PhD DISSERTATIONS OF
NORTHWESTERN ITALIAN PHILOSOPHY CONSORTIUM – FINO

Sensitive to Reasons: Moral Intuition and the Dual Process Challenge to Ethics

Dario Cecchini

(2022)

Examination Committee: Prof. Maria Silvia Vaccarezza, Dr. Hanno Sauer, Prof. Fabrice Teroni

UNIVERSITÀ DEGLI STUDI DI GENOVA
UNIVERSITÀ DEGLI STUDI DI TORINO
UPO UNIVERSITÀ DEL PIEMONTE ORIENTALE
UNIVERSITÀ DEGLI STUDI DI PAVIA

Sensitive to Reasons

Moral Intuition and the Dual Process Challenge to Ethics

Dario Cecchini

Acknowledgments

The path to this dissertation has been long, intense, and full of twists. I would have never neither started nor finished this project without the generous help and support of a considerable number of people.

The first mention goes to my supervisor Maria Silvia Vaccarezza who encouraged me to work on moral intuition and moral psychology. Encountering her Neo-Aristotelian approach has been fundamental in making my metaethical view more concrete. I am very grateful to Maria Silvia for patiently reading and commenting on every draft and fool idea that preceded this final version.

I also wish to thank Hanno Sauer for accepting me as a visiting scholar in his research group on moral progress between February and June 2021. During these months spent at Utrecht University, Hanno's supervision has been very precious for my work. In particular, his advice has been fundamental for structuring and communicating my ideas at best.

Another special mention goes to Mauro Imbimbo who introduced me to moral philosophy at high school and followed my philosophical path over the years. I hope to have inherited his analytic approach and love for philosophical argumentation.

The spread of the pandemic made social life difficult in these years. Notwithstanding, I had the opportunity to work and exchange ideas with many wonderful friends and colleagues. I thank (in alphabetical order): Paolo Babbiotti, Federico Bina, Charlie Blunden, Michele Bocchiola, Giulia Cantamessi, Chiara Cecconi, Mario De Caro, Ian Carter, Michel Croce, Arianna Dini, Francesco Emanuelli, Corrado Fumagalli, Roberto Gronda, Carline Klijnman, Matilde Liberti, Sarin Marchetti, Marco Miglino, Edoardo Peruzzi, Matteo Rategni, Paul Rehren, Damiano Simoncelli, Luca Stroppa, Lorenzo Testa, Silvano Zipoli Caiani, Federico Zuolo.

Portions of this thesis have been presented at numerous conferences and seminars (online and in-person). Among them I mention the *Philosophy in Progress Conference* at the University of Nottingham, the *10th European Congress of Analytic Philosophy* organized by Utrecht University, the 2020 *Annual National Conference of PhDs in Philosophy* organized by the Collegio Fondazione San Carlo, the *47th Annual Meeting of*

the Society for Philosophy and Psychology, the *2021 annual conference of the Italian Society for Analytic Philosophy* held in Noto, the Ethics and Politics work in progress seminar of FINO, the work in progress seminar of the department of philosophy at the University of Genoa, and the reading group on moral progress organized by Utrecht University. I thank the organizers of such appointments for the opportunity to present my ideas and the audience for the helpful comments.

Two chapters in this thesis are based on previously published articles. Chapter 1 is a shortened version of "Moral intuition, strength, and metacognition" *Philosophical Psychology* (2022). Chapter 2 is a slightly revised and extended version of "Dual-Process Reflective Equilibrium: Rethinking the Interplay between Intuition and Reflection in Moral Reasoning", *Philosophical Explorations* (2021), 24 (3): 295-311. I am thankful to the anonymous referees of these journals for their valuable comments.

My final thoughts go to those people without whom this PhD would not have been even imaginable: my parents, Roberto and Silvia, and my grandmothers, Germana and Maria Letizia, whose unconditional support and love gave me the chance to pursue my dreams; my brother Emilio and my friends from Florence, who made these hard times sweeter.

Contents

Introduction

Moral intuition

On February 3, 2016, Giulio Regeni, twenty-eight years old, was found dead on the side of a road in Cairo. The body presented signs of extreme torture: wounds on the face and the rest of the head, swelling in the hands, and bruises all over. After a few days, an autopsy confirmed that Giulio had been severely tortured with extreme violence before dying. Giulio had been an Italian PhD student at Cambridge University. He had gone to Cairo to research Egyptian trade unions. The motives of Giulio's death are still unclear, and his family is fighting for truth and justice (Deffendi and Regeni 2020).

It is likely that no one who hears Giulio's story needs to reflect to conclude that what happened was a crime against humanity. Knowing the vivid details is sufficient to understand the fact is morally wrong. This phenomenon is called *moral intuition* by philosophers and psychologists.

This dissertation constitutes a comprehensive philosophical and empirically informed investigation of moral intuition. I aim to explain the psychological features of moral intuition, its role in moral reasoning, its cognitive function, and how it guides conduct.

The concept of moral intuition has received much attention in moral psychology in recent decades. In the literature, moral intuition is generally understood as an immediate and compelling representation of a moral fact or proposition. However, beyond this minimal characterization, the notion takes on different meanings. In particular, two widespread views appear to be at odds. On the one hand, some philosophers (Huemer 2005, Bengson 2015, Audi 2015, Chudnoff 2013) understand moral intuition as a kind of "intellectual perception" and stress its importance for moral theory; according to this view, moral intuition is an "intellectual given" and it can be used as evidence for certain moral principles. On the other hand, some psychologists and empirically minded philosophers (Haidt 2001, Sinnott-Armstrong, Young and Cushman 2010, Greene 2014, Railton 2014) emphasize how moral intuitions are influenced by

emotions and by cultural and individual differences. In the first part of this dissertation, I argue that these two opposite views share common ground. According to the conceptualization I defend, moral intuition is defined by two fundamental features: *automaticity* and *strength*.

Automaticity is a complex concept that comprises a cluster of mental properties such as unconsciousness, being uncontrolled, quickness, and efficiency (Bargh 1992, Moors 2016). Moral intuitions are automatic mental states because they derive from processes that are largely automatic—that is, tending to be not controlled, fast, and effortless. Automaticity captures the spontaneous and immediate aspect of moral intuition, which distinguishes moral intuition from slower and effortful reflective judgments.

Beyond being based on automatic processes, moral intuitions are also *strong* mental states. That is, intuitive representations, such as the intuition that torture is wrong, are experienced as compelling such that they incline the subject to assent to their content. Intuitive strength distinguishes moral intuitions from "shallow" automatic mental states, such as guesses or quick hypotheses.

Though scholars widely agree that intuitions tend to be strong mental states, the strength of moral intuitions in particular has not received sufficient attention. More specifically, an account of moral intuitions that explains why they vary in strength and why they are stronger than shallow automatic responses is missing. I fill this gap by defending a *metacognitive account* of moral intuition. According to my account, intuitive strength denotes the degree of *confidence* of a subject in a moral representation. I show how this account accommodates the phenomenology of strong intuitions, sheds light on their role in moral reasoning, and is supported by empirical evidence.

Two processes into one moral mind

The history of moral psychology in the last forty years is not devoid of twists. In the 1980s and 1990s, many philosophers (Korsgaard 1996, Scanlon 1998) and psychologists (Kohlberg 1981, Turiel 1983) agreed that ethics is a deliberative practice. According to this view of ethics, not only is reasoning the main path to moral judgments and decisions, but reasoning is the process that confers moral meaning on conduct and behavior. In other words, a decision has no moral status unless motivated by an explicit process of reasoning, in which an agent balances the different values and principles at

play.

In the 1990s and 2000s, the "affect revolution" in moral psychology (and other disciplines) undermined the widespread rationalist view of ethics. Converging evidence suggested that moral judgment and behavior go hand in hand with emotions (for example, indignation, compassion, anger) rather than reasoning. According to this sentimentalist view of ethics, feelings play the leading part in moral thinking (Haidt 2001, Nichols 2004, Prinz 2007). However, in recent times, some re-evaluations of the evidence have put moral reasoning back in charge (Sauer 2017, May 2018): the role of emotion in ethics is less dominant than it appeared, and reflection can educate and regulate emotions to make them better fit in particular situations. Perhaps, these re-evaluations suggest, emotion and reasoning coexist and jointly contribute to moral judgment.

Influenced by "dual process" theories of the mind (Kahneman 2011, Evans and Stanovich 2013), some authors have come to a synthesis: moral thinking is governed by two distinct types of processes: one fast, automatic, and emotional, and one slow, reflective, and controlled (Greene 2013, Cushman 2013, Sauer 2019, Craigie 2011). However, such a conclusion leaves some questions open: How do the two processes interact? Do they coexist in peace or conflict? Call these issues the *dual process challenge* to ethics.

According to Greene's influential theory, automatic and reflective processes have conflicting outcomes: automatic thinking leads to characteristically deontological judgments, whereas reflection favors characteristically utilitarian judgments (Greene 2013). This view finds support in distinct lines of evidence involving "sacrificial" moral dilemmas, in which deontological and utilitarian judgments are opposed to each other. A classic example is the footbridge version of the trolley dilemma, in which a bystander faces two incompatible options: pushing a fat man off a footbridge to stop a trolley that is headed directly toward five men, or not doing anything and letting five men die (Thomson 1985, 1409). Much evidence reveals that the judgment that pushing the man is not permissible is fast and cognitively effortless and involves the activation of brain areas related to emotion; by contrast, the judgment that it is permissible is slower and requires effort (Greene 2014).

In line with Greene's theory, Cushman's dual process model states that automatic and reflective thinking can conflict because they are based on opposite types of decision-making (Cushman 2013). Reflective moral responses are favored by "model based"

decision-making, according to which the subject compares different courses of action and chooses the one they expect to produce the best outcome. By contrast, automatic moral decisions are "model free"—that is, driven by an assessment of the action immediately available and based on trial-and-error learning. This theory explains why automatic processes tend to generate deontological judgments and reflective processes generate utilitarian judgments.

In contrast to those who portray this bellicose picture of the moral mind, some authors have pointed out that automatic cognition and reflection cooperate most of the time. Specifically, they emphasize how automatic processes can be shaped by conscious beliefs through habituation and education. For instance, certain outcomes of utilitarian reasoning can become automatic if repeatedly arrived at (Sauer 2017, Fine 2006, Pizarro and Bloom 2003). Another relevant case showing the influence of reasoning on automatic processes is the regulation of emotions (Helion and Pizarro 2015, Helion and Ochsner 2018), through which a subject can change the intensity, duration, or type of an affective reaction to make a more appropriate moral judgment. For example, scholars have documented that encouraging people to reappraise their feeling of disgust makes their condemnation of moral violations less severe (Feinberg, et al. 2012).

In sum, much has been done in recent years to advance our understanding of how moral reflection can affect automatic processes. However, the regulation (or overriding) of automatic cognitions by reflection cannot be the only kind of cooperation between the two types of processes. Sometimes automatic cognition regulates moral reasoning. Indeed, automatic processes do not always lead to impulsive or spontaneous reactions; they can lead to extensive reflection. If sufficiently sensitive to the features of a situation, a moral agent knows when and how to switch from the automatic to the deliberative mode. However, in what circumstances? With what capacities? This "bottom up" interaction between automatic cognitions and deliberation has been insufficiently investigated. This dissertation addresses this unexplored issue in moral psychology.

One of my core claims is that moral intuition plays a pivotal role at the interface between automatic and deliberative processing. More specifically, I argue that the level of confidence with which automatic intuitions are generated is a crucial predictor of whether reflection is activated. In simpler words, in normal conditions, the stronger and more confident an intuition, the less likely a subject engages in moral reasoning;

conversely, weaker and less confident intuitions favor the activation of deliberative processes.

My contribution to the dual process challenge supports the idea that automatic and reflective processes tend to cooperate in the moral mind. Rather than conflicting with automatic cognitions, moral reasoning strongly depends on an automatic sensitivity to reasons. On the other side, the purpose of moral reflection is to rationalize pre-reflective intuitions by providing articulated and accessible reasons.

Moral intuitionism

A widespread pessimism has surrounded moral intuitions in recent years. It has been argued that moral intuitions are not reliable because they are influenced by framing effects (Rehren and Sinnott-Armstrong 2021), hypersensitive emotions such as disgust (Kelly 2011), or hyposensitive emotions such as compassion (Västfjäll, et al. 2014). All this empirical evidence undermines *moral intuitionism*—that is, the claim that accepting moral intuitions is justified under normal conditions.

In the second part of this dissertation, I consider how moral philosophers can vindicate the rationality of moral intuitions in the face of the recent skeptical challenges. For this purpose, I evaluate different strategies. First, I discuss whether subjects can track the reliability of moral intuitions through their level of confidence. I show that subjects can prevent possible biases and avoid irrelevant factors if the probability that moral intuitions are true is proportional to the subjects' level of confidence. However, I also argue that this hypothesis is insufficiently supported by current evidence. Then, I consider whether moral intuitionism can be supported on different grounds from the pure truth-conduciveness of moral intuitions. In particular, I base an argument on the idea that accepting moral intuitions is the most rational option that a subject has, given her limited cognitive resources and the illegitimacy of deferring to others.

My understanding of intuitionism diverges from the common conception of moral intuitionism as "foundationalism" (Sturgeon 2002, Väyrynen 2008, Audi 2004). The purpose of this work is not to discuss whether moral intuitions can ultimately ground moral knowledge to respond to the discussed "regress of justification problem". Rather, here I understand accepting moral intuitions as a *reasoning conduct*—that is, a way in which a reasoner manages her cognitive resources to reach a reflective goal. Reasoning conducts include moral theorizing but also ordinary reasoning in which a subject has to

make a moral decision. The more modest aim of this study is to evaluate whether accepting moral intuitions is epistemically legitimate (i.e., rational), in light of the goals of a correct moral inquiry.

Investigating the role of intuition in moral knowledge may sound premature since the existence of moral knowledge is a matter of metaethical dispute. According to nihilist accounts of ethics, such as error theory or emotivism, there are no ethical truths and, thus, speaking about moral knowledge is not legitimate. This study does not tackle metaphysical questions concerning the existence of moral facts; the issue is too big to be seriously discussed in a dissertation centered on moral psychology. I assume that moral knowledge is possible. This makes the second part of the dissertation incompatible with nihilst views of ethics. Nevertheless, since I adopt no specific metaphysical view of moral facts, moral intuitionism, as defined here, might be compatible with different theories of moral truth and objectivity (e.g., naturalism, non-naturalism, expressivism, and constructivism).

Automaticity in action

One of the most robust conclusions of contemporary psychology is that deliberation is expensive: reasoning requires time and much attention. In addition, there is ample evidence that human beings are cognitive misers: they tend to rely on effortless automatic thinking and activate reflection very parsimoniously (Kahneman 2011, Stanovich 2018). Given this, it is not surprising that most moral conduct is based on automatic processes, rather than reasoning (Haidt 2001, Narvaez and Lapsley 2005). In everyday life, people deliberate rarely and make many moral decisions by force of habit or based on emotions or intuitions. Sometimes, automatic processes guide moral conduct even at odds with beliefs and reasoning. This is attested to by the large literature on implicit biases, which predict people's behavior more accurately than conscious beliefs do (Frankish 2016). However, automatic processes are not necessarily synonymous with impulsive and bad conduct but can generate spontaneous competent behavior. For example, ordinary experience tells us that there are cases of "inadvertent virtue"—that is, acting rightly in spite of bad conscious beliefs (Arpaly 2003). Automaticity has two sides: occasionally, automatic processes lead to competent and virtuous behavior, but often their outcomes are impulsive and biased (Brownstein 2018).

That automatic processes, sometimes in conflict with deliberative attitudes, pervade the motivation of moral behavior raises a philosophical challenge. Traditional theories of moral motivation are based on the concepts of conscious desires, beliefs, and deliberation. These concepts appear inadequate to explain how moral action can be motivated automatically—that is, without the mental effort of deliberation. Another relevant problem is how to explain how moral behavior can be skillful yet automatic. How can emotions and intuitions be educated and trained such that they produce virtuous behavior? The third part of my dissertation addresses these challenges.

In response to the *automaticity challenge* (Sauer 2017, 51-83), I defend a theory of moral motivation based on the concept of *caring*—that is, a sentiment of regard or concern toward an object. The concept of caring has been introduced by some authors in philosophy of mind and action (Shoemaker 2003, Jaworska 2007, Seidman 2009, 2016, Brownstein 2018, 101-122); however, the notion needs to be refined. I develop a more detailed account of caring, and I apply it to the moral domain. I show how sentiments of caring about moral standards, sometimes in conflict with reflective desires and beliefs, can motivate action.

After defending a caring-based account of motivation, I outline a theory of *moral sensitivity* to explain how behavior can be automatic and competent at the same time. I understand moral sensitivity as the possession of a set of skills regulated by a moral standard. Countering a widespread objection (Zagzebski 1996, Rees and Webber 2014, Small 2021), I show that the concept of skill is not in conflict with that of moral motivation; sentiments of caring toward the standards of a domain of performance are required for the learning, exercise, and possession of skill.

Plan of the work

This dissertation is a contribution to the field of empirically informed metaethics (Prinz 2015), which combines the rigorous conceptual clarity of traditional metaethics with a careful review of empirical evidence. More specifically, this work stands at the intersection of moral psychology, moral epistemology, and philosophy of action.

The study comprises six chapters on three distinct (although related) topics. Each chapter is structured as an independent paper and addresses a specific open question in the literature.

As mentioned, the first part concerns the psychological features and cognitive function of moral intuition. Chapter 1 ("Moral intuition, strength, and metacognition") is focused on the concept of *intuitive strength*, which is one of the defining features of moral intuition. I provide a metacognitive account of intuitive strength and show why such a view is preferable to emotional or quasi-perceptual accounts. Then, in Chapter 2 ("Dual process reflective equilibrium"), I will discuss the interplay between intuition and reflection in moral reasoning. I will contend that the influential "default-interventionist" model of reasoning, theorized by Greene (2013), is insufficiently supported by the evidence. In light of some recent studies, I outline an account of moral reasoning in which intuition and reflection are not in conflict but cooperate to reach a reflective goal. I call this model *dual process reflective equilibrium*.

The aim of the first part is descriptive, i.e., it argues for an accurate understanding of moral intuition and reasoning in light of the available empirical evidence. In contrast, the second part addresses a normative question: is a subject epistemically justified in forming a belief on the basis of a moral intuition? Skeptics of moral intuition argue that accepting moral intuitions should be the exception rather than the rule to the extent that epistemically defective processes determine the content of moral intuitions. Chapter 3 ("Moral intuitionism and the reliability challenge") introduces the recent empirical challenges to the reliability of moral intuitions and elaborates a promising strategy for defending intuitionism. In short, I consider whether subjects can track the reliability of their intuitions with their confidence. In Chapter 4 ("The argument from limited cognitive resources"), I evaluate a different strategy to defend moral intuitionism. Specifically, I develop an argument according to which accepting moral intuitions is legitimate because it is the most rational option that a subject has, given her limited resources.

The third and final part of the dissertation concerns the role of moral intuitions in action. As mentioned, the influence of automatic processes on moral conduct raises different challenges to moral philosophy. The first challenge is to explain how a subject can be motivated by certain values without the mental effort of deliberation. Chapter 5 ("Caring, moral motivation, and automatic conduct") tackles this issue. Chapter 6 ("Moral sensitivity as skillful automaticity") aims to explain how moral agents can be sensitive to good reasons through automatic mental processes.

I

MORAL INTUITION AND REASONING

Chapter 1

Moral intuition, strength, and metacognition[1]

1. Introduction

The concept of moral intuition has received much attention in moral psychology and philosophy in recent decades. In the literature, moral intuitions are generally understood as fast and automatic moral representations that spontaneously arise in the mind (Haidt, 2001). They contrast with reflective judgments, which are slower and require deliberation. Moreover, many authors state that moral intuitions are *strong*, *compelling*, and *stable* mental states:

> When we refer to moral intuitions, we will mean *strong*, *stable*, immediate moral beliefs. (Sinnott-Armstrong, Young and Cushman 2010, 246, my italics)

> It [moral intuition] can persist in the face of contrary conscious judgment, while still remaining in some degree *compelling* or motivating and thus such that we are *reluctant to give it up or ignore it* (Railton 2014, 815, my italics)

> Such [intuitive] appearances are spontaneous and *compelling* propositionally contentful experiences that result from merely thinking about a proposition or a set of propositions. (Kauppinen 2013, 361, my italics)

Despite the wide consensus on this aspect, the strength of moral intuitions has not received sufficient attention. In recent decades, moral psychology has focused mostly on the automatic aspect of moral intuitions, neglecting their strength. It is unclear why moral intuitions vary in strength and why they are stronger than other shallow automatic responses. This chapter addresses these questions.

Some philosophical accounts of moral intuition understand intuitive strength as the emotional intensity of a moral representation (Kauppinen 2013, Railton 2014) or as

[1] This chapter is a shortened version of the work published in Cecchini (2022).

"presentational phenomenology" (Chudnoff 2013, Bengson 2015). In contrast with these views, I offer a metacognitive account of intuitive strength, according to which the strength of moral intuitions denotes the level of confidence of a subject. I will define confidence as a metacognitive appraisal determined by the fluency with which a subject processes information from a morally salient stimulus. I will show how this naturalist account explains the phenomenology of strong intuitions and their cognitive function.

The chapter proceeds as follows. In Section 2, I introduce the concept of moral intuition and its salient psychological features: automaticity and strength. Then, in Section 3, I describe the importance of strength of moral intuitions, its phenomenology, and its cognitive function. In Section 4 I defend the metacognitive account of moral intuition and review the empirical evidence for the view. Then, I distinguish intuitive strength from emotional intensity (Section 5) and presentational phenomenology (Section 6). Finally, I consider some open questions and future lines of empirical research (Section 7) such as the determinants of cognitive fluency in moral intuitions and the rationality of intuitive confidence.

2. Moral intuition: automaticity and strength

A moral intuition is defined by its moral content and certain psychological features. The moral content is constituted by a conscious representation of something or someone as *wrong*, *right*, *good*, or *bad*. For example, moral intuitions can represent a particular moral fact (e.g., what happened in Egypt to Giulio Regeni is wrong), a general proposition (e.g., happiness is the most fundamental good), or a midlevel principle (e.g., breaking promises is wrong).

Concerning its psychological features, no wide consensus has been reached in the literature about the type of mental state that constitutes moral intuition. Scholars in moral philosophy and psychology disagree on whether moral intuition is a type of belief (Sinnott-Armstrong, Young and Cushman 2010), emotion (Kauppinen 2013), or intellectual seeming (Huemer 2005, Bedke 2008). However, one can reasonably assume that at least two salient properties characterize a moral intuition as intuition, independently of what type of mental state it can be reduced to. Arguably, these salient mental features are *automaticity* and *strength*.

Moral intuition can be classified as an *automatic* mental state insofar as it derives from processes that are to a large extent *autonomous*—that is, not requiring conscious

guidance once triggered (Bargh 1992, Evans and Stanovich 2013). An intuitive representation is typically generated by a morally salient stimulus (e.g., two hoodlums torturing a cat) that triggers a series of unconscious mental associations leading to the moral representation (e.g., the representation of the act as wrong).[2] This mental process tends to be fast and not controlled by the subject. Moreover, the formation of a moral intuition requires little cognitive effort since the mental process does not need conscious guidance.

