, 104987.

Wascher C. A. F., & Bugnyar, T. (2013). Behavioral Responses to Inequity in Reward Distribution and Working Effort in Crows and Ravens. *PLoS ONE, 8*(2), e56885.

Watson, G. (2004). *Agency and Answerability: Selected Essays*. Oxford University Press.

Weiner, B., & Peter, N. (1973). A Cognitive-Developmental Analysis of Achievement and Moral Judgments. *Developmental Psychology, 9*(3), 290–309.

Widerker, D. (2000). Frankfurt's Attack on the Principle of Alternative Possibilities: A Further Look. *Noûs*, *34*(s14), 181–201.

Wilkinson, G. (1984). Reciprocal Food Sharing in the Vampire Bat. *Nature, 308*, 181–184.

Williams, B. (1973). A Critique of Utilitarianism. In J. J. C. Smart & B. Williams (Eds.), *Utilitarianism: For and Against* (pp. 77–150). Cambridge University Press.

Zagzebski, L. (2003). Emotion and Moral Judgement. *Philosophy and Phenomenological Research, 66*(1), 104–124.

Zawidski, T. (2013). *Mindshaping: A New Framework for Understanding Human Social Cognition*. MIT Press.

Zelazo, P. D., Helwig, C. C., & Lau, A. (1996). Intention, Act, and Outcome in Behavioral Prediction and Moral Judgment. *Child Development, 67*(5), 2478–2492.

Zola, E. (1885). *Germinal*. Oxford World's Classics. Translation by Peter Collier. Edition of 2008.