The automatic information processing behind an intuition is not always retrospectively accessible to the subject. This is attested to by some studies on moral judgment of taboo violations (Haidt 2001) and studies testing the doctrine of double effect (Cushman, Young and Hauser 2006, Hauser, et al. 2007). A popular and commonly discussed example of inarticulate moral intuition concerns the story of two siblings (Julie and Mark) that decide to have sex for fun, just once in their life, and without any apparent biological or psychological consequences.[3] In Haidt and colleagues' study, many of the interviewed subjects had the intuition that Julie and Mark's behavior is wrong, but they were not able to explain why they believed that it is wrong.[4] This opacity to introspection has led some authors to define an intuition as "a sense of knowing without knowing" (Epstein 2010, 296).

---

[2] Simon (1992, but see also Seligman and Kahana 2009) describes the unconscious mental process behind an intuition as a kind of *recognition*: a certain situation provides a cue, the cue gives the subject access to information stored in memory and the information provides the answer to the situation (155).

[3] Here is the whole story: "Julie and Mark are brother and sister. They are travelling together in France on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At the very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy making love, but they decide not to do it again. They keep that night as a special secret, which makes them feel even closer to each other. What do you think about that? Was it OK for them to make love?" (Haidt 2001, 814).

[4] Here the point is not whether the subjects were right in judging the behavior (probably, they were, since Julie and Mark's behavior is risky and irresponsible), but rather how able they are in defending their intuitions. However, I will have more to say about the rationalization of moral intuition in the next chapter.

Automaticity is a well-studied mental phenomenon beyond the moral domain. Recent reviews point out that automaticity should not be considered as a monolithic concept but as an umbrella term comprising different related mental properties such as unconsciousness, lack of control, efficiency, and quickness (Evans and Stanovich 2013, Moors 2016). Moreover, the evidence suggests that automaticity is not an absolute property but gradable and context sensitive. However, it is largely accepted that intuitions are more automatic mental states than are paradigmatic reflective moral judgments, which are slower and require effort.

The concept of automaticity captures the spontaneous immediate aspect of moral intuition. However, stating that moral intuition stems from automatic mental processes is not enough to capture its psychological features. Paradigmatic cases of moral intuitions are also "compelling": intuitions capture the subject's attention such that their content is hard to ignore; as a result, the subject is inclined to assent to the content of the intuitions, sometimes even in the face of contrary reflective considerations (Kauppinen 2013, Railton 2014). This felt compellingness, which I denote as *intuitive strength*, is the second essential feature of moral intuition and the subject of the present chapter.

3. The importance of strength in moral intuition

Intuitive strength is a gradable property. This means that a subject can have intuitions that vary in strength along a continuum (Andow 2016). For example, a subject can have a very strong intuition that killing babies is wrong and a weaker intuition that turning the switch in the trolley dilemma is permissible. Regardless, what is conventionally called intuition must possess some degree of strength. Arguably, a moral intuition is experienced as stronger than "shallow" automatic responses, such as guesses or quickly generated hypotheses (Bengson 2015). By definition, if one responds to a problem by guessing, one does not find one's own answer particularly convincing. Similarly, if one formulates a quick hypothesis, one will be very disposed to revise it in case of counterevidence. Compared with these experiences, intuitive representations appear more insightful and convincing; for this reason, the subject is more inclined to assent to them and more reluctant to abandon them in case of counterevidence.

The intrinsic features of intuitive strength are not easy to describe from a phenomenological point of view, since moral intuitions are very diverse. How a subject experiences a strong intuition can vary according to its *object*; for example, intuitions of

general moral propositions (e.g., that solidarity is good) are usually "colder" than intuitions of particular moral facts. In addition, the felt strength of an intuitive representation can be very different according to the *context*; for instance, having the intuition that torture is wrong while constructing a moral theory and having the same intuition while reading the story of Giulio Regeni might be very different experiences. However, in both cases, the subject's immediate moral thought is accompanied by a sense of veridicality and credibility.

What is also constant, regardless of the object of the intuition and the context of reasoning, is the *cognitive* function of strong intuitions. More precisely, the strength of moral intuitions helps subjects assign credibility to certain moral contents. Through the perceived strength, a subject can assess the likelihood of certain moral representations and filter her beliefs accordingly. The stronger an intuition, the more the subject will be disposed to consider its content as true and endorse it. Thus, strong intuitions, such as the intuition that torturing is wrong, tend to be *stable*—that is, resistant to situational factors or counterevidence (Zamzow and Nichols 2009, Wright 2010, Wright 2013).

The cognitive guidance provided by intuitive strength influences moral reasoning. More specifically, intuitive strength appears to be an important factor in how moral reasoners regulate whether to activate reflection. In normal conditions, the occurrence of a strong intuition does not incline the subject to reflect on its content: to the extent that strong intuitions are perceived as likely and credible, she does not feel much pressure to justify their validity. This allows reasoners to spare cognitive resources for conclusions that appear more controversial. However, things change when a strong intuition is challenged. Since moral intuitions can be highly recalcitrant, if a strong intuition is questioned by another source, the reasoner will be reluctant to abandon it and will likely spend many cognitive resources (i.e., time, attention, and available information) to defend the content of the intuition.

Moral theorizing can be considered as a particular case of reasoning conduct in which a subject must manage certain cognitive resources to justify some moral propositions. Many authors consider the objects of strong intuitions (for example, the propositions "Torturing is wrong" and "Happiness is good") as fundamental moral truths and tend to justify them only if challenged. Therefore, even in moral theory, as a particular case of moral reasoning, reasoners tend to rely on intuitive strength to decide what moral propositions appear credible and activate cognitive resources accordingly.

Although strength is an essential phenomenological aspect of moral intuition and performs an important cognitive function, it has not received sufficient attention in moral psychology. An accurate descriptive account of moral intuition should explain why moral intuitions are strong mental states and why certain moral intuitions are stronger than other ones. In particular, to capture intuitive strength, an account should explain three things. First, it should capture the phenomenology of intuitive strength, in light of the diversity of moral intuitions. That is, it should explain why strong intuitions are experienced as more compelling than other automatic representations. Second, it should explain the cognitive function of intuitive strength—that is, how the strength of moral intuitions guides the subject in forming beliefs and activating reflection. Third, it must explain how intuitive strength combines with automaticity. In other words, it should explain how the strength of an intuition can be generated by automatic processes—that is, fast, unconscious, uncontrolled, and effortless processes.

4. A metacognitive account of intuitive strength

In this section, I argue for a *metacognitive account* of intuitive strength.[5] The explanation I offer is straightforward: the strength of moral intuitions denotes the level of subjective *confidence* about a certain moral representation. Intuitive confidence, as I understand it, results from the fluency with which an intuition is processed.

Before showing the main advantages of this view (4.2), I briefly introduce the literature on metacognition and metacognitive feelings (4.1).

4.1 From metacognition to metacognitive feelings

Metacognition is commonly known as the capacity of "thinking about thinking". More precisely, metacognition comprises a set of skills and strategies through which a subject can represent or evaluate her own thoughts in a context-sensitive way (Proust 2013, 4). The key characteristic of metacognitive judgments and processes is that they are not directly related to stimuli; they have mental representations as objects.

[5] The importance of metacognition for moral intuition has also been emphasized by Clavien and FitzGerald (2017). However, in my account, the content of a moral intuition does not need to be challenged by reflection for a subject to experience the strength of the intuition.

Little research has been done about the evolutionary origins of metacognitive skills. Some ethological evidence shows that, in addition to humans, only some primates exhibit metacognitive capacity (Proust 2013, 77-109). This suggests that metacognition is a quite recently evolved capability whose purpose is to allow creatures not to be completely stimulus bound (Metcalfe 2008). Metacognition enables a subject to suspend the immediate response that an afferent environment elicits, thus favoring self-regulation and hypothetical thinking. Perhaps, as has been recently argued, metacognition has a social function: it favors suprapersonal decision-making by enabling broadcasting and sharing of private mental states (Heyes, et al. 2020).

Metacognitive processes involve two different levels of cognition: an *object level*, constituted by various kinds of cognition, and a *metalevel*, which comprises the subject's normative standards and constraints (van Overschelde 2008). The metalevel *monitors* and *controls* the object level; that is, a subject evaluates whether the information processing fits some normative standards and can intervene to regulate it whenever required. This typically occurs in memory tasks (e.g., recalling capital cities) in which a subject monitors whether a certain solution sounds correct and adjusts it accordingly.

Importantly, for present purposes, it has been argued that normative standards do not need to be represented to monitor and control object-level representations (Proust 2013). A documented "core metacognition" develops early on in humans, and through it a subject can *automatically* evaluate and regulate her own cognitions (Goupil and Kouider 2019). Affect, for instance, is a relevant and documented mode by which people automatically assess their cognitive processing in a context-sensitive way (Efklides 2006); feelings of familiarity, knowledge, confidence, or difficulty are crucial to providing information to a subject about how properly she is conducting a cognitive task.

According to a large body of evidence, the most important determinant of metacognitive feelings is cognitive (or processing) *fluency*—that is, the sense of ease with which a subject processes information. It has been documented that this subjective experience tends to determine people's metacognitive assessments in a wide variety of tasks (e.g., memory tasks, visual tasks, imagery tasks, decision-making) (Alter and Oppenheimer 2009). Subjects tend to have positive metacognitive feelings whenever information is fluently and easily processed, independently of the content.

Metacognitive feelings play a crucial role in the activation of cognitive resources; whereas positive feelings (e.g., familiarity, confidence, satisfaction) signal that cognitive

effort can be reduced, negative feelings (e.g., difficulty, uncertainty) inform a subject that more effort is required to attain a certain goal. For example, in a memory task, if a certain answer sounds familiar, a reasoner will stop looking for the solution; in contrast, if the reasoner feels that the answer is on the tip of her tongue, she will increase her effort.

The importance of metacognitive feelings in regulating cognitive resources has been documented in some studies on dual process reasoning. According to Thompson and colleagues' theory of *metacognitive reasoning*, when a subject faces a reasoning problem, her automatic processing generates two distinct outputs: an initial answer to the problem and an accompanying sense of correctness of the answer—that is, a *feeling of rightness* (FOR) (Thompson, Turner and Pennycook 2011, Thompson and Morsanyi 2012). A FOR is a metacognitive experience that measures how much a subject feels confident about her intuitive response; it is determined by the fluency of information processing—that is, how easily the answer comes to mind. Importantly, a FOR is predictive of the quality and the extension of a subject's reflective engagement in reasoning problems. That means that if the FOR is low (i.e., the subject feels less confident about her intuition), it is more likely that the subject will reflect more extensively and tend to provide a correct answer; in contrast, if the FOR is high (i.e., the subject feels confident), it is less likely that the subject will rely on effortful information processing to check her answer.[6]

4.2 Intuitive strength as confidence

When a subject faces a moral problem, she can rely on different cognitive mechanisms to quickly generate a moral response. However, in parallel with this automatic recognition of the relevant stimuli, the subject also monitors how *fluently* the information is processed. The level of cognitive fluency, in turn, affects her confidence in the resultant representation. That is, the more fluently and easily a certain stimulus is processed, the greater the confidence in the automatic response.

To the extent that there are various degrees of fluency, confidence in moral representations varies along a continuum. A given moral representation is experienced by the subject as *intuition* once it exceeds a certain threshold of confidence. Therefore, what

[6] Complementary to Thompson's studies, Gangemi and colleagues (2015) investigated whether subjects experience a "feeling of error" (FOE) in addition to biased intuitions.

differentiates moral intuition from weak automatic cognition is a substantial difference in confidence (figure 1).

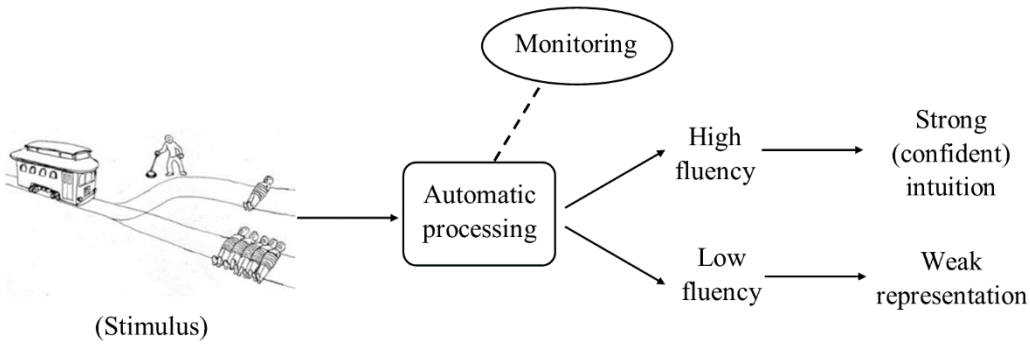

Figure 1. If the cognitive processing of a certain stimulus is fluent, the subject will have a strong, confident moral representation—that is, a moral intuition. Conversely, if the cognitive processing is not fluent, the subject will have a weak, unconfident representation.

From this reconstruction we can see that intuitive strength is nothing but the level of confidence in a moral intuition—that is, a metacognitive appraisal of the cognitive fluency with which the intuition is generated. Accordingly, moral intuitions are as strong as the cognitive processing that generates them is fluent.

The phenomenology of confidence is not easy to discern. In a "calm" reasoning context, confidence is typically experienced as a FOR or likelihood that accompanies a certain plausible response. However, the FOR tends to be overwhelmed by moral emotions in intuitions elicited by vivid and emotional scenarios; in these cases, the metacognitive feeling lurks in the background. Nevertheless, even emotional intuitions are typically generated with great fluency; the speed and easiness of the information processing generated by a moral emotion lead the subject to confidently praise or blame an action. For example, when a subject observes two hoodlums torturing a cat, the feeling of indignation favors quick information processing that leads to the confident intuition that the action is morally wrong. Therefore, regardless of whether it is due to a fast conceptual inference or a moral emotion, cognitive fluency favors strong moral intuitions, characterized by a high level of confidence.

Importantly, the perceived confidence differentiates strong intuitions from weak mental states, such as guesses or automatic hypotheses. Indeed, the latter are usually accompanied by low or no confidence. Compare, for example, two responses to the trolley dilemma. Daniel has little familiarity with the problem but perfectly understands the dilemma and is sympathetic with the victims; nevertheless, he has no idea of what is the

right thing to do for the bystander. With scarce conviction, he answers that turning the switch is wrong but just because he remembers that his high school professor said that it is wrong. Mary, in contrast, is a convinced Kantian familiar with dilemmatic situations of this kind; thus, she has the strong intuition that killing one person to save five is not permissible. Arguably, what distinguishes Daniel's guess from Mary's strong intuition is the cognitive fluency with which the two answers are generated. Daniel exerts much effort to respond to the moral problem and to recall relevant information; as a result, his thought that the bystander is wrong in turning the switch is characterized by a sense of uncertainty. In contrast, Mary quickly and fluently answers the problem in virtue of her Kantian belief; thus, her intuition is strong.

Not only does the metacognitive account explains the phenomenology of strong intuitions across different contexts and objects of intuition, but it also explains the cognitive function of intuitive strength. Indeed, it is plausible that subjects tend to detect the presence of a credible representation in a stream of thought by the elicited FOR. In other words, by monitoring cognitive fluency, a reasoner can spot the most plausible intuitions among different considered propositions. Consistent with the literature on metacognitive feelings, it is also likely that a reasoner is guided in activating cognitive resources by her level of confidence about her intuitions. The more a subject is confident about her answer, the less inclined she will be to reflect and revise her response; conversely, the less confident she is about her answer, the more inclined she will be to reflect and provide an unstable answer. Such predictions are supported by different lines of evidence.

Some recent studies on moral judgments of sacrificial dilemmas are worth mentioning (Bago and De Neys 2019, Vega, et al. 2020). These works confirm the key hypotheses of the metacognitive account. First, the level of intuitive confidence (FOR) turns out to be predictive of the subjects' rethinking time after the intuitive judgment and the tendency of the subjects to revise the initial judgment. Second, the subjects' confidence about their initial responses is correlated with the fluency with which the judgments come.

The metacognitive account finds also support in some studies on the stability of moral intuitions. Interestingly, Zamzow and Nichols (2009, 373-374) found that confident answers to moral dilemmas are less subject to "order effect" than less confident answers. This means that confident moral intuitions are more stable than unconfident

responses. Similar results were obtained by Wright (2010), which reported that confident intuitions are resistant to change in the order of presentation of some moral dilemmas. The strong link between confidence and stability is also confirmed by a series of studies, in which it is shown that confident replies are not affected by suggesting to the subjects the existence of disagreement among experts on the evaluated moral problems (Wright 2013).

Finally, a series of studies investigating the cognitive style of political extremism provides indirect evidence for the correlation between confidence and strong intuitions (van Prooijen 2021). By definition, extremists have strong and stable intuitions about moral and political issues; indeed, the extremist strongly identifies herself in some ideology and is hardly flexible to the circumstances. Importantly, a large body of evidence shows that extremist attitudes result from a simplistic construal of the world. This latter tends to generate excessive confidence about social and political problems, which explains why extremists have strong and stable moral intuitions.

A typical objection to the idea that we can reduce intuitive strength to confidence is that confidence can hardly be reconciled with the automaticity of intuitions (Bengson 2015, 721). In the case of "recalcitrant" intuition, the objection goes, a subject is confident that the intuition is false; however, the intuition persists despite being overridden by reflection. For example, suppose one reads Haidt's vignette of Julie and Mark and has the intuition that the two siblings are wrong. Then, suppose that, after reflecting on the story, the reader becomes convinced that their behavior is permissible;[7] however, the intuition that they are wrong persists even though the reader confidently believes that the intuition is mistaken. Therefore, subjective confidence appears to be in tension with intuitive strength.

To respond to this objection, it is important to consider the literature on metacognition, which distinguishes between two types of metacognitive appraisal: explicit judgments of confidence and metacognitive feelings. Explicit judgments of confidence are reflective and controlled by the subject, whereas metacognitive feelings, such as the FOR, are generated automatically. Thus, the two types of appraisal can

[7] This happens in Paxton and colleagues' study, in which a significant number of participants revised their negative judgment toward Julie and Mark when encouraged to reflect (Paxton, Ungar and Greene 2012).

conflict. A subject can reflectively assign a low level of credibility to a certain intuition but, at the same time, feel that the intuitive response possesses some likelihood of being true in virtue of the processing fluency.

For example, consider the famous "bat-and-ball" problem (Frederick 2005):

> A bat and a ball together cost $1.10. The bat costs $1 more than the ball. How much does the ball cost?

If one knows that the correct answer is $0.05, one reflectively states with certainty that the answer is $0.05. Nevertheless, given the spontaneity and fluency of the answer "$0.10", it is possible that the reflective metacognitive assessment conflicts with the perceived likelihood of "$0.10" being correct. Such perceived likelihood is an automatic metacognitive feeling resulting from the fluency of the answer. Although one can ignore and override it, the feeling persists as long as the answer is fluently processed once considered.

Similarly, the case of Julie and Mark is designed to dissociate by ad hoc provisos two concepts often linked to each other: incest and biological or psychological harm. The association between incest and moral wrongness is so fluently generated by the majority of subjects that it generates a confident moral intuition. Therefore, although a subject accepts Haidt's ad hoc provisos, the intuition that Julie and Mark are wrong can still be compelling because incest was frequently associated with harm in past experience. These examples illustrate how metacognitive feelings can conflict with reflective evaluations of confidence.

To summarize, according to the metacognitive account, the strength of moral intuitions denotes the level of perceived confidence of a subject about an intuitive representation. This account provides a convincing explanation of the phenomenology of strong intuitions and their cognitive function. The metacognitive account is consistent with the possibility of recalcitrant intuitions to the extent that reflective judgments of confidence can conflict with automatic metacognitive feelings. Therefore, the metacognitive account can in principle explain the three explananda of intuitive strength. As I will argue in the next sections, understanding intuitive strength as confidence helps distinguish it from other features such as emotional intensity and presentational phenomenology.

5. Intuitive strength and moral emotions

In this section, I consider and reject the hypothesis according to which the strength of moral intuitions should be understood as emotional intensity. To discuss this hypothesis, I proceed from a commonly accepted view of moral emotions and a brief review of the evidence on their influence on moral judgment (5.1). Then, I outline the main arguments for the reductionist account of intuitive strength and consider some objections (5.2).

5.1 Introducing moral emotions

The nature of emotion is subject to great discussion in philosophy and psychology. However, there is a substantial convergence among philosophical and experimental theories on the fact that emotions are complex mental episodes constituted by the dynamic integration of different components such as an appraisal of the situation, action preparation, physiological responses, expressive behavior, and subjective feelings (Scherer and Moors 2019, Deonna and Teroni 2012). Within this common ground, there are rival theories of emotion. Here I embrace a *cognitivist* account of emotion, according to which a representation of an emotionally salient object is intrinsic to an emotional episode. However, I will not take a side within the different cognitivist views provided in the literature (e.g., the *attitudinal* or the *perceptual* view).

Particularly relevant for the present purpose is the relationship between emotions and evaluative properties. Emotions are said to be evaluations of particular salient situations. Said otherwise, by emotions, a subject can be aware of the presence of certain evaluative properties (Deonna and Teroni 2012, 40-41). For example, by experiencing fear, a subject can apprehend the *dangerousness* of a situation; by anger, a situation is apprehended as *offensive* or *unfair*; by admiration, a particular act or person is evaluated as *admirable*. This is the case, as I will argue (Ch. 5), because emotions are tightly connected to the subjects' caring and concerns.

Given the intimate link between emotions and evaluative properties, it is unsurprising that the importance of emotions for moral judgment has often been emphasized in the history of philosophy (Aristotle 2004, Hume 2007) and, more recently, by some psychologists and empirically minded philosophers (Haidt 2001, Nichols 2004, Prinz 2007). However, it is necessary to premise that not every emotion is responsive to morally relevant evaluative properties. For instance, emotions such as sadness or amusement do not detect moral properties nor motivate moral beliefs. An emotion is said

moral whenever it promotes or detects conduct that violates or conforms to a moral standard (Prinz 2007, 68). Typically, emotions that do satisfy such conditions are guilt, shame, admiration, anger, disgust, pride, compassion, and gratitude. Moral philosophers and psychologists that stress the influence of emotion on moral judgment refer to this specific class of emotions, rather than affective phenomena in general.

Moral emotions can be divided into two main families: *other-directed* and *self-directed* (Haidt 2003, Prinz 2007). Other-directed moral emotions are typically elicited by other people's violation of or conformity to moral norms. Anger, disgust, indignation, admiration, compassion, and gratitude belong to this class of emotions. By contrast, self-directed moral emotions are triggered by a subject's own violation of or conformity to moral norms. Emotions of this kind are shame, guilt, pride, and dignity.

The influence of moral emotions (both other- and self-directed) on moral judgment is supported by different lines of empirical evidence. Some studies showing the modulation of moral judgment through the elicitation of disgust are considered to provide crucial behavioral data in favor of sentimentalist accounts of moral judgment. Specifically, Wheatley and Haidt (2005) found that hypnotizing some subjects to experience a flash of (morally irrelevant) disgust makes their judgments of moral violation more severe. Schnall and colleagues (2008), instead, employed four different methods to induce disgust: applying a smelly spray in the vicinity of the participants, questioning the subjects on filthy desks, inducing memories of disgusting experiences, and showing a revolting movie clip. Consistent with moral sentimentalism, all these manipulations have affected the severity of moral judgments.[8] Additionally, another study (Schnall, Benton and Harvey 2008) has shown that participants primed with words related to cleanliness and purity tend to make less harsh moral judgments.

The link between affect and moral judgment is also supported by some neuroimaging data. In their influential fMRI study, Greene and his team registered the activation of areas of the brain associated with emotion (such as the dorsolateral prefrontal cortex, DLPC) concurrently with perceived deontological violations in the trolley dilemma (Greene, Sommerville, et al. 2001). Consistent with this result, in more recent times, Decety and Cacioppo (2012) found that the amygdala is activated in the

[8] A similar effect has been obtained by inducing incidental feelings of anger, which produces harsher moral judgments against other people (Seidel and Prinz 2013).

early stages of moral judgment (between 122-180 ms); this suggests that emotion acts as a fundamental antecedent factor to moral cognition by alerting a subject to morally salient aspects of a situation, such as intentional harm (Decety and Cacioppo 2012, 3072). As has been argued (Huebner, Dwyer and Hauser 2008), since neuroimaging data are only *correlational*, they cannot decisively establish a causal connection between emotion and moral judgment. Nonetheless, they constitute a relevant case in favor of moral sentimentalism.

Investigations of the moral capacities of people affected by psychopathy have been considered a decisive way to test the constitutive role that emotion plays in moral judgment. According to James Blair's influential theory, a characteristic of psychopathy is dysfunction of the amygdala, which affects psychopaths' capacity to experience moral emotions such as guilt, empathy, or concern for others (Blair, Mitchell and Blair 2005). This emotional impairment would explain the impulsive and antisocial behavior shown by some psychopaths. In support of this hypothesis, Blair has shown that subjects affected by psychopathy tend to be insensitive to the distinction between the violation of moral norms (e.g., hurting another person) and conventional rules (e.g., wearing socks of different colors) (Blair 1995, Blair, Mitchell and Blair 2005, 57-59).[9]

Finally, moral sentimentalism is supported by some data from neuropsychology. Some studies have observed that people with a damaged ventromedial prefrontal cortex (VMPFC) judge moral violations without actual harm more permissible than healthy subjects (Koenigs, et al. 2007, Young, et al. 2010). Given the known emotional impairment of vmMPFC patients, this evidence might suggest that affect plays a crucial role in the detection of moral violations.

Therefore, as this brief review suggests, the influence of emotion on moral judgments is undoubtedly well documented. In what follows, I do not deny that moral emotions are necessary for the formation of moral judgment and intuition. Nor do I discuss the empirical claim that emotions strongly affect the *content* of moral intuitions. Rather, the specific purpose of this section is to consider how moral emotions are relevant for the strength of intuitions.

---

[9] However, this hypothesis has been challenged by a more recent counterevidence (Aharoni, Sinnott-Armstrong and Kiehl 2014).

5.2 Disentangling emotional intensity and confidence

It is critical to distinguish moral emotions, such as anger, guilt, and compassion, from metacognitive feelings. Moral emotions are *first order* feelings, which are directed toward morally salient facts: their object is constituted by violation or promotion of moral norms. In contrast, feelings of confidence, familiarity, or uncertainty are *second order* feelings directed toward first order cognitions, which constitute their object. Metacognitive feelings can have as objects different kinds of moral evaluations, such as a judgment or a moral emotion itself.

Although conceptually distinguishable, the intensity of moral emotions and the confidence of moral intuitions are tightly related. Converging evidence highlights that the more a scenario elicits a moral emotion, the more the subjects have confident intuitions. For instance, people tend to more confidently condemn a surgeon who saves five sick people by harvesting organs from a healthy person than a bystander who kills a man to save five by pulling a lever on a track (Zamzow and Nichols 2009, 373-374); probably, this is because the transplant triggers more sympathetic concern toward the victim than the impersonal sacrifice does.

Given the correlation between moral emotions and intuitive confidence, some authors seem to suggest that intuitive strength can be reduced to the emotional intensity of a moral representation (Kauppinen 2013, Railton 2014). This is certainly an attractive move since there are several points of connection between moral emotions and strong intuitions. First, the phenomenology of moral emotions resembles one of strong moral intuitions. Indeed, emotions such as guilt, shame, and disgust are experienced as compelling and direct attention toward the emotional object. Second, concerning the cognitive function of intuitive strength, the reductionist account can provide a natural explanation, considering the role of emotion in modulating subjects' attention (Vuilleumier 2005, Brady 2013, 16-25). Plausibly, emotional representations guide the formation of evaluative beliefs by capturing the subject's attention. Moreover, emotional experiences can consume the attention if challenged, inducing the subject to look for reasons. For example, an agent who thinks that Julie and Mark's behavior is right but still feels disgusted by it will spend time and pay attention to understand why the feeling persists despite the contrary considerations. For these reasons, moral intuitions accompanied by high emotional intensity are likely strong and stable representations, while intuitions with low emotional intensity tend to be weak and unstable.

Reductionism also explains how intuitive strength can be generated automatically. Emotions are generated by automatic processes: one cannot control the arousal of a moral emotion, nor does one always have clear access to the processing that leads to the emotion. For these reasons, emotions can be *recalcitrant*—that is, in tension with reflective beliefs (Benbaji 2013, Brady 2013). For instance, a subject can believe that donating money to charity is not obligatory but still may be inclined to believe that it is right because of felt compassion for poor people. The recalcitrance of moral emotions can explain the disruptive nature of strong intuitions, which can go against already-settled moral beliefs.

Another significant argument for reductionism appeals to the alleged *motivational force* of moral intuitions. Compared with other types of intuition, moral intuitions seem to play a major role in guiding conduct. Indeed, moral intuitions are said to be intrinsically motivating (Kauppinen 2013, 366): experiencing a strong moral intuition disposes the subject to act in conformity with some moral representation. For example, the subject who observes two hoodlums torturing a cat and has the strong intuition that it is wrong will be disposed to intervene to stop the torture. The reductionist account provides a straightforward explanation of the action-guiding role of strength of moral intuitions. Strong moral intuitions, reductionists point out, motivate a subject to act because they are emotionally charged. Experiencing a relevant moral emotion can produce in a subject a certain "action tendency" toward the emotional object (Frijda 2007, 33-34). For instance, the felt anger or indignation about the torture inclines the subject to intervene.

In sum, these are the main reasons why the strength of moral intuitions is often understood as reducible to emotional intensity. However, I disagree with this reductionist view.

It is important to consider that the influence of moral emotions on intuition varies greatly by type of moral scenario (Ugazio, Lamm and Singer 2012). For example, moral violations involving *personal force* tend to generate more emotional intuitions (Greene 2014). Emotional moral intuitions tend to be, but not necessarily are, stronger than unemotional intuitions. Some interesting experiments conducted by Nichols and Mallon (2006) show that clear violations of moral rules with minimal emotional force generate strong intuitions comparable to the ones responsive to personal-force violations.

The hypothesis that emotional intensity and intuitive strength can diverge is suggested by the ordinary experience of strong and clear moral intuitions with low

emotional force. This happens, for example, when one considers general moral propositions, such as the proposition that benevolence is a virtue or that freedom and happiness are fundamental values. In support of this hypothesis, it is worth mentioning the influential neuroimaging study conducted by Kahane and colleagues (2012), which investigated the neural base of "intuitive judgments", defined as immediate and unreflective responses to moral problems. In contrast with Greene and colleagues (2001), Kahane and colleagues did not observe that activation of brain areas associated with emotions (amygdala and DLPC) was correlated with intuitive judgments, regardless of the judgments' content (deontological or utilitarian). Rather, clear and strong intuitions turned out to be correlated with the visual and left premotor cortex.

As regards the supposed motivational force of strong intuitions, I argue that the strength of a moral intuition and its motivational power should be considered as distinct properties. Intuitive strength concerns how likely and credible a certain moral representation is and performs the function of guiding the formation and justification of beliefs. In contrast, the motivational power of a moral intuition, which is probably correlated with its emotional intensity, is the disposition of the intuition to generate actions consistent with the moral representation. The motivational power of intuitions serves to adapt the conduct of an agent to the specific demands of particular situations in light of the agent's values and concerns. Given these definitions, there is no a priori reason to identify the two distinct features of moral intuition with each other. Therefore, as long as convincing evidence for their correlation is not provided, one should consider motivational power and intuitive strength as separate.

Intuitive strength and motivational power are dissociable also because they depend on distinct individual traits. How a subject experiences strong intuitions depends mostly on her *cognitive style* or *thinking disposition*—that is, how much she tends to trust her immediate responses and hunches, how confident she feels about her moral beliefs, how much she is disposed to reflect on a moral problem. By contrast, how much an agent is motivated by her intuitions depends on her sentiments of concern for moral standards (Ch. 5). Importantly, this hypothesis finds support in some empirical evidence. In five studies, Ward and King (2018) find that harsher condemnation of provoked harm is significantly correlated with Faith in Intuition (FI), which measures the tendency of a subject to rely on intuitions in reasoning problems (Epstein, Pacini, et al. 1996). The experimenters compared the influence of FI on the participants' moral judgment with

other traits, among them religiosity, emotional intensity, disgust sensitivity, and emotional reappraisal. None of these variables alone predict the severity of people's judgments. Thus, the strength of intuitions about moral violations seems to depend on the subjects' cognitive style, rather than their sensitivity to emotions or their ideology.

The hypothesis that treats intuitive strength and emotional intensity as separate properties dependent on different functions and traits is also attractive from a theoretical point of view. If the strength of moral intuition is distinguishable from motivational force, one can reject the reductionist account of strength while conceding that moral intuitions are intrinsically motivating and more emotional compared with other types of intuitions.

If intuitive strength cannot be reduced to emotional intensity, why do emotional intuitions tend to be stronger? A plausible hypothesis is that moral emotions, as great cognitive facilitators, tend to increase the processing fluency of an intuition. Moral emotions activate immediately after the perception of the stimulus (Decety and Cacioppo 2012) and direct attention toward morally salient details (e.g., the intentional killing of a man), thus favoring rapid information processing. To the extent that emotional processing is rapid and fluent, the resultant representations tend to be more confident than the average.

However, moral emotions are not the only determinant of cognitive fluency. An important documented aspect is the *paradigmaticity* of a moral scenario, that is the degree to which it presents a case with a clear exemplification of a moral concept (Wright 2010, 500). Exemplar cases of violations or promotions of moral norms trigger an immediate unconscious inference from the perceived stimulus to a moral concept. Thus, the fluency of the inference tends to favor confident intuitions.

In addition to the paradigmaticity, *familiarity* can affect the fluency of a moral judgment: repeated experience of a moral problem probably contributes to making subjects feel more confident about their moral representations. Importantly, the familiarity with a moral problem can decrease the emotional intensity; for example, after reading 500 times the transplant dilemma, the sympathy for the victim might be less intense than the first time. However, the decrease of the emotional intensity with prolonged experience does not necessarily entail a reduction of the confidence of moral intuition; on the opposite, sometimes more familiarity increases confidence. This suggests, once again, that intuitive confidence and emotional intensity can diverge.

6. Intuitive strength and the perceptual analogy

In recent years, some authors (Bengson 2015, Chudnoff 2013) have defended a sophisticated "quasi-perceptual" theory of intuition, according to which intuitions are intellectual perceptions. This influential account emphasizes some strong analogies between intuitions and sensory perceptions. More specifically, according to quasi-perceptualism, both intuitions and sensory perceptions are *presentational states*.[10] Like sensory perceptions, intuitions present the world as being in a certain way. In other words, intuitions provide the impression that things stand in the way they are represented; for example, if one has the intuition that killing babies is wrong, one has the vivid impression that killing babies is wrong.

Advocates of quasi-perceptualism relate the fact that intuitions are strong, stable, and compelling mental states with the claim that they are perceptual-like presentational states. More precisely, quasi-perceptualism seems to argue that intuitive strength *just is* presentational phenomenology or, according to a weaker interpretation, that the strength of intuitions *supports* the view that intuitions are intellectual perceptions. I disagree. The strength of intuitions does not denote presentational phenomenology; nor does intuitive strength entail that intuitive representations have presentational phenomenology. Intuitions are not sui generis intellectual perceptions, but just automatic cognitions more confident than the average, or so I argue in this section.

This is not the place to discuss the quasi-perceptual theory in detail. I just consider the most influential argument for quasi-perceptualism, namely the analogy between visual illusions and recalcitrant intuitions. A popular example of a visual illusion is the Müller-Lyer illusion, in which two parallel straight lines of the same length are shown, but the top line looks longer. Importantly, the illusion persists even when one knows that it is an illusion (figure 2). Similarly, in the case of recalcitrant intuition, a subject has the intuition that *p*, although she knows that *p* is false. This can be the case, for example, when one knows the correct answer to the bat-and-ball problem is $0.05 but still has the intuition that the answer is $0.10.

---

[10] Quasi-perceptualism is closely related to the view of moral intuition as intellectual seeming (Huemer 2005) since a presentational state is a specific type of seeming (Bengson 2015, 729-730).

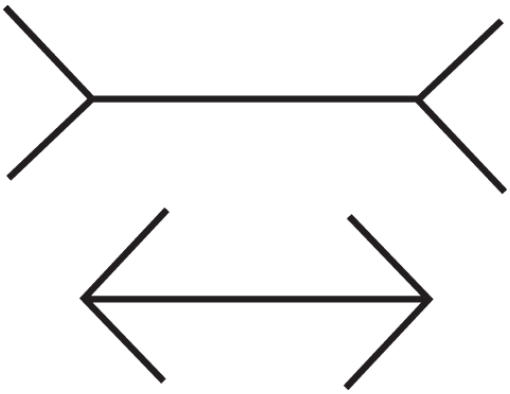

Figure 2. The Müller-Lyer illusion.

According to advocates of quasi-perceptualism, the analogy between visual illusions and recalcitrant intuitions suggests that intuitions are like perceptual states in presenting certain contents as veridical, independently of reflective beliefs. However, there are some important dissimilarities between visual and cognitive illusions.

First, compared with sensory perceptions, intuitions are by large more sensitive to cultural and individual differences. Visual illusions are experienced across a wide variety of cultures and individual traits. I am not saying that sensory perceptions are cognitively impenetrable;[11] however, they are certainly less malleable than intuitions, which are extremely variable according to habits, social norms, and individual traits (Zamzow and Nichols 2009, Sauer 2017).

The second dissimilarity, probably related to the first one, is that the strength of intuitions tends to decrease slightly when the intuition is overridden or challenged by reflection.[12] In contrast, the vividness and veridicality of an illusory perception remain unchanged. After learning that in figure 2 the two lines have the same length, the impression that the upper line is longer remains as strong as before learning the illusion. This means that intuitions are more sensitive to reflective considerations than perceptual states are.

Third, and finally, the susceptibility of intuitions to habits and reflective considerations reveals a substantial disproportion between visual illusions and cognitive

[11] Some evidence highlights that susceptibility to geometrical illusions can vary by culture (Segall, Campbell and Hesrskovits 1963).

[12] This phenomenon was reported in Wright (2013), which reports a slight decrease in strength of intuitions after inducing instability.

biases. While there are just a few visual illusions and they are easily recognizable, cognitive biases are manifold and their origin is much more disputed.

The metacognitive account of intuitive strength offers straightforward explanations of such dissimilarities. Since the confidence of intuition is determined by processing fluency, the subject plays an active part in it, although she cannot directly control it. How fluent cognitive processing is depends on different factors, such as the subject's beliefs, past experience, and cognitive style. That explains why the resultant confidence is susceptible to biases and to cultural and individual traits. Moreover, contrasting reflective considerations and accidental circumstances may interfere in the information processing, making it less fluent. More precisely, if a subject is aware that a certain reasoning problem (e.g., the bat-and-ball problem) might be tricky, reading the problem may trigger conflicting solutions ($0.10 versus $0.05). The more intuitive answer may prevail, but the detection of the conflict affects the processing fluency and the confidence of the final response (De Neys 2018).

Note that the metacognitive account captures the most important similarities between intuitions and perceptions highlighted by quasi-perceptualism. I mean, for example, the automatic and potentially disruptive nature of strong intuitions and their role in the formation of beliefs (Bengson 2015). However, unlike quasi-perceptualism, the metacognitive account stresses the active role of the subject in generating intuitive strength. This aspect, as I will argue (Ch. 3), has significant implications on the question of the reliability of moral intuition.

7. Conclusion

Moral intuitions are automatic mental states characterized by a certain level of strength. In this chapter, I have offered a metacognitive account according to which the strength of moral intuitions denotes the level of subjective confidence about a moral representation. Intuitive confidence is determined by the fluency with which the information is processed.

The explanation I have provided demystifies moral intuition by understanding it an automatic cognition more confident than the average, rather than a sui generis mental state. The metacognitive account explains the phenomenology of strong intuitions and their cognitive function.

I have argued that intuitive strength cannot be reduced to emotional intensity, to the extent that moral emotions are not the only determinant of processing fluency. Finally, I have contended, in contrast with quasi-perceptualist accounts of moral intuition, that intuitive strength does not entail that intuitions are presentational states.

Chapter 2

Dual Process Reflective Equilibrium[13]

1. Introduction

Dual process theories of the mind (Kahneman 2011, Evans and Stanovich 2013) have been very influential in the development of moral psychology in recent decades. In particular, the dual process framework helps explain the coexistence of two distinct pathways to moral judgment: intuition and reflection. However, it is still unclear from the literature how the two processes interact in moral reasoning. Competing hypotheses address this concern.

Greene (2008, 2014) and other authors (Suter and Hertwig 2011, Conway and Gawronski 2013) seem to understand moral intuition and reflection as conflicting cognitions: intuitive thinking would elicit heuristic and deontological responses, whereas reflection would favor balanced and utilitarian judgments. This view fits a "default-interventionist model" of reasoning (Kahneman 2011, Evans 2019). The key assumption of the default-interventionist view is that an intuitive heuristic response is generated from a reasoning problem by default and afterwards, reflective processes intervene to check the heuristic response if the problem is complex and unfamiliar. Accordingly, deontological judgments would derive from uncorrected heuristic intuitions, while consequentialist judgments would result from reflective interventions that override the heuristic responses. This model plainly explains Greene's core claims and is consistent with some theories that understand moral intuitions as heuristics (Sunstein 2005, Gigerenzer 2008, Sinnott-Armstrong, Young and Cushman 2010).

This chapter reviews the evidence for the default-interventionist view of moral reasoning and proposes an alternative account of how intuition and reflection interact in the moral domain. I will show that the evidence for the default interventionist view is inconclusive and has been challenged by a growing amount of counterevidence in recent

[13] This chapter is a slightly revised and extended version of the work published in Cecchini (2021).

years (Bialek and De Neys 2017, Gürcay and Baron 2017, Bago and De Neys 2019, Rosas and Aguilar-Pardo 2019, Vega, et al. 2020). In addition to the recent empirical findings, also a close examination of the literature on psychopaths favors an interdependent rather than conflicting view of the two types of information processing (Maiese 2014). In this view, which I call *dual process reflective equilibrium,* intuition and reflection cooperate in moral reasoning to reach a reflective goal, which is supposedly normative justification. In sum, on the one hand, the scope of moral intuitions extends to selecting relevant information and calling for reflection whenever a problem presents conflicting aspects; on the other hand, the purpose of moral reflection is to rationalize pre-reflective intuitions to provide articulated and accessible reasons.

The chapter is structured as follows. In Section 2, I will outline the default-interventionist view of moral reasoning (MDI, for brevity); I will show that MDI finds support in two distinct hypotheses: (i) type 1 processes tend to elicit heuristic and deontological responses, and (ii) type 2 processes tend to correct heuristic responses by utilitarian judgments. In the following section, I will discuss the empirical evidence for MDI. Firstly, I will consider empirical studies on moral reasoning that involve manipulations of deliberation time and cognitive resources (3.1). Then, I will consider some studies on psychopaths that could provide evidence for MDI (3.2). In Section 4, I will show some counterevidence that challenges MDI's core hypotheses. Section 5 outlines my account of moral reasoning as dual process reflective equilibrium and describes the functions of type 1 and type 2 processing in moral reasoning. Finally, in Section 6, I will discuss some hypotheses on why moral reasoners tend to rationalize their intuitions.

2. Dual process morality and the default interventionist view

As argued (Ch. 1), one of the main findings in the field of moral psychology in the last decades is that emotional processes play a key role in moral judgment (Damasio 1994, Haidt 2001). Greene and his team have perfected this view by pointing out that two distinct pathways to moral judgment are possible: emotional processes (i.e., moral intuitions), which lead to characteristically deontological judgments, and reflective processes, which lead to characteristically consequentialist judgments (Greene, Sommerville, et al. 2001, Greene 2008, Greene 2014). A relevant challenge left by Greene's dual process theory of moral judgment is to understand how intuition and

reflection interact. If one observes moral behavior, one notices that moral reasoners intelligently know when and how to switch between intuitive and reflective thinking, according to the context and their abilities. However, in what circumstances? By what capacities?

Thus far, in the present research, I have not established a neat distinction between moral intuition and judgment. Here, conventionally, I assume that intuition is a moral representation deriving from automatic and often emotional processes. In contrast, I will refer to moral judgment as the conclusion of moral reasoning. Accordingly, in the present chapter, moral judgment is not identical to moral intuition but involves some interplay between intuition and reflection. The mode of interaction between the two processes, i.e., how moral judgment is generated, is the focus of the present discussion.

"Default-interventionist" model of reasoning (Kahneman 2011; Evans 2019) is the most influential account of the interplay between intuition and reflection. The core claim of this theory is that reasoning is constituted by a serial interplay of "type 1" and "type 2" processes.[14] Type 1 processing is characterized as a fast and automatic cognition that does not require working memory; by contrast, type 2 processing is a controlled and slower cognition that engages working memory (Evans and Stanovich 2013; Evans 2019). In summary, according to the default-interventionist view of reasoning, when a subject faces a reasoning problem, she automatically activates type 1 processing, which provides a heuristic response by default; subsequently, according to her motivations and cognitive abilities, the subject can activate a form of controlled type 2 processing and, possibly, correct her heuristic response. The key assumption of this view is that "reasoners are conceived as cognitive misers who try to minimize cognitive effort" (De Neys 2018, 48). Since analytic engagement is cognitively effortful, the subjects tend to avoid type 2 processing interventions and instead rely on type 1 responses as much as possible. However, this is problematic when the reasoning problem is complex and unfamiliar: in these cases, as the heuristic and bias literature shows (Kahneman, Slovic

[14] Some theories assume that the two processes correspond to two different cognitive systems (dual-system theories). This hypothesis is stronger since it presupposes that type 1 and type 2 processes are located in two different areas of the brain with different evolutionary histories (Evans and Stanovich 2013).

and Tversky 1982, Kahneman 2011), heuristic responses are inadequate, and consequently, a missed type 2 intervention can lead to biased answers.

If faithfully applied to the moral domain (figure 3), a default-interventionist view of reasoning explains Greene's empirical findings quite plainly: emotional deontological judgments derive from uncorrected type 1 heuristic responses, whereas consequentialist judgments result from type 2 interventions that override heuristic responses (figure 3).

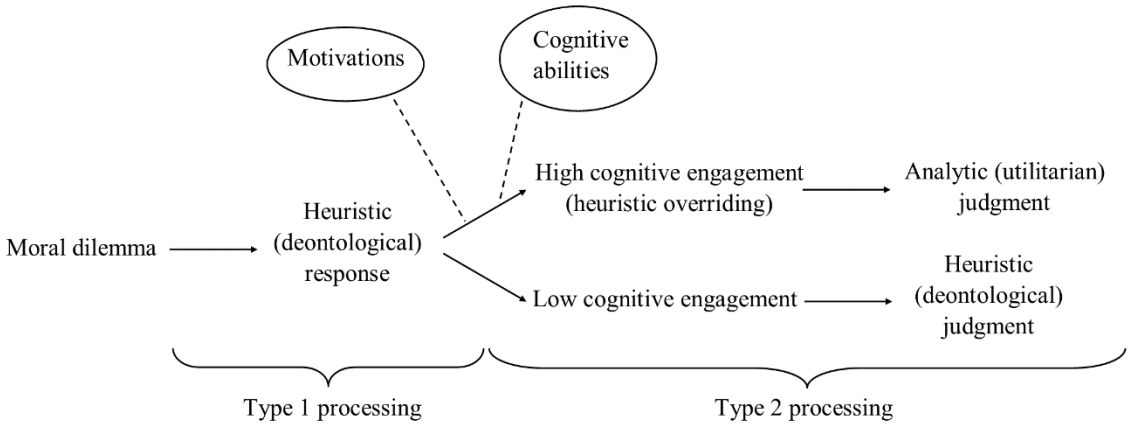

Figure 3. The default-interventionist model of moral reasoning.

Moral default-interventionism (MDI) is grounded in two distinguishable hypotheses. The first hypothesis is that type 1 processes in moral reasoning work through *moral heuristics*, whose purpose is to provide a quick and cognitively effortless answer to a moral problem.

Heuristics are mental shortcuts or rules of thumb that unconsciously substitute a relatively inaccessible attribute ("target attribute") with a more easily accessible attribute ("heuristic attribute") (Sunstein 2005; Gigerenzer 2008; Sinnott-Armstrong et al. 2010). In the moral domain, the target attribute is a moral term (for example, "being wrong") that is usually supported by a set of contributory reasons; the heuristic attribute is a good or wrong-making fact of the situation that is usually associated with that moral value (for example, the intentional killing of a man). When a moral heuristic activates, the subject unconsciously substitutes the moral term with the heuristic attribute. This process makes moral decisions easier since a decision that requires the balancing of different competing reasons is made by considering only one reason. Indeed, heuristics are "fast and frugal": they are designed to provide a quick solution by making only a small part of the information salient and ignoring available counterevidence (Gigerenzer 2008, 4). This yields some cognitive gain since taking into account all the relevant information takes

much more time and cognitive effort. However, the gains in efficiency may entail a loss in the accuracy of the answer provided: if a problem is complex, a correct answer requires a balancing of different competing relevant facts, and one-reason decisions might be inadequate. Finally, heuristics are *adaptive* insofar as they are malleable by the social environment in which they are developed (Gigerenzer 2008, 5). This means that heuristics tend to provide correct responses when they deal with familiar problems; nonetheless, when they are not shaped by trial-and-error experience,[15] heuristics offer inadequate answers.

Therefore, moral heuristics might be useful adaptive tools for moral judgment, but they go beyond the reasoning sphere that concerns complex and unfamiliar moral problems. For this kind of problem, heuristic shortcuts likely lead the individual astray, suggesting inaccurate moral judgments.[16] On this basis, the proponents of MDI theorize (and this is their second core hypothesis) that the function of reflection in moral reasoning is to correct type 1 heuristic responses. In unfamiliar moral problems, competent moral reasoners manage to activate type 2 processes and override their heuristic response by cognitive engagement. By contrast, incompetent reasoners, which cannot accomplish this task, provide heuristics-based judgments. This means that the goal of moral reasoning is mainly corrective: reasoning problems require a type 2 intervention to override a former intuitive emotional response. This view coheres with the idea that type 1 responses are heuristic because, to the extent that they favor inadequate answers to unfamiliar environments, the most adaptive choice that a reasoner can do is correct them by type 2 processes.

The main consequence of MDI is that intuition and reflection are considered *conflicting* in unfamiliar dilemmas: intuition favors heuristic and characteristically deontological responses; by contrast, reflection favors balanced and characteristically utilitarian judgments. Note that, even in deontological judgments, a minimum type 2 intervention occurs insofar as the deontic reasoner manages to formulate a moral

---

[15] This kind of learning may come from genetic transmission, cultural transmission or personal experience. According to Greene (2014, 714) these are the only sources by which moral automatic processes, understood as heuristics, can be informed.

[16] As Greene argues, "it would be a cognitive miracle if we had reliably good moral instincts about unfamiliar moral problems" (Greene 2014, 715).

judgment. However, the relevant point is that in deontological judgments, the degree of cognitive engagement is not high enough to override the heuristic response.

In the next sections, I will review the evidence for MDI.[17] I will discuss first the empirical findings in favor of MDI; then, I will consider some recent counterevidence.

3. Evidence for MDI

Psychologists employed different kinds of methodologies to investigate the dual process nature of moral reasoning. Two distinct lines of evidence may support MDI. Some studies manipulated the subjects' deliberation time and cognitive resources to induce intuitive or reflective thinking (3.1). Other relevant studies tested the utilitarian nature of psychopaths (3.2).

3.1 Studies on moral reasoning

A "sacrificial dilemma" asks whether a moral norm can be legitimately violated to obtain a greater good. For instance, in the footbridge or "fat man" version of the trolley dilemma (Thomson 1985, 1409), a bystander faces two incompatible options: pushing a fat man off a footbridge to stop a trolley that is headed directly towards five men, or not doing anything and letting so five men die. The footbridge version of the trolley dilemma, unlike the standard version, has the effect to make the moral violation more emblematically salient, to the extent that the bystander has to kill the man by using "personal force" (Greene, Cushman, et al. 2009, Greene 2014). So designed, *personal* sacrificial dilemmas polarize utilitarian and deontological reasons and, thereby, they can be involved as reliable tools to investigate the different inclinations that contribute to moral judgment.

As I argued, the core tenet of MDI is that utilitarian judgments require a cognitively effortful reflection to override deontological heuristic responses. On this basis, MDI predicts that interfering with the subjects' availability to reason would favor heuristic deontological responses to sacrificial dilemmas. By contrast, allowing time and

[17] My review will focus on the heuristic and deontological nature of intuition, as well as the utilitarian and corrective nature of reasoning. This chapter does not question Greene's fundamental hypothesis that moral intuitions are emotional.

cognitive resources to deliberate would favor utilitarian judgments. This hypothesis is clearly stated by Greene (2008):

> According to the view I have sketched, people tend to have emotional responses to personal moral violations, responses that incline them to judge against performing those actions. That means that someone who judges a personal moral violation to be *appropriate* (e.g., someone who says it's okay to push the man off the bridge in the *footbridge* case) will most likely have to override an emotional response in order to do it. This overriding process will take time, and thus we would expect that "yes" answers will take longer than "no" answers in response to personal moral dilemmas like the *footbridge* case. (44).

In order to test Greene's corrective hypothesis, experimenters have involved two distinct strategies: manipulating the available time to judge for the subjects and increasing cognitive load of the moral judgments by imposing a distracting task. Limited time or the presence of a distracting task may compromise the possibility to reflect and to override an intuitive response. Thus, through these manipulations, subjects are induced to rely on their intuitive or reflective thinking.

Greene and colleagues (2008) by firsts collected confirmations for MDI. The experimenters presented some moral dilemmas to the participants and instructed a group of them to perform a distracting cognitive task while deliberating;[18] the experimenters observed that, under the cognitive load condition, the response time for utilitarian judgments significantly increased, while no influence was noticed in the response time of deontological reasoners. However, this evidence does not provide strong support for MDI, insofar as the cognitive load condition did not have the effect to decrease utilitarian judgments, but only to extend their response time. However, by replicating similar conditions, Conway and Gawronski (2013) did obtain the predicted decrease of utilitarian responses and observed that deontological judgments were not affected by the cognitive load. Similar effects have also been found by Trémolière and colleagues (2012) but in conditions of stronger cognitive load: in order to knock out people's cognitive resources, the experimenters induced them to think about death. Suter and Hertwig (2011) adopted

[18] The task consisted of detecting a number "5" within a stream of numbers scrolling across the screen.

a different methodology. They set time-limits in deliberation to force the subjects to provide intuitive judgments. Importantly, the experimenters found that when reflection is encouraged and more time is allowed for deliberation, people tend to give more utilitarian judgments, whereas the number of deontological responses increases under time pressure. These results are consistent with the view that consequentialist judgments require slow and cognitively effortful reflection to correct intuitive deontological responses.

The evidence mentioned so far turned out to be inconclusive. An empirical study conducted by Trémolière and Bonnefon (2014) shows that the effect of cognitive load on utilitarian judgments disappears once the number of lives that one can save increases (100 or 500). On this basis, Trémolière and Bonnefon (2014) advance the hypothesis that the more lives one can save, the less demanding is to endorse a sacrifice. This suggests, consistent with Kahane et al. (2012), that utilitarian responses can be elicited automatically if they are not so "counterintuitive" to require reflection. But the most relevant limit of the aforementioned empirical studies is that they do not provide direct evidence that reflective utilitarian judgments truly result from the overriding of a deontological intuition. Nor do they provide direct evidence that deontological intuitions are insensitive to utilitarian reasons. As I argued, these predictions are crucial for MDI. Yet, the methodologies involved so far do not seem to be sufficiently fine-grained to test the core hypotheses of MDI.

3.2 Are psychopaths more utilitarian?

As Blair and colleagues define it, "psychopathy is a disorder that consists of multiple components ranging on the emotional, interpersonal, and behavioral spectrum" (Blair et al. 2005, 7). The psychopath individual tend to be characterized by impulsivity, conduct problems, and by a callous and unemotional interpersonal style (Blair et al. 2005, 8). However, what mainly distinguishes a psychopath from a non-psychopath individual is not antisocial behavior per se but the emotional impairment. Importantly, the emotional dysfunction that psychopaths manifest alters their capacity to experience moral emotions, such as guilt, empathy, or concerns for the others.

Empirical studies with psychopaths might be a relevant source of evidence to understand the nature of moral reasoning. To the extent that psychopath individuals have an important emotional dysfunction, investigating how psychopaths reason about moral problems could reveal the role of type 1 affective processing in moral reasoning.

Quite paradoxically, MDI hypothesizes that, in a sense, psychopaths provide more rational responses before sacrificial dilemmas than healthy individuals (Young, Koenigs, et al. 2012). Insofar as their capacity to have emotional intuitions is impaired, psychopaths would consider violations of moral rights less seriously than non-psychopaths; consequently, psychopaths would need less cognitive effort to override a deontological intuition and provide a utilitarian judgment. Consistent with this hypothesis, Koenigs and colleagues found increased utilitarian responses before sacrificial dilemmas (personal and impersonal) in psychopaths (Koenigs, Kruepke, et al. 2012) and subjects with damaged ventromedial prefrontal cortex (VMPFC, an area of the brain associated with emotions) (Koenigs, Young, et al. 2007). However, these results were contradicted by other studies (Glenn, et al. 2009, Cima, Tonnaer and Hauser 2010) that found that psychopaths' moral judgments are in line with non-psychopaths. Young and colleagues (2012) investigated psychopaths' sensitivity to harm and found mixed results: compared with healthy individuals, psychopaths find accidental harms more permissible, but their moral judgments towards attempted and intentional harms do not significantly differ from non-psychopaths.

The most problematic aspect concerning the evidence from psychopathy is that the capacity of psychopaths to perform genuine moral reasoning is highly disputable. As Kennett (2006) argues, psychopath individuals manifest relevant inabilities in performing practical reasoning; for instance, they have difficulties in assessing the consequences of their actions, detecting conflicts between their desires, and choosing the appropriate means for their ends. This is consistent with a body of evidence highlighting that psychopaths have serious cognitive and attentional deficits that impede them to consider alternative actions and long-term goals (Hamilton, Racer and Newman 2015).

Some authors (Kennett and Fine 2008, Damm 2010, Sauer 2017, 185-191) emphasize that psychopaths tend to commit several inconsistencies in moral reasoning such as, for example:

> When asked if he experienced remorse over a murder he'd committed, one young inmate told us, "Yeah, sure, I feel remorse." Pressed further, he said that he didn't "feel bad inside about it." (Hare 1993, 41)

> When asked if he had ever committed a violent offense, a man serving time for theft answered, "No, but I once had to kill someone." (125)

These statements reveal that psychopaths might not possess the capacities to deeply understand moral concepts. Due to their lack of empathy, it is possible that they do not assign the proper meaning to evaluative terms, such as "remorse" or "violent offence". Likely, psychopath individuals fail to grasp the action-guiding nature of the moral concepts they employ. Perhaps, this deficiency does not compromise their performance in providing "yes or no" answers to moral dilemmas, insofar as psychopaths may superficially understand certain moral rules and coldly apply them to specific cases. However, psychopaths' cognitive and emotional deficits could affect their ability to articulate consistent reasons (after all, this is what real-life moral reasoning requires). Therefore, psychopaths' inability to perform practical reasoning, together with their incapacity to assign proper meanings to moral concepts, do affect their competence in reasoning about moral problems. This suggests that utilitarian reasoning in psychopath individuals may have only the guise of genuine moral reasoning.

According to Maiese (2014), the literature on psychopathy favors an integrated rather than conflicting view of moral intuition and reflection. Possibly, psychopaths' troubles with moral reasoning are due exactly to their incapacities to have correct moral intuitions. As mentioned, psychopaths manifest impulsive and disinhibited behavior: in virtue of their impaired attentional capacity, psychopath individuals fail to pause and reflect about the maladaptive nature of their actions; they cannot catch relevant stimuli from the context that in healthy people would lead to response evaluation and self-regulation (Hamilton, Racer and Newman 2015, 773). This suggests that type 1 affective processing may have a crucial role in regulating moral deliberation. Specifically, moral intuitions could be indispensable to detect possible situational cues that guide a subject to inhibit impulsive behavior and provide a more reflective response. Therefore, according to this reading, studies on psychopathy highlight how moral reflection *fundamentally depends* on intuition.

In sum, on a close examination, the literature on psychopathy does not provide strong support for MDI. Empirical studies which tested the utilitarian tendency of psychopaths provide mixed results. Moreover, the capacity of psychopaths to perform moral reasoning is questionable, and that favors an interdependent view of moral intuition and reflection.

4. Counterevidence

MDI is challenged by an increasing amount of counterevidence. In recent years, some empirical studies have replicated conditions of time pressure and cognitive load and have found distinct results from the studies mentioned in the former section. Moreover, other studies, by involving new fine-grained methodologies, have tested MDI's core hypotheses.

Tinghög et al. (2016) conducted a large experiment to replicate both conditions of time pressure and cognitive load. No significant effect on utilitarian judgments was found by the experimenters in conditions of cognitive load and under time pressure. Gürcay and Baron (2017, study 3) replicated the same conditions of Suter and Hertwig's study but they found an opposite tendency: under time pressure, utilitarian intuitions are more frequent than deontological intuitions. A similar tendency is also present in Rosas and Aguilar-Pardo's study (2019). The authors show that utilitarian judgments tend to increase with more extreme time pressure (between 18-26 seconds after reading the dilemma) than what had been imposed by the previous experiments.[19]

As I said, the use of new experimental methodologies favored more decisive tests for MDI. An interesting method is provided by mouse-tracking technology: the interviewed subjects, after reading about a moral dilemma, have to indicate their moral judgment (deontological or utilitarian) by moving the mouse from the centre to one side of the screen. By studying the curvature of the trajectories of the mouse, the experimenters can assess the level of confidence of the subjects' responses. Consistent with MDI's corrective hypothesis, utilitarian responses were expected to swing more than deontological answers, insofar as they would result from the correction of a deontological intuition. Nevertheless, Koop (2013), as well as Gürcay and Baron (2017, study 4), found that switches occur in both directions. This provides evidence that both deontological and utilitarian reasoners are sensitive to the conflict of reasons. These results are consistent with Bialek and De Neys (2016, 2017), which measured people's judgments by two parameters: confidence and decision time. The experiments show that deontological judgments are slower and less confident than utilitarian judgments. Moreover, a significant increase in confidence is observed when the moral scenario presented is not

[19] In contrast with those studies, Mata (2019) reported that consequentialist reasoners tend to feel more conflicted than deontological reasoners.

conflicting, that is, when the deontological and utilitarian theories converge on the same solution. This suggests that deontological judgments seem sensitive to utilitarian reasons in conflict scenarios.

Bago and De Neys (2019) involved a "two response" paradigm (Thompson, Turner and Pennycook 2011) to test the corrective hypothesis properly. In this approach, first, the participants are instructed to answer to a moral dilemma as quickly as possible with the first response that comes to their mind (a type 2 processing activation is thereby ruled out in this task); afterwards, the moral problem is presented again, and the subjects can take as much time as they need to reflect and give a second answer. If, as MDI predicts, moral reasoning is corrective, utilitarian second answers should be preceded by type 1 deontological answers. However, the empirical results contradict MDI's prediction: most utilitarian responses are already given in the first quick answer; therefore, the utilitarian reasoners do not need to override an intuitive deontological response. In addition, if one looks at the non-correction rate between the first and the second response, one notices that they end up being quite high (70-90%) independent of the nature of the judgments; this suggests that the influence of reflection in moral judgment is not corrective but confirming; hence, the corrective hypothesis of MDI is contradicted even when ruling out the correspondence between type 1 processing and deontologism and between type 2 processing and consequentialism. Vega and colleagues (2020), adopting the "two-response" paradigm, substantially confirmed the results from Bago and De Neys (2019): few participants in the second response revised their moral judgments from one category to another, and revisions from deontological responses to utilitarian were not more common than revisions in the opposite direction.

All the evidence mentioned so far seems to go toward a common direction which can be summarized in two distinct points. First, the content of a moral judgment does not appear to be a reliable indicator to predict its reflective or intuitive nature: the evidence shows that confident moral intuitions can be utilitarian and, conversely, deontological judgments can be slow, reflective, and sensitive to utilitarian reasons. This point challenges Greene's dual process theory that posits a correspondence between intuition and deontological judgments on the one side, and between reflection and utilitarian judgments on the other side. Second, how much people engage in moral reasoning seems to depend on the conflicting nature of the dilemma: the more utilitarian and deontological reasons conflict, the less confident are the resultant intuitions and the more the subjects

activate type 2 processing. For instance, strengthening the utilitarian reason, by increasing the number of people that one can save, increases the confidence of utilitarian intuitions; by contrast, strengthening the deontological reason, by including family members in the sacrifice, generates the opposite effect to decrease the confidence of utilitarian intuitions and, thereby, to favor reflection. The stronger reason (e.g., utilitarian or deontological) tends to prevail in the final response given by the subject; however, this does not mean that the subject is insensitive to contrary reasons: this is testified by the low confidence of the response and by the time that the subject takes before judging. The personal or impersonal nature of the dilemma (Greene 2008) did not result to be so effective on confidence as the strength of the reasons involved.

5. Towards a dual process reflective equilibrium

In the previous sections, I discussed the empirical evidence for the two core hypotheses of MDI. All considered the examination of the different kinds of evidence puts into question the empirical claim that intuition favors heuristic and deontological responses, as well as the empirical claim according to which reflection tends to be corrective. This leaves room for a major reconsideration of the nature of moral reasoning and the distinct roles of the two types of information processing.

According to MDI, the role of moral intuitions is crucial, but not within the reasoning sphere, where, given the unfamiliarity of the problems, moral heuristics need to be overridden. Nonetheless, the view of moral intuition as heuristic seems limited in light of the recent evidence, which highlights that intuitions can be unconfident and sensitive to utilitarian reasons. Therefore, the role of type 1 processing in moral reasoning should be reconsidered. The role of type 2 processing should be rethought as well, insofar as its corrective nature has been put into doubt by recent empirical findings. Additionally, I have shown that a plausible reading of the literature on psychopaths favors an integrated account of affective and reflective processes in moral reasoning.

In what follows, I offer some possible interpretations of the two types of information processing in moral reasoning. Contrary to Greene's theory, my account does not posit a correspondence between the type of information processing and the moral content of the response. Nor does my account proceed from a correlation between the type of processing and the epistemic correctness of the outcome judgment. That would constitute a serious "normative fallacy" (Evans 2019, 387). An intuitive judgment is not

necessarily more biased and irrational than a reflective judgment; on the other side, moral deliberation does not always provide more rational responses than intuitive judgments. Trivially, people can have very accurate intuitions, as well as moral reasoners can fail to deliberate.

Unlike MDI, the picture I offer is not conflicting: the roles that type 1 and type 2 processes perform are complementary and interdependent. On the one hand, the task of type 1 processing consists in tracking relevant information from a moral problem and calling for type 2 processing whenever conflicting reasons are at stake; on the other hand, moral reflection aims to rationalize the cognized information in order to achieve a justified judgment. Inspired by the influential philosophical method (Daniels 2016), I call this kind of interplay between moral intuition and reflection *dual process reflective equilibrium*.

5.1 Type 1 processing in the moral domain: reasons tracking and conflict detection

It is largely acknowledged that one of the hallmarks of automatic thinking is its efficiency: automatic processes allow the mind to select relevant information quickly and without much cognitive effort (Kahneman 2011). Moral intuitions constitute an indispensable tool for filtering and selecting information in accordance with some endorsed values, without the mental effort of deliberation. This *reasons tracking*[20] process enables the moral agent to determine which facts from within a large amount of available information to pay attention to and which to ignore (Maiese 2014).

Various automatic mental processes can contribute to reasons tracking. Moral emotions are probably the most efficient processing. Indeed, by experiencing anger, disgust, or guilt, a subject can become immediately aware of the salience of a certain object. In addition, some evidence suggests that the perceptual system can be attuned to detect morally salient facts (Gantman and van Bavel 2015). Importantly, the affective and perceptual systems are not cognitively impenetrable; over time and habituation, reflective beliefs can influence emotions and perceptions (Sauer 2017). Therefore, even reasoning can indirectly contribute to the automatic recognition of reasons.

---

[20] By "reason", I mean a salient fact that favour a certain course of action (Mantel 2018). I will come back to the concept of practical reason in Chapter 5.

The task of type 1 processing is not limited to tracking morally salient information. A crucial metacognitive task it performs consists in monitoring whether the incoming information favors conflicting responses. As the evidence suggests, this *conflict detection* process occurs at the pre-reflective stage of reasoning (Thompson et al. 2011; Bago and De Neys 2019; Vega et al. 2020).

Conflict detection is possible to the extent that moral reasons are processed with various strengths at the *intuitive* level of reasoning (Bago and De Neys 2019). The difference in the strength of the competing reasons is positively correlated to the level of confidence of the resultant intuition (Gürcay and Baron 2017; Bago and De Neys 2019; Vega et al. 2020). That means, in other words, the more a kind of reason (e.g., utilitarian) is prevailing upon the others, the more confidently the subject is inclined to provide a utilitarian response. Confidence of intuitions, in turn, predicts the cognitive engagement that subjects undertake to process the evaluative conflict. Said otherwise, the more an intuition is confident, the less likely the subjects engage in moral reasoning. That means that, if intuition is unconfident, it is more likely that the reasoner takes more time to reflect before endorsing a certain judgment; by contrast, if the problem does not present competing reasons and the resultant intuition is confident, it is likely that the subject provides a spontaneous judgment.

All this evidence coheres with the metacognitive account of moral intuition, according to which moral intuition is characterized by two components: a moral content and a metacognitive feeling. The content of intuition is constituted by a prevailing inclination towards a certain moral response (e.g., consequentialist or deontological). In addition, the moral content is accompanied by a metacognitive feeling: a feeling of confidence that is influenced by the presence (or the absence) of a detected conflict between reasons. The metacognitive feeling accomplishes the function to incline the reasoner to provide a fast and spontaneous response or a slow reflective judgment. So conceived, moral intuitions, rather than mere heuristics, play an important role at the interface between type 1 and type 2 information processing.

How able a subject is in detecting conflict among reasons depends on her *metacognitive sensitivity*, i.e., the capacity to correctly calibrate intuitive confidence according to the context. Individual differences in such an ability may depend on moral beliefs. For instance, utilitarian reasoners will be more confident in their utilitarian

intuitions.[21] In addition to moral beliefs, individual differences in *cognitive style* might significantly contribute to a subject's metacognitive sensitivity. I mean, for example, how much a subject considers different courses of action in making an ethical decision, how much a subject tends to trust her intuitions, how much a subject tends to review the elements of an ethical dilemma, how much a subject asks herself what is important before engaging in the decision-making process.[22]

5.2 Moral reflection as rationalization

Based on the evidence discussed in this chapter, moral reasoning, rather than being corrective, tends to confirm intuitive responses. This does not mean that type 2 process cannot be corrective but is still "the exception rather than the rule" (Bago and De Neys 2019, 1794). This means that moral reflection tends to be *rationalizing*.

Rationalization has a bad reputation in philosophy (Audi 1985, D'Cruz 2015). It is usually understood as tendentious reasoning by which a subject distorts facts to invent an explanation that casts her behavior in a favorable light. In particular, it has been argued that reflection in defense of moral intuitions tends to be biased, self-deceptive, and blind to counterevidence (Haidt 2001). However, in contrast with this widespread view, I understand the term "rationalization" with no pejorative meaning.

Here I embrace Cushman's account of rationalization. Cushman defines rationalization as a kind of "representational exchange", i.e., "the process of translating information from one psychological system or representational format, into another" (Cushman 2020, 9); specifically, the function of rationalization is extracting implicit information from unconscious and adaptive systems (such as instincts, social norms, habits or emotions) and making it accessible through conscious representations (i.e., beliefs or desires). The aim of rationalization is not to produce an accurate reconstruction of the effective causal mechanism that has generated implicit cognition; rather,

---

[21] Gürcay and Baron (2017) developed a Rasch model to predict the response time of a moral judgment in a sacrificial dilemma; interestingly, the model combines the tendency of a subject to provide a yes answer to a sacrificial dilemma with the tendency of a problem to elicit a deontological response.

[22] The *Moral Metacognition scale* (McMahon and Good 2016) provides an overall measurement of the subjects' thinking dispositions toward moral problems.

rationalization is a "useful fiction" (Cushman 2020, 7) to the extent that it makes useful information from adaptive systems accessible by constructing beliefs and desires as fictional motives of behavior.

The typical representational exchange that occurs in moral reasoning is between automatic and mostly affective moral representation (provided by intuitions) and reflective judgments. More precisely, when a subject rationalizes her intuitions, she transforms the received information into a more articulated and accessible moral judgment. To this purpose, the subject makes use of conscious moral beliefs (e.g., deontological or consequentialist principles), even though such beliefs were not part of the causal process that generated the intuition.

Therefore, according to this neutral understanding of rationalization, the fact that moral reflection tends to be rationalizing does not undermine the rationality of moral reasoning, *per se*. Whether moral reasoning is irrational depends on the epistemic attitude with which the reasoner rationalizes. Accordingly, one can state that there are two kinds of rationalization: rationalization "of the good kind", in which a reasoner articulates a sound and compelling justification and rationalization "of the bad kind" (or *confabulation*), i.e., "demonstrably inaccurate, tendentious after the fact reasoning" (Sauer 2017, 68).

Empirical evidence on how people rationalize moral intuitions are mixed thus far. In their influential study, Haidt and colleagues presented to a group of subjects some disgust-eliciting scenarios involving dead pets eating, sex with dead animals, incestuous intercourses, and other taboo violations (Haidt 2001); many of the interviewed subjects judged the situations as morally wrong and yet, when questioned, failed to provide any compelling reason in support of their judgment. Haidt called this phenomenon "moral dumbfounding". Nevertheless, a more recent replication of the experimental conditions did not obtain the moral dumbfounding effect (Royzman, Kim and Leeman 2015). Other relevant studies reported that the accuracy of moral justification varies according to the type of judgment (Cushman, Young and Hauser 2006, Hauser, et al. 2007); for example, the majority of the subjects were able to justify the distinction between harm provoked by action and omission, whereas only a few subjects were able to justify judgments based on the doctrine of "double effect". Finally, a recent study showed that a large number of participants were able to identify the reasons for their moral judgments with great accuracy and specificity (Farsides, Sparks and Jessop 2018). In sum, these data suggest

that not all, nor the majority of, people confabulate when they justify intuitions. Therefore, moral reflection can be rational, although post-hoc rationalization.

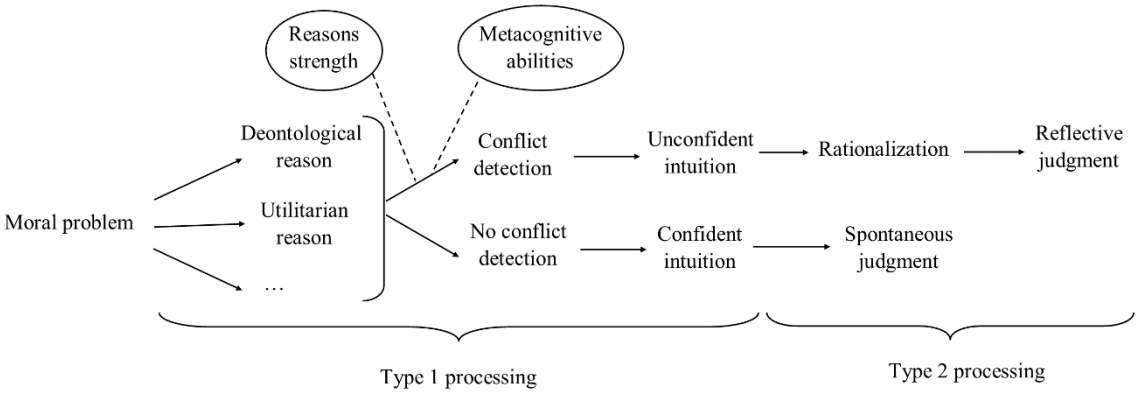

Figure 4. Dual process reflective equilibrium.

6. The puzzle of rationalization

The fact that moral reasoning tends to be rationalizing raises an important puzzle: why do moral agents spend cognitive resources in rationalizing judgments already triggered by automatic processes? For what purpose do people rely on reasoning after less confident intuitions if its scope is not to correct them? The evidence that moral reflection tends to be rationalizing may suggest that the purpose of moral reasoning goes beyond judgments per se: possibly, the moral reasoner aims to *justify* her judgments. However, what are the benefits of moral justification? In addition, why do people tend to justify more extensively challenged intuitions? In what follows, I present some hypotheses that address these issues.

Undoubtedly, rationalization provides some benefits for the individual. For example, it tends to reduce *cognitive dissonance*, that is a mismatch between thought and action or between moral beliefs (Festinger 1962). Cognitive dissonance is psychologically aversive and motivates people to achieve consonance. Rationalized moral beliefs and desires that fit automatic thought and behavior reduce psychological discomfort due to cognitive dissonance and, thereby, are beneficial for the individual. Another individual advantage of moral rationalization is that it helps the agent to act consistently (Summers 2017, S27-S28). If one acts in a certain way and then rationalizes the motives, one assumes the outcomes of reflection as the real motives of the action. As a result, the provided justification will exert some pressure on the reasoner to act in conformity with the identified reasons in future situations. For example, suppose I

donated money to a beggar passing by the street. The real explanation why I donated is because I was scared; nonetheless, I conclude by rationalization that I donated money because the beggar was suffering. To the extent that I sincerely identify myself with this explanation, I will feel pressure to be consistent and to make charity in a future similar situation.

The benefits for the individual could not be enough to explain why people rationalize so extensively. Perhaps, the main advantages of moral rationalization can be observed at the *social* dimension of reasoning. Indeed, making certain moral intuitions explicit and accessible can favor communication among reasoners. In an intersubjective context, individual moral intuitions can be challenged, and the reasoner is forced to articulate sound and convincing reasons to defend her viewpoint (Sauer 2017, 85-127). Importantly, the social dimension of moral reasoning can correct some individual cognitive biases (Mercier and Sperber 2017). The reasons are manifold. First, people are more accurate in evaluating others' arguments than the ones provided by themselves (231). Second, social communication is beneficial for reasoning insofar as, in an interactive discussion, people exchange many short arguments, and this permits them to reach longer and sound arguments with little individual cognitive effort (224). Third, social reasoning tends to favor coherent justifications since it is easier to convince an interlocutor of a claim by showing that the claim is coherent with what the interlocutor believes (194).

The claim that social reasoning can favor intuition overriding finds support in an influential study conducted by Paxton and colleagues (2012). The authors presented to the participants some scenarios that elicit strong intuitions, among which Julie and Mark vignette. The goal of the study was to observe whether the subjects would change their intuitive judgment by allowing more time to deliberate and by suggesting different counterarguments. The "weak" argument is the following:

> A brother-sister relationship is, by its nature, a loving relationship. And making love is the ultimate expression of love. Therefore, it makes perfect sense for a brother and sister, like Julie and Mark, to make love. If more brothers and sisters were to make love, there would be more love in the world, and that is a good thing. If brothers and sisters were not supposed to make love, then they wouldn't be sexually compatible, and yet they are. Brothers and sisters who don't want to make love should at least try it once. There is nothing wrong with trying something once. Thus, it wasn't morally wrong for Julie and Mark to make love. (170f)

In contrast, this is the "strong" argument:

> For most of our evolutionary history, there were no effective contraceptives, and so if siblings slept together they might conceive a child. Children born of such closely related parents would have a lower than normal likelihood of surviving. Thus, feelings of disgust toward incest probably evolved to prevent such children from being born. But in Julie and Mark's case, two kinds of contraception were used, so there was no chance of conceiving a child. The evolutionary reason for the feeling of disgust is, therefore, not present in Julie and Mark's case. Any disgust that one feels in response to Julie and Mark's case cannot be sufficient justification for judging their behavior to be morally wrong. (170)

Interestingly, only the strong counterargument was effective in inducing the subjects to override the intuition that Julie and Mark's behavior is wrong. This means that the subjects' deliberation was sensitive to the quality of reasoning.

In sum, the documented benefits of intersubjective reasoning suggest that moral reasoning is corrective at the social level. Accordingly, the purpose of rationalization would be to prepare the individual to reflect with other individuals; the exchange of reasons, in turn, would tend to challenge and override problematic intuitions.

7. Conclusion

This chapter has discussed the evidence for the default interventionist view in the moral domain and, consistent with recent empirical findings, has offered an alternative account of the interplay between intuition and reflection in moral reasoning.

The main predictive hypothesis of MDI is that intuitive thinking favors heuristic deontological responses, whereas reflection enhances utilitarian judgments resulting from the overriding of a deontological intuition. This view finds support in some empirical studies that manipulate time limits and cognitive resources and in some studies that highlight that psychopaths are more utilitarian than healthy people are. However, the evidence turned out to be inconclusive and challenged by an increasing amount of counterevidence. These latter empirical studies suggest that the content of a moral judgment is not predictive of its reflective or intuitive nature; rather, the conflicting or non-conflicting nature of the dilemma predicts cognitive engagement. Moreover, an appealing interpretation of the literature on psychopaths suggests that moral intuitions and reflection are integrated and not conflicting.

In light of the examination of the evidence, I sketched some possible alternative interpretations of the distinct roles of intuition and reflection in the moral domain, conceiving of them as interdependent rather than conflicting (*dual process reflective equilibrium*). In my account, type 1 processing encompasses two levels of cognition. In the first level, different kinds of reasons are recognized and processed with various strengths (*reasons tracking*). In the second metacognitive level, the presence of an evaluative conflict is monitored (*conflict detection*). This second level of cognition is crucial to send a signal of doubt that calls for more effortful information processing. Instead, type 2 processing accomplishes the task of rationalizing pre-reflective moral intuitions.

The rationalizing nature of moral reasoning is problematic for moral psychology. I suggested different hypotheses to explain why individuals do not tend to correct their moral intuitions. The most promising explanation states that rationalization is beneficial because it favors an intersubjective exchange of reasons.

II

MORAL INTUITIONISM

Chapter 3

Moral intuitionism and the reliability challenge

1. Introduction

Moral intuitionism is the view according to which accepting moral intuitions is justified in the absence of contextual defeaters. In recent years, intuitionism has been attacked by debunking arguments showing that intuitions are unreliable because they are influenced, for example, by framing effects (Rehren and Sinnott-Armstrong 2021, Wiegman, Okan and Nagel 2012, Schwitzgebel and Cushman 2012, Sinnott-Armstrong 2008), hypersensitive emotions, such as disgust (Kelly 2011), or hyposensitive emotions, such as compassion (Västfjäll, et al. 2014). I will call this the *reliability challenge* to moral intuitionism.

This chapter elaborates on a promising strategy for responding to the reliability challenge. In short, I will appeal to the fact that moral intuitions come with different levels of strength. The experimental results on which debunking arguments are based show the unreliability of a *weak* and *unstable* type of moral intuition concerning complex and artificial moral scenarios. In contrast, the reply goes, moral intuitionism concerns *strong* intuitions, not subject to change under realistic circumstances. Such a defense of intuitionism has been explicitly or implicitly suggested by many authors (Shafer-Landau 2008, Wright 2010, Bengson 2013, Liao 2008). However, it has not yet been sufficiently explored.

In the present chapter, I will refer to confidence to capture intuitions' strength. On this basis, I will argue that to tackle debunking arguments, intuitionists need a convincing explanation of why people's metacognitive sensitivity should be reliable. As long as intuitionists do not provide evidence in favor of the reliability of intuitive confidence, the reliability challenge remains open. This is, in brief, my contribution to the debate.

The chapter proceeds as follows. Section 2 introduces moral intuitionism. In the following section, I present some debunking arguments and how they undermine intuitionism. In Section 4, I explain how intuitionism can appeal to the role of intuitive

confidence in defending the reliability of moral intuitions. Then, Section 5 discusses whether confidence is conducive to moral truth.

2. Moral intuitionism

Intuitionism has a longstanding history in moral philosophy. It was a dominant theory in the late nineteenth and early twentieth centuries (Sidgwick 1874, Moore 1903, Ross 1930). Many metaethicists, especially from the 1950s to the 1990s, regarded intuitionism as an untenable theory because it was based on a mysterious mental faculty (cf. Mackie 1977). Since more plausible accounts of moral intuition were offered, intuitionism has been reevaluated and defended by prominent scholars in recent decades (Stratton-Lake 2002, Audi 2004, Huemer 2005).

Although, traditionally, ethical intuitionism has been considered a synonym of non-naturalist realism, here I understand it as an epistemological claim. Specifically, I define intuitionism as the view that *accepting moral intuitions is justified in absence of contextual defeaters*. Such a definition is silent about the metaphysics of moral facts, which will not be discussed here. The truth of intuitionism entails the existence of moral knowledge, which is incompatible with nihilist accounts of ethics, such as error theory or emotivism. However, the possibility of moral knowledge is compatible with different accounts of moral objectivity and truth (e.g., non-naturalism, naturalism, expressivism, constructivism).

As argued in the present research, moral intuitions are automatic mental states. Merely *having* a certain intuition cannot be subject to epistemic evaluation.[23] Rather, what can be assessed is the deliberative act of accepting or endorsing one's own intuition. By "accepting moral intuitions", I mean, for instance, *forming* a belief on the basis of intuition, as well as *maintaining*, *endorsing*, or *sustaining* a belief on the basis of intuition. Intuitionism, as understood in the present research, is the claim according to which such mental acts are epistemically legitimate.

To my knowledge, no intuitionist considers moral intuitions indefeasible. Rather, it is largely acknowledged among intuitionists that, in a moral inquiry, considerations of

[23] One could object that one can be blamed for *not having educated intuitions*. However, the object of evaluation here is the character of the subject and not the single intuition, which is the focus of the present discussion.

different kinds can prevent subjects from accepting a certain intuition. Specifically, there are two ways to defeat an intuition. First, a consideration can *undermine* the source of the intuition. For example, suppose I have the intuition that LGBT adoption is wrong; the consideration that I know that I have grown up in a strongly catholic family might undermine the reliability of my intuition. Second, intuition can be *outweighed* by stronger considerations. For instance, consider Haidt's vignette of Julie and Mark (Ch. 1). One could have the intuition that incest is wrong, but the fact that there is a convincing reasoning that concludes that Julie and Mark's conduct is permissible might outweigh the intuition. Importantly, intuitionists point out that both types of defeaters (undermining and outweighing) are *contextual*. This means, in other words, that accepting moral intuitions is *generally* permissible, and in the absence of defeaters, a subject can trust her own intuitions.

According to some authors (Väyrynen 2008, Sturgeon 2002, Cowan 2013), ethical intuitionism is the view that there is some non-inferential moral knowledge. In this account, intuitionism is synonymous with *foundationalism*. However, the adopted definition of intuitionism diverges from this understanding. The purpose of this and the following chapter is not to understand whether moral intuitions can ultimately ground moral knowledge to respond to the discussed "regress of justification problem".[24] Rather, here accepting moral intuitions is understood as a *reasoning conduct*, i.e., a way in which a reasoner manages her cognitive resources to reach a reflective goal. Reasoning conducts include moral theorizing but also ordinary reasoning, in which a subject has to make a moral decision in limited time. I will assume that a reasoning conduct is placed in a social dimension, in which the reasoner exchanges arguments with other agents and receives feedback from them. The aim of the present discussion is to evaluate whether the practice of accepting moral intuitions is epistemically legitimate (i.e., rational) in normal conditions. As reasoning conduct, accepting moral intuitions is legitimate whenever it is *epistemically responsible*, given certain cognitive resources and proper epistemic goals such as moral knowledge or, as I will contend later (Ch. 4), moral understanding.

[24] I do not think the question itself is very clear. The concept of "non-inferential" is ambiguous, and the structure of moral justification can vary according to the reasoning context (Timmons 1999, 178-246).

3. The reliability challenge to moral intuitionism

Skepticism (or pessimism) of moral intuitions states that accepting intuitions is not justified. The skeptics object to moral intuitionism that intuitions' defeaters are not merely contextual, but rather intuitions per se tend to be mistaken (Sinnott-Armstrong 2008). In this view, accepting moral intuitions should be the exception rather than the rule: only in certain contexts (e.g., when a subject is very familiar with a moral problem) can a subject rely on her intuitions. In particular, many authors have defended such skepticism by arguing that intuitions are *unreliable*.

A mental process is reliable whenever it is *truth-conducive*, i.e., tends to generate true beliefs. Undoubtedly, the reliability of moral intuitions is relevant for the epistemic justification of intuitive beliefs. If intuitions systematically lead to believe false propositions, then accepting intuitions is hardly justifiable. For this reason, it is important to assess the truth conduciveness of moral intuitions to consider their epistemic justification.

Evaluating how much moral intuitions lead to true beliefs is not a simple task, to the extent that moral truths are subject to normative disputes. A possible method consists of observing whether moral intuitions can be influenced by factors that are uncontentiously morally irrelevant; if people's intuitions are affected by morally irrelevant factors, then one can conclude that they are not highly reliable. Therefore, accepting this methodology, the question of the reliability of moral intuitions can be empirically investigated.

Following this method, some authors have developed different kinds of "debunking arguments" to question the reliability of moral intuitions. A debunking argument proceeds from a descriptive premise, which states that a certain set of moral beliefs is based on a process P; then, it assumes that P is epistemically defective and concludes that the set of moral beliefs is unjustified (Sauer 2018). In what follows, I shall consider three paradigmatic examples of arguments that debunk the reliability of moral intuitions.

One of the most influential debunking arguments in recent years appeals to studies reporting that the content of moral intuitions is subject to *framing effects* (Rehren and Sinnott-Armstrong 2021, Wiegman, Okan and Nagel 2012, Schwitzgebel and Cushman 2012). In particular, some studies report that the *words* through which a moral scenario is described affect people's judgment. For example, subjects are more prone to positively

judge a certain action if told how many people the action *saves* than if they are told how many people it *kills*, even though the outcome is the same. Other studies show that the *order* in which different scenarios are presented influences moral intuitions. For example, it has been documented that manipulating the order of presentation of scenarios involving harm brought by action or by omission has some effect on how bad people regard the omission.[25] Since the mere framing (e.g., the words or the order) of moral scenarios is morally irrelevant, moral intuitions are influenced by irrelevant factors, or so the debunking argument from framing effects suggests.

On the basis of evidence on framing effects, skeptics conclude that moral intuitions are unreliable. The argument can be summarized as follows:

(P1) S arrives at the intuition that *p*, on the basis of frame F.
(P2) Had S received frame F', S would have arrived at the intuition that ¬*p*.
(P3) There is no morally relevant difference between F and F'.
(C1) Therefore, S arrives at the intuition that *p* influenced by irrelevant factors.
(P4) If S arrives at the intuition that *p* on the basis of irrelevant factors, then the intuition is unreliable.
(C2) The intuition that *p* is unreliable.

Another influential debunking argument proceeds from the empirical premise that a substantial number of moral intuitions are affected by disgust. This hypothesis finds confirmation in the already mentioned studies that highlight how inadvertently induced disgust produces harsher moral intuitions (Ch. 1). However, this empirical fact is not enough to debunk moral intuitions: it must also be argued that disgust is not conducive to moral truths.

As Kelly has argued, the function of the human disgust system was not originally moral, but it was designed to protect people from parasites and poisons in food (Kelly 2011). To the extent that the cost of ingesting a contaminated food is typically higher than wasting an opportunity for eating, the mechanism on which disgust relies is extremely prudent; many innocuous objects trigger a repulsive reaction just because they are unknown or suspicious. Over time, the disgust system has adapted to protect people from

[25] Note that what is assumed as irrelevant here is not the distinction between action and omission, but rather the mere order in which an action scenario and an omission scenario are presented.

violations of moral norms. Thus, moral disgust, being based on the same original prudent mechanism, tends to induce "hypersensitivity" toward moral violations such that many deviant behaviors or strange-looking people become immoral (Sauer 2018, 39). Although not completely off the track, this process produces a great number of moralistic false intuitions. For this reason, one can conclude that human disgust makes intuitions concerning wrongdoing unreliable.

The debunking argument from disgust can be summarized as follows:

(P1) Moral intuitions are influenced by disgust.
(P2) The disgust system is based on a hypersensitive mechanism.
(C1) Therefore, moral intuitions are influenced by a hypersensitive mechanism.
(P3) If a process is influenced by a hypersensitive mechanism, then the process is unreliable.
(C2) Moral intuitions are unreliable.

Finally, another relevant studied case concerns the difficulty of intuitive thinking to deal with great numbers. According to some evidence, people tend to show a diminished sensitivity in perceiving the moral relevance of catastrophic loss of life. While we experience a great aversion in attending the loss of one or a few lives, we do not feel much difference between 800 and 1000 lives. This means that our intuitive aversion is not proportional to the number of victims. This might make intuitions unreliable in the perception of risk (Slovic and Västfjäll 2010) and in altruistic and charitable behavior when the number of needy individuals grows (Västfjäll, et al. 2014). Västfjäll and colleagues call this phenomenon "compassion fade". In contrast to the case of disgust, the evidence on compassion fade suggests that moral intuitions can be influenced by a system of empathic concern that is *hyposensitive* because it does not activate in many relevant circumstances (or, at least, not proportionally to what some situations require) (Sauer 2018, 40).

In sum, debunking arguments, such as the argument from framing effects, from disgust, and from compassion fade, undermine the truth conduciveness of moral intuitions. If moral intuitions tend to produce false moral propositions, it is reasonable to conclude that accepting intuitions is unjustified and intuitionism is false. Call this the *reliability challenge* to moral intuitionism.

4. A reply to the reliability challenge

Different replies to the reliability challenge have been provided in the literature. Rini, for instance, has argued that debunking arguments entail a regress; since they are based on a normative premise (i.e., that a certain psychological process is not truth conducive), they are vulnerable to second-order debunking arguments about the reliability of the normative premise; the second-order debunking argument must rely on another normative premise, and so on (Rini 2016). Another employed strategy consists in contesting the debunking arguments empirical premises by arguing that the effect of epistemically defective processes on intuitions is low or not sufficiently strong to cast doubt about the reliability of moral intuitions. In support of this claim, Demaree-Cotton conducted a meta-analysis of the available data in the literature and estimated that only 20% of the reported moral judgments are vulnerable to framing effects (Demaree-Cotton 2016). Concerning the evidence on disgust, recent meta-analyses have revealed that the amplification of moral wrongness by disgust is not as robustly supported as it might appear (Landy and Geoffrey 2015, May 2018, 30-33): the emotional effect was found only among particular subgroups of people and, when detected, the influence is not strong.

Surely, *how much* subject to biases moral intuitions are is crucial for intuitionism, and only careful empirical research can answer this question. However, a purely quantitative approach may neglect some important features that determine the reliability of a mental process (Weinberg 2007). Consider, for example, sensory perceptions. These latter are usually considered as reliable not because they are infallible but rather mainly because the subjects know under what conditions they can trust them; the subjects know, for instance, that perceptual experiences do not deserve much epistemic credit when it is foggy, or one is under effect of drugs or alcohol; thus, perceptual mistakes can be easily prevented. In a similar vein, it is important to assess whether the subjects can prevent cognitive biases by tracking the reliability of their intuitions in a context sensitive way. For this purpose, one should consider under what conditions intuitions are reliable, beyond assessing their absolute susceptibility to biases. Therefore, *when* moral intuitions are reliable is as important as how much they are reliable.[26]

[26] On the basis of such considerations, Bengson, Cuneo and Shafer-Landau (2020) distinguish between reliability and trustworthiness, which entails "conscientious reliance" on a cognitive process.

Following this line of reasoning, rather than contesting the *depth* of debunking arguments, the objection I will consider questions the *target* of the arguments, that is, the class of moral intuitions that would be unreliable. The objection goes as follows. As argued in the previous chapter, moral intuitions can be distinguished according to their different level of strength. Intuitions vulnerable to framing effects or emotional influence are *unstable* or *weak*. For example, studies reporting framing effects involve complex moral dilemmas, which plausibly generate uncertainty in the subjects. Given the complexity and difficulty of the problem, it is also unlikely that subjects do not assign much weight to their automatic thoughts. Then, it is not surprising that intuitions deriving from situations of such kind tend to be influenced by irrelevant factors. In contrast, moral intuitions on which agents typically rely are strong and persistent over time and thus more resistant to the influence of irrelevant circumstances. Therefore, debunking arguments may target the wrong class of automatic responses, namely a class of moral thoughts that the subjects preventively know not to deserve much epistemic credit.

This argumentative strategy has been implicitly or explicitly suggested by different authors (Shafer-Landau 2008, Bengson 2013, Wright 2013) but has not been sufficiently explored. For instance, Shafer-Landau, in reply to the argument from framing effects, points out that there is a class of moral intuitions, such as intuitions about the wrongness of torture, rape, or deliberate humiliation, that are unlikely to be vulnerable to external influences:

> They are genuine moral beliefs, and the evidence about framing effects casts no doubt on their reliability. Neither does this evidence impugn the reliability of more specific, entirely uncontroversial moral beliefs of the sort I introduced at the beginning of the chapter. These are beliefs that are (for almost everyone) not subject to framing effects: They are invulnerable to change under realistic circumstances. (Shafer-Landau 2008, 92).

In line with such considerations, Liao contends that the experimental evidence against the reliability of moral intuitions does not consider the distinction between "superficial" and "robust" intuitions, which are the real justifiers in philosophical theorizing:

> some might think that one should distinguish between surface intuitions, which are ''first-off'' intuitions that may be little better than mere guesses; and robust

> intuitions, which are intuitions that a competent speaker might have under sufficiently ideal conditions such as when they are not biased. In other words, when philosophers assert that 'Everyone would agree that ...' or 'Intuitively, we would all find it obvious that ...' or 'It is clear to us that ...', the 'we' and 'us' should be interpreted as applying only to competent speakers in certain non-distorting conditions. (Liao 2008, 256)

In a similar vein, Bengson argues that debunking arguments do not distinguish between "unstable answers", i.e., guesses or quick hypotheses, generated by unfamiliar and not commonsensical scenarios, and "stable answers", i.e., genuine intuitions elicited by commonsensical and familiar scenarios (Bengson 2013, 522-523). Note that Bengson does not discriminate between strong and weak intuitions but separates intuitions from non-intuitions, which he calls "blind answers". However, the kernel of the argument is the same: debunking arguments miss the target because the class of judgments they undermine is not the same as the one to which intuitionism refers.

The considerations just outlined are based on the descriptive claim that moral intuitions are experienced with different degrees of strength. To support such a claim, how one understands moral intuition is decisive. In particular, there are three possible accounts to capture intuitions' strength. First, one might appeal to the fact that intuitions have *presentational phenomenology* (Bengson 2015, Chudnoff 2013). In this view, the strongest intuitions are those that present some content that strike most as true. Alternatively, one may understand intuition strength as the degree of *emotional intensity* of an intuition (Railton 2014; Kauppinen 2013). Third, one can understand intuitive strength as a high degree of *confidence* (metacognitive account). In earlier chapters (1-2), I have provided different motives to prefer this latter view rather than the quasi-perceptualist and emotional accounts of moral intuition. Therefore, in what follows, I will refer to strong intuitions as intuitions accompanied by a substantial degree of confidence.

Claiming that moral intuitions are experienced with different levels of confidence is not enough to defend intuitionism from debunking arguments. One must show that confident intuitions are also *stable*. In other words, it should be proven that how a subject relies on moral intuitions is proportional to the level of initial confidence. Indeed, the confidence and stability of an intuition are two logically distinct concepts. While confidence (or strength) is a subjective experience resulting from a metacognitive appraisal, stability is a behavioral measure denoting how resistant to external influences

a moral intuition is, i.e., how a subject is disposed to keep her opinion despite the disturbance of irrelevant factors. Whether there is a correlation between these two measures cannot be taken for granted.

In the previous chapters, I have provided some behavioral evidence showing a direct correlation between intuitive confidence and stability. In particular, the evidence collected by Wright (2010, 2013) reports that the induction of instability is effective only on unconfident intuitions, resulting from noncommonsensical scenarios (Ch. 1). Moreover, evidence on dual-process reasoning shows that how confident a subject feels about an intuition regulates the activation of reflection (Ch. 2). Such data, consistent with intuitionism, support the claim that only *some* moral intuitions (i.e., less confident intuitions) are vulnerable to irrelevant factors (Wright 2010, 492). Through metacognition, the subjects are implicitly aware of the likelihood of intuition and can activate reflection accordingly. In this fashion, they can protect their beliefs from biases and irrelevant factors.

To summarize, according to a promising line of argument, debunking arguments do not affect moral intuitionism to the extent that they do not prove the unreliability of strong and stable intuitions, but just the weaker and less stable intuitions. Such a defense of intuitionism requires two descriptive claims: first, that moral intuitions are experienced with different degree of strength (i.e., confidence); and second, that intuitive confidence tracks intuitions' stability. Both claims seem to be supported by the evidence.

5. Vindicating the reliability of moral intuitions

If intuitive confidence tracks intuition stability, this means that the subjects tend to accept their moral intuitions *proportionally to their level of confidence*. To put it more bluntly, the more a subject feels confident about an intuition, the more she is disposed to accept it, i.e., to maintain or endorse it; conversely, the more an intuition lacks confidence, the greater the subject is disposed to revise it. Such an empirical hypothesis is warranted by the empirical evidence discussed in earlier chapters and the preceding section.

However, the defense of intuitionism sketched in the preceding section is, importantly, incomplete. Indeed, the empirical claim that confidence tracks stability cannot be enough. To vindicate the reliability of moral intuitions, another empirical hypothesis is required: *the hypothesis that intuitive confidence is truth tracking*. In other words, it must be proven that confidence is a reliable indicator of moral truth. If this is

the case, the more a subject feels confident about an intuition, the more likely to be true such intuition is. Thus, through confidence, subjects can track moral truth and reliably accept their moral intuitions in proportion to their level of confidence.

On this basis, intuitionists can construct the following "vindicating argument"[27] in favor of moral intuitionism:

(P1) The subjects accept moral intuitions proportional to their level of confidence.
(P2) Intuitive confidence is truth tracking (i.e., reliable).
(P3) If P1 and P2, then the subjects accept their intuitions proportionally to their reliability.
(C1) The subjects accept their intuitions proportionally to their reliability.
(P4) If C1, then intuition-based beliefs are justified.
(C2) Intuition-based beliefs are justified.

The key premise of the argument is P2, which claims that confidence tracks moral truth. Before discussing P2, let me prevent a possible objection.

In contrast with the line of argument just sketched, one could point out that the content of strong and confident intuitions is usually trivial and philosophically uninteresting. Indeed, on average, people's most confident intuitions concern uncontentious moral principles, such as "torture is wrong", or "happiness is good". Rather, one could argue, what advances moral knowledge are intuitions that go beyond common sense and the "comfort zone" of the members of a community.

There are two possible lines of reply to the objection above. First, one should not underestimate the importance of commonsensical intuitions in moral reasoning practices. In particular, strong intuitions can help define and clarify ethical concepts. Accepting common intuitions is useful to anchor moral discussions in uncontroversial premises. Importantly, this practice allows the reasoners to spare cognitive resources for debating unclear and controversial issues. Second, as Wright (2016, 573) observes, one should consider that metacognitive sensitivity, by which a subject calibrates the confidence of her intuitions, can be refined with experience. Thus, expert moral reasoners (not

[27] On the opposite of a debunking argument, a vindicating argument aims to defend the legitimacy of a class of beliefs (in the present case, of intuition-based beliefs) by pointing out that the psychological process on which they are based is reliable (Sauer 2018, 209).

necessarily philosophers) can have confident but skillful intuitions about nontrivial moral propositions (I will elaborate this topic in more detail in Chapter 6).

6. Is confidence truth tracking?

As argued, the claim that intuitive confidence is truth tracking is the key premise to vindicate the reliability of moral intuitions. Intuitionists can employ two different strategies to support this hypothesis. Both strategies are not devoid of problems.

A first option consists in vindicating the reliability of confidence by appealing to evolution. Miscalibration of confidence leads to erroneous judgments or decisions due to overestimation of one's capabilities or underestimation of tasks and risks. Therefore, it would be odd that evolution has favored unreliable metacognitive sensitivity. Nonetheless, although intuitively plausible, the hypothesis that confidence accuracy is evolutionarily adaptive is contentious. In contrast with this claim, some authors have argued that overconfidence is evolutionarily advantageous because it encourages an individual to claim resources she could not otherwise win in the case of conflict, and it keeps the individual from walking away from conflicts they would surely win (Johnson and Fowler 2011, 319). In addition, it is contentious whether *moral* truth is consistent with evolutionary fittingness. This broad metaethical question cannot be discussed here; nevertheless, it is worth noting that the appeal to evolution to defend the reliability of moral intuitions' confidence probably requires a strong naturalist metaphysics showing the adaptive nature of moral facts.[28]

If the appeal to evolution is problematic, intuitionists might need to find more direct evidence showing the reliability of confidence in moral intuitions. Some of the studies mentioned in Chapter 1 (Zamzow e Nichols 2009, Wright 2010, Wright 2013) show that confident intuitions are less influenced by framing effects compared with unconfident intuitions. In addition, the confidence of moral intuitions seems to be sensitive to the degree of conflict between moral reasons, which can be considered as an important objective feature of the difficulty of a moral scenario (Bago and De Neys 2019, Vega, et al. 2020). This is good news for advocates of intuitionism. However, the evidence on people's metacognitive sensitivity does not justify great optimism (Koriat 2007, 303-307). Rather, ample evidence in different domains suggests that meta-

[28] See, for example, Sterelny and Fraser (2016).

ignorance (i.e., people's ignorance about their own ignorance) is widespread. The reasons are manifold.

First, ignorance is often *invisible* (Dunning 2011, 51-56). For any complex task, there is a class of relevant information (e.g., potential problems and risks) that a subject does not know she does not know because she cannot even conceive it. Because of such "unknown unknowns", people tend to overestimate the possessed knowledge necessary to accomplish a task.

A second cause of inaccurate metacognitive assessments is the subjects' tendency to rely on "reach-around" knowledge in domains in which they are completely ignorant:

> people take cues from the social situation they are in and their general world knowledge to cobble together enough apparent information to form an impression. That is, people reach back or around to any knowledge they have that might appear to be relevant, and then use it to impose some meaning on the questions they are asked and then to form a judgment. (Dunning 2011, 258)

This hypothesis is supported by evidence showing that people express knowledge about topics that were completely invented by researchers.

Third, metacognitive appraisals are often miscalibrated insofar as the subjects tend to rely on misguiding cues to assess the competence of a judgment. How quickly and fluently a judgment comes to mind is not necessarily a relevant indicator of accuracy; this is testified by the literature on heuristics and biases, which shows how the most fluent response to a reasoning problem can be often wrong (Kahneman, Slovic and Tversky 1982). Nor is the familiarity with a problem always synonym of competence; as I will argue (Ch. 6), the quality of experience matters just as much as the quantity in the acquisition of skills. Therefore, "metaheuristics" such as fluency and familiarity can produce an illusion of skillfulness (Kahneman 2011).

The fourth and most important reason to be pessimistic about metacognitive sensitivity is that meta-ignorance goes hand in hand with ignorance. People with substantial deficits in their knowledge in a given domain should not be able to recognize their incompetence in that domain, given the absence of a better term. In other words, incompetent people tend to overestimate their competence exactly because of their lack of competence. This hypothesis, known as the *Dunning-Kruger effect*, has been documented in a wide range of tasks and skills (Dunning 2011). If confirmed in the moral domain, the Dunning-Kruger effect would be particularly problematic for intuitionism;

one would conclude that as long as moral intuitions are unreliable in their content, there is no reason to consider their level of confidence reliable. In short, the reliability of confidence is directly dependent on the reliability of the content. This contrasts intuitionists' attempt to defend intuitions from debunking arguments according to which people can protect their beliefs from biases and irrelevant factors through metacognition.

In sum, both the evolutionary explanation and the employment of general evidence on metacognitive sensitivity are not easy strategies to argue for the reliability of moral confidence. As long as it does not offer a convincing explanation of why metacognition should be reliable, intuitionism remains exposed to second-order debunking arguments that show how intuitive confidence is subject to epistemically defective processes. Therefore, there is still work to do for intuitionism to tackle the reliability challenge.

7. Conclusion

This chapter has discussed a promising line of reply to the reliability challenge to moral intuitionism. According to the considered counterargument, the effect of irrelevant factors is limited to weak (i.e., unconfident) intuitions, and since the subjects tend to accept moral intuitions proportional to their level of confidence, the harm of irrelevant factors to the subjects' moral beliefs is negligible.

The vindicating argument elaborated in this chapter has the merit of shifting the question of the reliability of moral intuitions from intuition content to intuition confidence. As argued throughout this research, the role of confidence is crucial to understand how much credibility a subject assigns to a certain moral intuition and how much she is disposed to revise it through effortful cognitive processes. However, to support their vindicating argument, intuitionists still lack a convincing explanation of why intuitive confidence should track moral truth. As long as such an explanation is not provided, the reliability challenge remains open.

Chapter 4

The argument from limited cognitive resources

1. Introduction

In the preceding chapter, I showed that the reliability of moral intuitions is an open empirical question. The present chapter discusses whether intuitionism can be true while conceding to the skeptics that intuitions are not particularly reliable. Specifically, I will consider whether accepting moral intuitions is a legitimate epistemic practice given the alternatives and the cognitive resources available for the subject. I will call this defense of intuitionism *the argument from limited cognitive resources*.

The argument I will outline in this chapter aims to go beyond the two most influential defenses of moral intuitionism: *reliabilism* (Bengson, Cuneo and Shafer-Landau 2020, Railton 2014) and *phenomenalism* (Bengson 2015, Chudnoff 2013, Huemer 2005). Unlike pure reliabilism, the argument from limited cognitive resources does not regard truth conduciveness as the only significant factor for the justification of moral intuitions. However, unlike *phenomenalism*, I will not consider the reliability of moral intuition to be epistemically irrelevant. Rather, I will show that moral intuitionism can be defended by outweighing intuitions' limited reliability by considering how intuitions can be conducive to *moral understanding*.

Different authors have recently focused on the notion of moral understanding (Hills 2009, 2016, Callahan 2018, Howard 2018). However, it has not yet sufficiently highlighted how this turn in moral epistemology can be significant for intuitionism. If, as has been argued, moral understanding and not mere knowledge is the most significant epistemic goal in moral inquiry, there is room for vindicating the legitimacy of moral intuitions notwithstanding their limited reliability; or so I will argue.

The chapter proceeds as follows. In Section 2, I will argue that both pure reliabilism and phenomenalism are problematic and that a different strategy to defend intuitionism is needed. In Section 3, I will outline the argument from limited cognitive resources, which proceeds from the assumption that a moral reasoner, given her limited resources, is committed to accepting some beliefs without full reflective consideration.

Among the possibilities that require little cognitive effort, the reasoner has to choose between deferring to other reliable agents or accepting her own intuitions. I will clarify what deferring to others means in Section 4. Then, in the following section (5), I will compare moral deference with moral intuitions in light of the conduciveness to understanding. Finally, I will state some concluding remarks in Section 6.

2. Justified although imperfectly reliable?

As defined in the preceding chapter, moral intuitionism is the claim according to which accepting moral intuitions is epistemically justified in the absence of contextual defeaters. In the recent literature, two dominant argumentative strategies have emerged in defense of intuitionism: some authors (Bengson, Cuneo and Shafer-Landau 2020, Railton 2014) have argued that accepting moral intuitions is justified because the latter derive from reliable mental processes (*reliabilism*); other authors (Bengson 2015, Chudnoff 2013, Huemer 2005) have argued that moral intuitions are justified by virtue of their phenomenology (*phenomenalism*). Either account is not devoid of problems.

In the preceding chapter, I showed how the reliability of moral intuitions has been put in doubt by some debunking arguments. Although they can appeal to the role played by intuitive confidence, intuitionists still lack convincing empirical evidence for the reliability of the subjects' metacognitive sensitivity. Such considerations put reliabilism in a quite unstable position.

According to phenomenalism, the question of the reliability of the mental processes behind moral intuition is irrelevant for its epistemic justification. Rather, phenomenalists argue that the mere fact that moral intuitions represent some content as true with a certain strength is a sufficient reason to believe that content, in the absence of defeaters. The most influential version of phenomenalism is *presentationalism*, according to which moral intuitions provide epistemic justification because they have presentational phenomenology (Chudnoff 2013, Bengson 2015). However, as I showed in Chapter 1, the descriptive claim that moral intuitions have presentational phenomenology is questionable.

Beyond the psychological doubts about the presentational phenomenology of moral intuition, phenomenalism is also objectionable on epistemological grounds. It is doubtful how the occurrence of a type of mental state (i.e., intuition) can alone be a sufficient epistemic reason. A plausible epistemology should ground the legitimacy of

moral intuitions from the reasoning context in which moral beliefs are formed. Such a context of inquiry comprises an epistemically valuable goal, the subjects' capacity, and some available time and information. Under these conditions, the reliability of a source cannot be considered an epistemically irrelevant fact but should be weighed on the basis of the goal and the features of the inquiry. Furthermore, phenomenalism seems to neglect the social dimension of moral reasoning. To justify moral intuitions, one should consider the possibility of moral reasoner to rely on other agents' knowledge. For these reasons, phenomenalism is an unsatisfying defense of intuitionism. In short, what is needed is a justification of moral intuition that *outweighs* its limited reliability, without ignoring the question, and considering the possibility of the reasoner to rely on other subjects.

Given the unpromising solutions offered by phenomenalism and reliabilism, in what follows, I shall explore a different strategy to defend moral intuitionism. I will proceed from the consideration that an agent needs to form moral beliefs under conditions of limited cognitive resources, such as time, attention, and accessible information. Accordingly, agents have to choose how to manage their scarce cognitive resources. In other words, they have to decide which moral propositions should be accepted as true and which require full reflective scrutiny. As I will argue, despite its imperfect reliability, intuitive confidence is still the most valuable source to automatically assess the likelihood of a certain proposition. Therefore, it is rational (i.e., epistemically legitimate) for a subject to accept her strongest moral intuitions and to inquire into those propositions about which she feels more uncertain. Call this *the argument from limited cognitive resources*.

3. The argument from limited cognitive resources

Moral reasoning requires much cognitive effort.[29] However, moral agents form beliefs in an epistemic field circumscribed in time and space. To form true and justified moral beliefs, the agent can rely on limited cognitive resources. The *limited time*, for instance, impedes the agent from reflecting on all the moral propositions considered in a context of inquiry. The agent has also *limited attention* and can scrutinize only a limited number of considerations at a time. In addition, the inquirer can access a *limited amount of*

[29] See Chapter 2 and, for a comprehensive reviews of the literature on this topic, Kahneman (2011), Stanovich (2018), and Mercier and Sperber (2017).

information to justify her moral claims; for instance, some morally relevant information is hardly accessible by the agent because it requires much nonmoral knowledge to be appreciated. In certain circumstances, the information is excessive, and a subject needs much time and attention to select the morally relevant evidence.

Given the limited time, attention, and available information, a moral reasoner can hardly provide full reflective consideration to every considered moral proposition. Consider, for example, an undergraduate student (Sarah) that, after studying on history books the XIX century Imperialism, starts reflecting on why colonialism is morally wrong. Sarah takes into account various aspects of the phenomenon, such as the extreme violence, the imposition of the colonizers' culture to the colonized people, and the violation of the territorial rights. Then, she concludes that what is specifically wrong in colonialism is the fact that the colonizers prevent the colonized people from the possibility of self-determining. In justifying the wrongness of colonialism, Sarah takes for granted that self-determination is a fundamental right of a people and violating it is morally wrong. Sarah does not have sufficient philosophical background to know why self-determination is a human right; nor does she have time to reflect about it, since the focus of her reasoning is colonialism. Therefore, given the circumstances, she must simply accept the proposition that self-determination is a human right as true.

Note that the problem of limited cognitive resources does concern *every* moral agent, not just the less epistemically virtuous. Consider, for instance, the expert moral theorist who finds herself in the most ideal context to extensively reflect on significant moral claims. Nonetheless, even the moral theorist, for reasons of limited time, cannot consider every significant moral proposition to construct the theory; therefore, she must assume a set of "initially credible judgments".[30]

---

[30] One could object that the coherence of the considered judgments with the other elements of theory suffices for their justification. However, this pure coherentism is problematic. As has been argued (Kelly and McGrath 2010, McGrath 2020), the mere coherence among considered judgments, principles, and background theories is not enough to justify a moral theory. The considered judgments must be *credible*, independently of their being coherent with the other elements of the theory.

At this point, in light of the conditions of limited resources, one could prescribe that a moral reasoner should accept *only* those moral propositions supported by full reflective scrutiny and suspend judgment on those that cannot be fully investigated due to limited time and capacities. However, this demand is hardly feasible. First, the claim entails a regress: drawing conclusions from reflection requires the acceptance of some premises; reflecting on the premises requires the acceptance of other premises, and so on. Second, suspending every belief without full reflective scrutiny is not always responsible. The ultimate goal of moral inquiry is practical; moral reasoning should orient moral behavior toward correct moral decisions. Some moral situations demand a decision, notwithstanding the limited time and the agent's capacity; not endorsing any belief without full reflective scrutiny can lead to action paralysis. For these reasons, a moral reasoner can legitimately accept some moral propositions without full reflective consideration.

By the latter claim, I do not intend to justify cognitive laziness. Needless to say, the responsible moral reasoner should be vigilant for the evidence for her moral beliefs, if possible. I also assume that the inquirer should be prudent in endorsing moral propositions; the acceptance of a moral proposition should never be unconditional but open to challenges and revisions.

The fact that moral agents think under conditions of limited mental resources does not exempt them from being epistemically responsible in endorsing moral propositions. Rather, since agents have limited time and capacities, they must rationally manage their resources. The agents should also be careful and responsible in accepting propositions without cognitive effort. For example, accepting a moral proposition by guessing is arguably an irresponsible behavior.

Excluding patently wrong behaviors, such as guessing, two rational epistemic sources remain available for the reasoner with limited cognitive resources. First, the reasoner can rely on her intuitive confidence to assess the likelihood of the considered moral propositions. Accordingly, the reasoner can accept her strongest moral intuitions without much reflective scrutiny and focus on those propositions about which they feel more uncertain. Second, alternatively, the reasoner can infer some moral beliefs from other reliable and trustworthy agents, i.e., she can accept some propositions without much reflective scrutiny on the basis of the epistemic authority of other agents. Importantly, both options require little cognitive effort: moral intuitions are representations deriving

from automatic mental processes that do not involve working memory; deferring to others, by definition, entails the acceptance of a deferred proposition without scrutinizing the reasons why it is true. Therefore, the moral agent can rely on these two alternatives even in the absence of time and capacities to reflect.

Note that the two epistemic sources (moral intuition and the other agents) are not always in conflict and not always both simultaneously available.[31] Whether it is permissible to accept moral intuition or defer to others depends in part on the specific reasoning context. Nevertheless, it is worth considering which of the two options is in principle preferable from an epistemic point of view. In the following sections, I will argue that there are good reasons to consider intuitions' acceptance more epistemically valuable than deference, although the former ensures less reliability than the latter.

If cognitive resources are limited and, among the possibilities that demand little cognitive effort, moral intuition is the most valuable, then it is reasonable to conclude, consistent with intuitionism, that accepting moral intuition is legitimate, in absence of defeaters. The argument from limited cognitive resources can be summarized as follows:

(P1) Moral agents think under conditions of limited cognitive resources.
(P2) If P1, then agents are rationally committed either to defer some moral propositions from other agents or accept some moral intuitions.
(C1) Therefore, agents are rationally committed either to defer some moral propositions from other agents or accept some moral intuitions.
(P3) Accepting moral intuitions is better than deferring to other agents.
(P4) If C1 and P3, then accepting moral intuitions is justified.
(C2) Therefore, accepting moral intuitions is justified.
(C3) Therefore, moral intuitionism is true.

Thus far, I have shown that there are sound reasons to assume premises P1 and P2, from which C1 follows. In the next sections, I will discuss P3, which is a key premise to infer C2 from C1 and P4.

---

[31] A further element of complication, which I will not discuss, is the influence of other agents on the individual's moral intuitions. Sometimes, it is hard to discern whether a certain moral belief is based on intuition or deference to the extent that a subject can have confident moral intuitions about a matter just because is influenced by the community where she lives.

4. Moral deference: clarifications

Before comparing moral deference with intuition acceptance, it is important to clarify what is meant here by "moral deference". As is usually understood in the literature, moral deference is a case of *pure* and *direct moral testimony* (Fletcher 2016, Lewis 2020a). Let us consider the key aspects of this definition in turn.

An agent forms a belief by testimony whenever she accepts a certain proposition on the basis of another agent's supposed authority. I will assume that the recipient of the testimony is a mature moral agent.[32] In addition, I will also take for granted that the testifier is an epistemic superior of the deferrer.[33] This means that the testifier has some epistemic authority concerning the deferred propositions. In other words, the testifier is reliable and trustworthy on the matter at stake. Note that this assumption does not entail any specific account of expertise but just the possibility for the moral reasoner to rely on the testimony of other agents who possess more knowledge (or understanding) than she does, concerning particular moral topics.

Supposedly, in case of moral deference, the content of the deferred belief is purely moral, such as, for example, the proposition *that stealing is wrong*. In contrast, a deferred belief whose content is just morally relevant information (e.g., *stealing causes pain*) does not count as a case of moral deference.

Deferring to epistemic authorities about descriptive morally relevant propositions is a widespread and typically legitimate practice. In a moral inquiry, nonmoral deference is crucial because it can, in part, supplement the agent's limited cognitive resources. For instance, in the previous example, Sarah needs to defer to the history book to acquire relevant information to understand why colonialism is wrong. However, obtaining descriptive knowledge by deference cannot be sufficient to infer moral conclusions: the subject needs to know by other sources that the deferred descriptive propositions are *morally relevant*.[34]

---

[32] This assumption rules out discussions about the legitimacy of deference in cases of children or immature moral agents.

[33] Deferring to another agent who is not epistemically superior would be trivially wrong from an epistemic point of view.

[34] This claim entails what Sturgeon has called "the autonomy of ethics", i.e., the fact that ethical knowledge cannot be fully grounded in nonethical knowledge (Sturgeon 2002). Importantly, the

An agent *directly* relies on another agent's testimony when the recipient forms or revises a moral belief *solely* (or *mainly*) on the basis of the authority of the testifier. Therefore, direct deference does not include cases of indirect reliance on testimony, in which the other's testimony helps the agent understand some moral proposition. This occurs, for example, in what Boyd has called *cooperative testimony*, in which the recipient's cognitive states play a justificatory role in the formation of a belief by testimony (Boyd 2020). Another example of indirect reliance of testimony is considering another's opinion as advice (Sliwa 2012); in this case, the testimony is taken by the subject as a suggestion to consider the reasons in favor of the testified proposition. In cases of indirect reliance on testimony, unlike direct deference, testimony plays a role in the formation of a moral belief, but the agent does not accept the proposition only on the basis of the other's authority.

Discussing the legitimacy of indirect forms of testimony goes beyond the aims of this chapter. Regarding the argument from limited cognitive resources, it is important to compare moral intuition with deference. For this purpose, moral deference should be considered in its direct form because indirect forms of testimony might presuppose the legitimacy of moral intuition. For example, cooperative testimony can occur only if the recipient and the testifier share the same background intuitions.[35] In considering some advice, a subject needs to check the appropriateness of the advice through intuition, reflection, or a combination of the two.

With this framework in mind, consider the following example of pure and direct moral deference:

> *Research project*: Jack is a graduate student who is working on a research project on normative ethics. In his project, Jack assumes that happiness is a fundamental moral good. Since the project is in the early stages, Jack has not fully investigated the reasons why happiness is a fundamental good, but his belief is mostly based on intuition. One day, in the department corridors, Jack encounters Margaret, an

---

autonomy of ethics does not necessarily entail a nonnaturalist metaphysics of moral facts, but can be explained by naturalist, constructivist, or expressivist accounts.

[35] My impression, considering the examples provided by Boyd (2020, 24), is that cooperative testimony consists of a combination of moral intuition and testimony about morally relevant information.

> esteemed professor of ethics, who has several publications in top-rated philosophy journals. Jack mentions his project to Margaret, in which he considers happiness to be a fundamental moral good. Margaret replies that she disagrees; according to her, freedom, not happiness, is the most fundamental moral good. However, Margaret is in a hurry, and she has no time to explain to Jack why freedom is a fundamental good. After this episode, Jack immediately changes his mind and thinks that freedom is the fundamental moral good.

In this vignette, Jack's deference to Margaret is purely moral because the deferred belief has moral content, namely, the proposition that freedom is a fundamental moral good. Were Margaret a psychologist and had Jack deferred from her the belief that the majority of people consider freedom as the fundamental good, Jack would defer just a morally relevant information.

The vignette is also a case of direct testimony to the extent that Jack defers the moral belief that freedom is a fundamental moral good solely on the basis of Margaret's testimony: he revises his project (supposedly not just for pragmatic reasons) not because Margaret has offered any reason for her moral belief but just because Margaret is an esteemed researcher in the field, and she believes the proposition. Moreover, it is assumed that Jack revises his belief without critically assessing the validity of Margaret's opinion: he blindly trusts Margaret's epistemic authority.

The example just described is particularly interesting for the purposes of this chapter because it illustrates a tension between moral intuition and testimony. In the next sections, I will explain why Jack is wrong in revising his intuition-based belief, even though Margaret's testimony is highly reliable and trustworthy.

5. Moral deference and intuition: a comparison

In this section, I will compare moral deference with the acceptance of moral intuition from the point of view of the goals of moral inquiry. In particular, I will consider whether accepting intuitions is more conducive to moral understanding than deferring to other reliable agents.

5.1 Pessimism about moral deference

The phenomenon of moral deference has been widely discussed in recent years. Many authors have argued that agents have pro tanto reasons to refrain from inferring moral

beliefs from others, regardless of the reliability and trustworthiness of the testifier's opinion (Hills 2009, Howell 2014, Fletcher 2016, Callahan 2018, Lewis 2020b). This view is defined as *moral deference pessimism*.

The most influential argument for deference pessimism is based on the concept of *moral understanding*. According to this technical notion, if a subject *understands why* a moral proposition *p*, then she grasps (or appreciates) the reasons why *p* is true. Such a mental state requires a set of reasoning skills, such as the capacity to follow an explanation of why *p*, the ability to explain why *p*, or the ability to infer that *p* given relevant information (Hills 2009, 2016). For example, if one understands why colonialism is wrong, one can follow a justification of why colonialism is wrong given by someone else; when questioned, one can justify by one's own why colonialism is wrong, and one can conclude that colonialism is wrong by observing particular manifestations of the phenomenon.

The nature of moral understanding has been hotly debated in recent years. In the present chapter, I adopt a nonreductionist, and in part sentimentalist, account of moral understanding. According to this view, moral understanding is not a species of propositional knowledge. Understanding why *p* is not reducible to knowing that *p*, nor knowing why *p*. Unlike mere propositional knowledge, understanding why entails the grasp of the connection between a moral proposition that *p* and the reasons why *p* is true. Moreover, in the account I adopt, grasping the reasons why *p* entails having a sentiment of concern toward *p*, beyond the intellectual grasp of the reasons. In other words, to fully understand why *p*, a subject must be disposed to experience a set of fitting emotions in response to manifestations of *p*. For example, if one fully understands why colonialism is wrong, one is capable to feel anger or indignation before specific cases of colonialism. Importantly, the emotional component constitutes another important difference between full moral understanding and mere propositional knowledge. Furthermore, the emotional component clarifies the assigned role of understanding in moral action, given the motivational power of emotions (see Ch. 5).[36]

It has been argued that understanding, and not mere knowledge (i.e., true and justified belief), is the most significant epistemic goal (Pritchard 2010, Hills 2016).

[36] For a full defense of a sentimentalist account of moral understanding, see Callahan (2018) and Howard (2018).

Understanding is particularly valuable in the moral domain for different reasons (Hills 2009). First, moral understanding, compared with mere knowledge, puts the subject in a better position to justify herself to others. Being able to communicate reasons is essential to moral reasoning, given its social dimension (Ch. 2). Typically, moral reasoning proceeds from challenges to certain moral propositions (Sauer 2017, 85-127); understanding why a challenged proposition is true is important to properly address the challenge. A sound moral reasoning does not simply conclude that something is good or bad but *explains* why something is good or bad. Second, moral understanding favors *morally worthy actions*, i.e., actions motivated by concern for reasons. To the extent that a moral inquiry must orient toward the performance of good actions, moral reasoning should pursue moral understanding for this purpose. Third, and finally, moral understanding is a valuable goal because fully appreciating moral reasons constitutes an essential part (although not sufficient) of a good character required to reliably act well. A virtuous person does not act on the basis of superficial propositional knowledge but is motivated by a stable understanding of reasons, including a deep concern for them. Therefore, such considerations suggest that moral understanding is more valuable epistemic goal than mere knowledge. Accordingly, it is reasonable to conclude that a responsible moral reasoner should pursue a reflective goal consistently with the standards of moral understanding.

Providing a full defense of the adopted account of moral understanding goes beyond the aims of this chapter. For the present purposes, it is important to observe how deferring moral propositions from others is strongly *in tension with the acquisition of moral understanding*. The reasons are manifold. First, deference discourages the deferrer from looking for moral reasons. To the extent that deferring moral propositions entails the acceptance of an already settled moral view, the deferrer will be less inclined to engage with the reasons for the accepted proposition (Callahan 2018). For example, in *Research project*, Jack, being confident about Margaret's authority on the matter, will be less motivated to look for a justification for the claim that happiness is a fundamental good; moreover, Jack will likely delegate to Margaret the task of defending the claim if questioned. Second, deferring a moral proposition from another agent can hardly transmit the moral sentiment related to the proposition (Fletcher 2016). For instance, if a subject accepts that one must behave kindly with other people just because her mother says so, it is unlikely that the subject feels the appropriate concern for others to behave kindly.

Arguably, the lack of moral sentiment impedes the deferrer from fully appreciating the reasons for a certain moral proposition.

In sum, since it discourages engagement with reasons and cannot convey genuine sentiments, deference can hardly be conducive to moral understanding. To the extent that moral understanding is particularly valuable for reliably acting well and developing a good character, one can conclude that deferring to others is not epistemically responsible for a mature moral agent, notwithstanding the reliability of the testifier.[37]

5.2 Moral intuition and understanding

The truth of moral deference pessimism cannot suffice for defending intuitionism. To use an analogy, suppose one has to choose between two cars (car A and car B). From the mere fact that car A is old and expensive, one cannot conclude that one should buy car B, insofar as car B might be even older and more expensive than car A. To assess whether car B is preferable, one has to consider the salient features of car B. In a similar vein, from the mere fact that there are strong reasons to refrain from deferring moral propositions, one cannot infer that the alternative (i.e., accepting moral intuitions) is legitimate, insofar as there might be even stronger reasons for not accepting moral intuitions. Therefore, to defend intuitionism, one has to compare moral deference with moral intuition. Specifically, since one has assumed moral understanding as the most significant epistemic value, one has to evaluate how accepting moral intuitions is conducive to understanding compared with deference to others. This question has rarely been addressed in the literature.

In the previous chapter, I showed how the reliability of moral intuitions has been undermined by debunking arguments. However, how much debunking arguments affect the reliability of intuitions is debated.[38] The issue is undoubtedly relevant because moral

---

[37] The moral understanding explanation is not the only possible explanation of the wrongness of moral deference. For instance, non-epistemic explanations have been provided in the literature (Howell 2014, Fletcher 2016). Since the goal of this section is to evaluate moral deference from an epistemic point of view, I will not discuss these rival accounts.

[38] Demaree-Cotton has estimated that only 20% of moral judgments are affected by irrelevant framing effects (Demaree-Cotton 2016). Recently, Rehren and Sinnott-Armstrong have published a study reporting that the effect is much larger (Rehren and Sinnott-Armstrong 2021).

understanding is *factive*: one cannot understand why *p* if *p* is false. Accordingly, the more intuitions are reliable, the more conducive they are to moral understanding.

Since the reliability of moral intuitions is uncertain and discussed, one cannot establish an exact estimate of the probability of an intuition of being true. However, one can reasonably assume that the probability of an intuition of being true is lower than the one of the opinion of a competent and reliable testifier. Nonetheless, truth conduciveness is not the only relevant aspect for the acquisition of moral understanding; other significant factors concerning engagement with reasons and the experience of moral sentiments are at play. What is the relationship between the latter and moral intuition?

As Lewis (2020a, 471) observes, having the intuition that *p* cannot ensure understanding why *p;* one can have the strong intuition that happiness is good, without being able to articulate the reasons why happiness is good. As argued in Chapter 1, although the content of an intuition is conscious, the subject tends to be unaware of the mental process that leads to the content. Accordingly, the reasons for an intuition-based moral belief are not immediately accessible by the subject. Thus, intuition-based and deferred beliefs are very close in this respect. Nonetheless, there is a connection between moral intuitions and reasons that is absent in moral deference. Specifically, moral intuitions, more than deference, tend to favor the subject's engagement with reasons. Intuitions are strong mental states; they represent certain content as credible and incline the subject to give assent to and rationalize their content. Accepting moral intuitions exerts some pressure on the subject to look for reasons in a future inquiry and justify herself to other subjects if questioned. For example, by accepting his intuition as a basis for the research project, Jack will be more motivated to engage in first person with the reasons in support of his intuition and to defend his view from future challenges.[39]

Accepting moral intuitions is clearly preferable to deferring to others as concerns the connection with moral sentiments. As shown in previous chapters, different types of empirical evidence report a correlation between moral intuitions and emotion. How much emotions influence intuitions and what role they play in the formation of an intuition are open questions; nevertheless, it is quite likely that a subject who has the intuition that *p*

[39] One could object that the reasoning favored by intuitions is just biased confabulation, not aiming at understanding but self-defense (Haidt 2001). In response to this doubt, I have stressed the benefits of rationalization in Chapter 2.

will be disposed to feel fitting emotions toward manifestations of *p*. For example, if one has the strong intuition that stealing is wrong, one will be inclined to feel indignation or anger before cases of stealing. Experiencing emotions related to moral propositions is very important to fully appreciate the reasons for the propositions. One can hardly understand the normative relevance of certain considerations if the grasp of the latter is not accompanied by appropriate feelings. Moreover, beliefs tightly connected with relevant moral sentiments are particularly significant to motivate the subject to reliably act well in different circumstances and to act on the basis of sincere moral conviction.[40]

In light of the relationship between moral intuition, engagement with reasons, and sentiments, one can conclude that accepting intuitions can be conducive to moral understanding, despite their limited reliability. Although moral intuitions cannot guarantee full understanding, accepting moral intuitions can constitute the *first step* for an autonomous inquiry into moral reasons, consistent with the standards of moral understanding.

How much intuition is conducive to moral understanding is an empirical question. However, stating that intuitions are more conducive than deferring to others does not appear to be very controversial. Of course, how much moral intuitions are reliable can make the difference. If they were completely unreliable and blind to reason, accepting intuitions would hardly be conducive to moral understanding; however, if intuitions have an average reliability, one can reasonably conclude that accepting them is more conducive to understanding than deferring to others.

Assuming moral understanding as the most valuable goal of moral inquiry, one can also conclude that accepting moral intuition is better than deferring propositions from other subjects. From this, given the conditions of limited cognitive resources, it follows that accepting some moral intuitions is legitimate in absence of defeaters.

6. Concluding remarks

This chapter has discussed an original argument in favor of moral intuitionism: the argument from limited cognitive resources. The argument states that to the extent moral agents think under conditions of limited time, attention, and available information, they are committed to accepting some proposition without full reflective consideration. Then,

---

[40] I will defend this claim in Chapter 5.

I have argued that among the possibilities that require little cognitive effort, accepting moral intuitions is the most epistemically responsible choice because it is more conducive to moral understanding than the most plausible alternative (i.e., deferring to reliable agents).

The conclusion of the argument is consistent with intuitions' limited reliability. However, the issue is not considered as irrelevant but outweighed by the importance of other factors of moral understanding, such as engagement with reasons and the appreciation of moral propositions through genuine sentiments. For these reasons, the defense of intuitionism I have discussed is more promising than pure reliabilism and phenomenalism.

Admittedly, the claim that moral understanding is more valuable than knowledge is not uncontentious. Skeptics can reject it and argue that knowledge is more important than understanding or that there is not difference between knowledge and understanding.[41] However, this objection would lead to the odd conclusion that moral deference is permissible. Consequently, skeptics would need to explain such an oddity. Posing this puzzle to the skeptics is certainly a virtue of the argument from limited cognitive resources.

Another merit of the argument discussed in this chapter is that it explains an apparent asymmetry between how people rely on intuitions in the moral and nonnormative domains. Whereas appealing to intuitions is quite widespread concerning moral questions, there is something odd in relying on intuitions regarding descriptive propositions. The asymmetry is because deferring to others for moral propositions under conditions of limited resources is not usually considered a legitimate option; in contrast, it is considered desirable to defer to experts concerning descriptive questions.

Despite the aforementioned merits, the argument from limited resources depends on the empirical claim that accepting intuitions is more conducive to understanding than deference. Intuitionists need to shed more light on this issue to strengthen their position.

This chapter concludes my research on moral intuitionism. What I discussed cannot exhaust the broad topic of intuitionism; nor can it suffice to provide a full defense

---

[41] According to reductionist accounts, moral understanding just is the possession of a high degree of moral knowledge (Riaz 2015, Sliwa 2017).

of an intuitionist moral epistemology. However, I hope I pointed the way to defend intuitionism from the skeptical challenge deriving from recent empirical studies.

III

THE AUTOMATICITY CHALLENGE

Chapter 5

Caring, moral motivation, and automatic conduct

1. Introduction

In recent decades, moral psychology has undermined the widespread view of ethics as deliberative practice. Indeed, prominent research in this field has highlighted that many morally relevant decisions derive from automatic processes rather than reasoning (Damasio 1994, Narvaez and Lapsley 2005, Haidt 2001). Such empirical research posits a metaethical challenge: how can actions based on automatic processes be motivated by *moral reasons*?

In the present chapter, to address the *automaticity challenge*, I defend an account of moral motivation based on the concept of *caring*, which has been recently introduced by some authors in philosophy of mind and action (Shoemaker 2003, Jaworska 2007, Seidman 2009, 2016, Brownstein 2018, 101-122). However, the notion of caring characterized thus far is still sketchy and needs to be refined. To address this concern, I will develop a more detailed account of caring, and I will apply it to the moral domain. Specifically, I will employ my account of *moral caring* to provide a better explanation of automatic moral action compared with rival proposals (Snow 2006, Sauer 2012).

The plan of the chapter is the following. In Section 2, I introduce the automaticity challenge, and I disentangle a descriptive and a normative interpretation of it. In the following section (3), I highlight the most salient limitations of the solutions to the automaticity challenge adopted in the literature. In Section 4, I outline a general account of caring; I will discuss the internal link between caring, emotions, and practical reasons, as well as the differences between caring and other attitudes. Section 5 develops a caring-based account of moral motivation—how it explains the occurrence of automatic actions and mismatches between caring and beliefs. Then, I reply to some relevant objections (Section 6), and finally, I show that moral caring cannot be sufficient as a normative theory of moral sensitivity (Section 7).

2. The automaticity challenge

In Hitchcock's movie *Rope*, two young Harvard scholars strangle a former classmate to death just before a dinner party. The murder is an intellectual exercise since they want to prove they are able to commit "the perfect murder". The two characters were inspired by the ideas of their former schoolmaster Rupert Cadell, who is attending the party, too. Indeed, during dinner, Cadell coldly explains his quasi-Nietzschean view, according to which killing another human being for manifest intellectual superiority is legitimate and desirable. Nevertheless, once Cadell discovers what his students did, he is shocked and ashamed. Although perfectly consistent with his abstract ideas about murder, Cadell condemns the gratuitous homicide and calls the police.

Such a story dramatically tells us how automatic processes, such as emotions, can motivate people to commit certain moral actions, in contrast with what people consciously believe. Cadell, horrified by the crime, condemns his students, although he believes that murder is permissible. This case is similar to the conduct of Huckleberry Finn in Mark Twain's novel (Arpaly 2003, 9). At a key point in the novel, Huck helps his friend Jim escape from slavery, even though he does not think that this is the right thing to do from a deliberative standpoint. Therefore, according to Arpaly's interpretation, Huck has the intuition that helping Jim is the right thing to do, despite a moral judgment that is in tension with it.

Importantly, the automatic mental processes at play in Cadell and Huck's actions do not seem completely blind and impulsive but are based on some valid reasons. In other words, some mental processes, although automatic and relatively independent of deliberative thoughts, seem to be *responsive to moral reasons*. That constitutes a philosophical problem, to the extent that, in a widespread view of ethics, moral reasons depend on reasoning and deliberation. In this view, reason responsiveness is in tension with automaticity. Following Sauer (2012, 2017, 51-83), I call this problem *the automaticity challenge*.

The automaticity challenge presses philosophers and psychologists to explain how actions based on automatic processes (automatic actions, for brevity) can be responsive to reasons despite their relative independence of deliberative attitudes. This question cannot be ignored by moral theorists since a large amount of evidence shows that many moral decisions are based on automatic processes rather than explicit reasoning (Narvaez and Lapsley 2005, Haidt 2001).

The reason responsiveness of automatic actions admits two distinct interpretations, one descriptive and one normative. According to a first interpretation, an automatic action is responsive to moral reasons whenever the action is based on considerations that count as moral from the perspective of the agent, considering her traits and attitudes. In this meaning, "reason-responsive" is understood as "dependent on *motivating reasons"*. In this interpretation, the automaticity challenge demands a theory of moral motivation that explains how actions can be both automatic and intentionally moral.

According to a second interpretation, an automatic action is responsive to moral reasons whenever the action is based on considerations that count as morally valid according to some normative standards. In this meaning, "reason-responsive" is understood as "dependent on normative (i.e., good) reasons". Therefore, in this interpretation, the automaticity challenge demands a theory of *moral sensitivity* that explains how an agent can track normative reasons automatically without the mental effort of deliberation.

I will address the question of moral sensitivity in the next chapter. In the present chapter, I will consider the descriptive challenge, which is more basic: according to many moral theories, every action based on normative reasons (i.e., morally sensitive) must be motivated by those reasons; conversely, not every action based on moral motivations is sensitive. Therefore, the question I will address in this chapter concerns how actions can be automatic and based on moral motivations.

3. Influential solutions to the automaticity challenge

Other authors have addressed the automaticity challenge in recent years. Sauer (2012, 2017, 51-83) understands the automaticity challenge as a problem for the effectiveness of moral reasoning. In response to this problem, he shows that automatic processes (e.g., emotions) can be influenced and modified by reasoning and conscious beliefs through habituation. As Kahneman (2011) points out, cognitive operations can migrate from System 2 to System 1 if repeated over time. This is what happens in moral education, in which agents learn to follow rules resulting from reasoning and regard for rational rules becomes automatic through repeated practice. In this fashion, moral reasoning can be effective in automatic actions, and the latter can be based on moral reasons.

I find Sauer's solution to the automaticity challenge too intellectualistic. The explanation assumes that moral reasons require present or prior reasoning. This assumption is problematic (Arpaly and Schroeder 2012). In short, as Arpaly and Schroeder argue, reasoning is a mental act that can be favored by reasons. In other words, there are reasons for and against deliberating according to the circumstances. The reasons for deliberating cannot be tracked by further reasoning (prior or present); otherwise, there would be a regress. This suggests that reasons might be independent of reasoning.

Another influential solution to the automaticity challenge is Snow's account of habitual virtuous action (Snow 2006, 2010, 39-62). To explain how actions can be automatic and virtuous, Snow appeals to Bargh's concept of "goal-dependent automaticity" (Bargh 1992). According to such theory, certain mental goals (e.g., driving home, washing teeth) are "chronically accessible" by virtue of habituation and practice. Chronically accessible goals are those that can be promptly activated without being mentally represented at the time to act. Behavior consistent with a chronically held goal can be triggered by the relevant stimuli. This explains how actions can be automatic and rational (i.e., dependent on an agent's goal). In the moral domain, certain evaluative goals, such as the goal of equity in social exchanges or the commitment to truth, can be automatically activated and thus generate behavior consistent with moral reasons.

Note that the concept of goal-dependent automaticity is sufficient to distinguish automatic behavior motivated by moral reasons from mere habitual behavior in accordance with reasons. Suppose one has the goal of being patient for many years. Then, one abandons the goal but continues to do the same things by habit. The action, although in accordance with moral reasons, is no longer motivated by reasons because it is no longer dependent on the moral goal (Snow 2006, 555).

However, I argue that goal-dependent automaticity is insufficiently precise to explain genuine moral behavior. First, goals can be pursued instrumentally. For instance, suppose John has the goal of being a good citizen because he does not want any trouble with the law. John's behavior is in accordance with moral reasons, dependent on a moral goal, but not genuinely motivated by moral reasons. Second, goals can be "coldly" pursued, even non-instrumentally. Suppose a scientist implants the goal of being virtuous into an android. As a result, the android mimics virtuous behavior, and her actions are based on a proper moral goal, yet it is unlikely motivated by moral reasons. Moral motivation requires a certain emotional attachment or regard toward a moral goal, in other

words, a *sentiment*. The sentiment toward a moral end disposes the agent to have certain feelings toward morally salient facts. Feelings, in turn, occur in different morally relevant situations. It seems that Snow's account misses this important aspect of automatic moral action.

In the following sections, I will outline a *caring-based* account of moral motivation to address the automaticity challenge. The account I will defend attempts to address the limitations of the accounts discussed in this section. First, I reject the intellectualist assumption that moral reasons must depend on present or prior reasoning. I will assume here that normative reasons are facts or considerations that favor certain courses of action (Alvarez 2010, Arpaly and Schroeder 2012, Scalon 2014, Mantel 2018). Such facts can be grasped automatically without the need for any present or prior reflection. Accordingly, motivating reasons are *modes of grasping* normative facts (i.e., reasons) by agents through beliefs, desires, or nondeliberative attitudes (Mantel 2018). This assumption prevents the objection from regress and makes the automaticity challenge less demanding, to the extent that there is no necessity of establishing a causal link between an automatic action and an explicit episode of reasoning or an explicit transmission of a moral norm. Second, I will try to go beyond the notion of goal-dependent automaticity by providing a more precise characterization of the attitude of regard for moral ends, how it relates to emotions, and how it manifests itself in particular circumstances.

4. Introducing caring attitudes

The concept of caring has been introduced by several authors in philosophy of mind and action (Shoemaker 2003, Jaworska 2007, Seidman 2009, 2016, Brownstein 2018, 101-122). However, the accounts provided thus far are sketchy and insufficiently precise. In this section, I develop a more detailed account of caring. I will understand caring as a sentiment toward an object (4.1). Then, I will spell out the relationship between caring and practical reasons (4.2) and compare caring with other attitudes (4.3).

4.1 Caring as sentiment

In English, "caring about something or someone" means feeling that something or someone is important and worth worrying about (*Oxford Dictionary*). Synonyms of caring about could be "being concerned about" or "having regard for'' something or

someone. Slightly different from "caring about" is the concept of "caring for", which means looking after or taking care of someone who is in need. In the present discussion, I will refer to the concept of caring only as "caring about".

The definition of caring about just stated comprises two relevant aspects. First, the object of caring is characterized as something *salient* from the perspective of the subject who cares about it. For example, if Susan *cares about* Mark, this means that Mark's health and happiness *matter* to Susan. Second, the importance of the object of caring is *felt*, that is, caring entails some *emotional vulnerability* of the subject toward what she cares about. On this basis, following other authors (Shoemaker 2003, Jaworska 2007, Seidman 2009, 2016, Brownstein 2018, 101-122), one can define caring as an *emotional disposition* or *sentiment*:[42] if a subject cares about something, she is disposed to experience a wide range of emotional reactions before specific facts or considerations that the subject associates with the object of caring. The different emotional reactions are connected with and unified by the object of caring, which can manifest in various circumstances. Importantly, by connecting them with an object, the sentiment makes the different mental episodes meaningful.

Consider, for example, Susan, who cares about Mark. It is likely that Susan will be worried when Mark is in danger and relieved once Mark escapes danger. Susan's worry and relief make sense in those situations *because Susan cares about Mark*; it is exactly Susan's caring about Mark that explains why Susan is worried in that context and relieved afterward. In other words, Susan's caring about Mark connects those different emotional episodes and makes Susan's mental change meaningful.

Caring can be broadly classified as a type of *attitude*. In social psychology, attitudes denote preferences or evaluative orientations toward objects (Maio, Haddock and Verplanken 2019). Although some philosophers (Deonna and Teroni 2012, Brownstein 2018) refer to them as occurrent mental states, I understand attitudes in their most common meaning, that is, as *dispositions* or *traits*. So understood, an attitude is

[42] Like Prinz (2007), I understand the terms "emotional disposition" and "sentiment" as synonyms.

characterized by a *valence*, which denotes the type of evaluation (positive or negative) toward the object, and a certain *strength*, i.e., how resistant to change the attitude is.[43]

A caring attitude has a positive valence, since caring about means to consider something or someone as important and worth worrying about; in other words, the object is regarded as *good* by the subject. It is important to note that although caring has a positive valence, the emotions that stem from caring can be positive or negative, according to the specific context. For example, suppose Mark cares about his dog. Then, it is likely that he will feel angry if someone beats his dog or sad if the dog suffers from a disease. Caring is the disposition that unifies different emotional episodes, positive or negative, into a coherent picture.

Caring is a quite strong attitude. Indeed, sentiments, by definition, are resistant to change. For instance, they tend to persist in the face of reflective considerations. A change in sentiment typically requires the acquisition of a habit.[44] As I will argue later in this chapter (4.3), understanding how caring attitudes relate to change is helpful to distinguish them from other types of attitudes.

Identifying a caring attitude is not an easy task. Surely, an agent's manifest behavior is crucial to understand what she cares about. However, as sentiment, attributing caring requires some psychological introspection. To disentangle the various carings that an individual possesses, scientific tools such as lab experiments, in which it is possible to manipulate environmental factors, might be helpful. It is also important to consider that not every action can be interpreted as an unequivocal manifestation of caring; conflicting carings are possible, as well as conflicting interpretations of a subject's behavior.

4.2 Caring and practical reason

Caring, as I defined it, is a sentiment that manifests itself through emotions. However, it is worth noting that caring entertains an important relationship with practical reasons. To the extent that an object of caring is seen as important, it is plausible that the subject sees the object as a source of reasons to act (Seidman 2009, 285-286). If a subject cares about

---

[43] The notion of attitude strength is distinct from the concept of intuitive strength discussed in Chapter 1.

[44] Surely, carings can vary in their strength. That means that, in other words, there are stronger and weaker carings.

X, the subject will be disposed to see facts or considerations that favor the good or the ill of X as reasons for herself to act. For instance, the fact that Mark is in danger is seen by Susan as a reason to help Mark; the fact that today is Mark's birthday is seen by Susan as a reason to buy him a present.

At first glance, the connection between caring and reasons appears to be in tension with the emotional nature of caring attitudes. However, the tension is only apparent since emotions are tightly related to practical reasons. Recall the important relation of emotion with attention (Ch. 1). As stated, emotions help the subjects direct their attention toward significant objects in light of subjective goals and concerns. The emotional objects are nothing but *evaluative construals* or *appraisals* that signal on certain reasons to act for the subject. For example, by fear, Susan automatically realizes that Mark is in danger, and she immediately recognizes the presence of a reason to help Mark.

In sum, by being disposed to feel a wide range of emotions, caring favors certain appraisals of particular situations. The appraisals track salient facts and considerations related to the object of caring that favor certain courses of action. Given this framework, the claim that caring is connected with reasons is not in conflict with its being a sentiment but a direct consequence of that.

For present purposes, it is important to note that caring attitudes, through emotions, track reasons in an automatic mode. This is possible to the extent that reasons do not need to be explicitly represented to be detected (cf. Mantel 2018). Practical reasons can be recognized through their "indicators", that is, salient stimuli that a subject associates with a practical end. This is what occurs in the activation of a caring attitude, in which a situational cue triggers an emotional reaction, which in turn inclines the subject to act in a certain way.

4.3 Caring and other attitudes

In addition to its definition and its relationship with reasons, it might be helpful to compare caring with other attitudes and mental dispositions.

A caring attitude, as I understand it, is a more malleable and sophisticated disposition than mere instinct. Typically, innate instincts and impulses are highly resistant to self-regulation and changes in habits. In contrast, caring attitudes can be acquired or lost. It is possible, for example, to begin to care about philosophy at some moment in

one's life, and it is possible to stop caring about video games. It might be possible, in extreme circumstances, even to stop caring about one's own parents.

Although carings can change, they are distinguishable from deliberative attitudes, such as beliefs and desires. Compared with the latter, carings are more resistant to reflective considerations. Beginning to care about an object is not something that one can decide or conclude through reflection. Likely, it is something that *happens*, to some extent. However, the difference between carings and deliberative attitudes is not only in strength. Another striking difference is that the object of caring is not necessarily conscious and transparent as is the object of beliefs and desires. People can care about something while not being fully aware of what they care about or having a full understanding of the reasons why they care about something. This does not mean that caring attitudes are completely inaccessible, just that they are relatively independent of what people believe and intentionally pursue.

According to the interpretation I embrace, what a subject cares about manifests itself through her behavior and feelings, regardless of what she reports to care about or reflectively judges as important. However, I do not mean to overemphasize the actual discrepancy between people's deliberative attitudes and people's carings. Deliberative attitudes tend often to give rise, in the long run, to coherent caring attitudes, and carings are typically accompanied by coherent deliberative attitudes. However, as I will show later (5.3), mismatches between carings and deliberative attitudes are possible, and conceptually distinguishing the two kinds of attitudes is crucial to explain these phenomena. *How much* carings fit deliberative attitudes remains an important empirical question, which goes beyond the scope of this chapter.

Carings are distinguishable from mere habits. First, carings are dependent on goals (i.e., the good of the object of caring), while habits can be acquired and maintained independently of goals. Second, unlike caring-based behavior, habitual behavior can be cold and mechanical; habits may involve no emotional vulnerability toward the task at hand. Therefore, it is possible to acquire certain habits without caring about the activity in which one is engaged.

The relationship between caring and habits is complex. On the one hand, habits can generate new caring attitudes. For example, suppose that I do not care much about a colleague of mine. However, by meeting and chatting with him every day at work, I become more attached to him and begin to care about him more. On the other hand, caring

attitudes can produce certain habits. For example, if I care about my grandmother, I am motivated to acquire the habit of visiting her every weekend. Furthermore, habituation processes can be normatively beneficial for carings; by regulating emotions, habits can help the agent track reasons related to the object of caring more efficiently. Nevertheless, it is worth noting that not every habit is caring preserving and beneficial: certain repeated and routinized behaviors make the subject's emotional vulnerability decrease and the conduct more inflexible, which is less sensitive to salient stimuli. In the next chapter, I will show that skills acquisition is a type of habituation that is normatively beneficial and caring preserving.

To summarize, caring is a mental disposition more malleable and subject to change than instincts and impulses. However, it is more resistant to change than beliefs and conscious desires. Compared with the latter, caring attitudes have a less conscious and transparent object. Unlike mere habitual behavior, caring is dependent on goals. Finally, habits can preserve carings and be beneficial for them according to the circumstances.

5. A caring-based account of moral motivation

With the general framework on caring in mind, we are now in a position to explain how caring attitudes motivate automatic moral actions. To this concern, I will identify the type of caring that produces genuine moral motivations (5.1). Then, on that basis, I will provide an explanation of automatic moral action and how it can conflict with deliberative attitudes (5.2).

5.1 Caring and moral motivation

Scholars refer to caring attitudes for different purposes. For instance, Jaworska (2007), like Frankfurt (1999), stresses the importance of caring for practical identity. In contrast, Brownstein (2018) argues for a caring-based theory of moral responsibility. My aim here is more modest: in what follows, I intend to defend a caring-based account of *moral motivation*.

That caring attitudes motivate people is implicit in the general characterization of caring I have provided in the preceding section. Carings are sentiments and have an internal connection with emotions. As argued, emotions involve the appraisal of salient facts that constitute practical reasons for the subject. However, it is worth noting that

emotions are not just cognitions of reasons, but they help the subject experience the *practical relevance* of the recognized reasons and be inclined to act accordingly. As Brownstein (2018, 46) puts it, affective reactions make the subject aware of some *tension* and orient her behavior toward the alleviation of such tension. In other words, feelings are like signals that an agent receives to adjust her conduct. Indeed, emotions tend to generate what Frijda (2007, 33-34) has called *action tendencies* or *readiness*, that is, motivational states that orient behavior in response to emotional objects. For example, fear inclines the subject who experiences it to *move away* the fearful object, anger prepares to *move against* the object, sadness disposes to *be helpless*. Therefore, given its important connections with emotions, reasons, and action tendencies, caring is a plausible candidate to explain how people are motivated to act.

Arguably, people are moved by what they care about. This is not to say that people are motivated *only* by what they care about. Rather, I am just saying that attributing a caring attitude to a subject is a sufficient condition to attribute to her a motivation to promote the good of what she cares about. However, caring has a peculiar type of motivational power. Compared with the motivation provided by deliberative attitudes, caring-based motivation is more enduring and resistant to change. For instance, the motivation to pursue the good of what we care about tends to persist in the face of reflective considerations. Moreover, motivations based on carings tend to manifest themselves through stronger emotions compared with more superficial types of motivation. Finally, motivational states connected with carings are easier to activate. It is not necessary to reflectively consider the motivational relevance of a certain situation; recognizing certain salient stimuli is sufficient to automatically trigger a relevant action tendency.

It is worth noting that a caring attitude excludes instrumental motivation toward its object. Suppose John seems to care about being a good citizen just because he does not want any trouble with the law. However, from this, it follows that John does not truly care about being a good citizen, since if the goal of being a good citizen conflicted with the goal of not being punished by the law in some circumstances (e.g., in case of civil disobedience), John would not manifest any emotional vulnerability toward the particular case relevant to the goal of being a good citizen. Then, John just cares about not being punished by the law; being a good citizen is only an instrumental goal to that caring but

not a deeper caring attitude. Therefore, it is not possible to care about an object instrumentally.

Not every caring attitude provides *moral* motivations. To motivate moral actions, a subject must care about a *moral standard*. By standards, I mean "concepts of possible and usually desirable states, including ideals, expectations, values, and goals." (Gailliot, Mead e Baumeister 2008, 475). For example, people can be moved by *ideal* standards, which include hopes and aspirations, or *ought* standards, which instead comprise duties and obligations. Moral standards can concern the individual (e.g., standards of kindness or honesty), the community (e.g., standards of justice and equality), or interpersonal relationships (e.g., standards of friendship or parenthood). What standards are morally relevant is debatable.

The commitment to a certain standard is, in my view, the hallmark of moral caring. However, it is important to note that, according to the account of caring I have adopted, caring about a standard does not require the full consciousness of the standard that one cares about. It is possible to care about a moral standard while not explicitly endorsing the standard or not being aware of the principles governing the standard. In other words, it is possible, according to my account of caring, to implicitly commit oneself to a moral standard through behavior or feelings consistent with the standard.[45]

Caring about standards motivates people through the occurrence of *moral emotions*, such as anger, indignation, admiration, shame, guilt, and pride. Moral emotions activate whenever a subject detects a particular situation that violates or conforms to a moral standard that she cares about. Importantly, each kind of emotion involves a specific appraisal of the situation and certain motivations to act. For example, through *anger*, the subject evaluates a situation as offensive or unjust and is inclined to punish and retribute the one who is responsible for the wrongdoing; through *admiration*, a subject evaluates a certain person or action as admirable and is motivated to emulate the object of admiration (see Table 1). It is important to note that the attitude of caring about a moral standard makes these emotional episodes intelligible: something is an object of anger because it violates a certain standard; someone is admirable because she conforms to the standard.

---

[45] Here, my account is in line with Horgan and Timmon's *morphological rationalism*, according to which certain moral standards ("principles" in their own words) can guide the conduct without being represented by the agents (Horgan and Timmons 2007).

Emotion	Appraisal	Motivations
Anger	The situation is offensive or unfair	Motivations to punish, retribute the responsible for the wrongdoing (Ask and Pina 2011)
Disgust	The situation is intolerable	Motivations to express condemnation of the wrong behavior and to protect the social order (Tybur, et al. 2013, Kelly 2011)
Shame	I failed to meet some standards	Motivations to hide, withdraw, or disappear our social presence (Tangney and Dearing 2002)
Guilt	I am responsible of a wrong situation	Motivations to confess, apologize (Tangney and Dearing 2002), to remedy (cooperative behavior) (Ketelaar and Tung Au 2003)
Gratitude	An individual did a good thing that benefits me	Motivations to be thankful, to return the favor (Haidt 2003)
Admiration	An individual or an act is admirable	Motivations to elevate and emulate the one responsible for the situation (Haidt 2003)
Sympathy and compassion	Another individual is in an unfair situation	Motivations to help the other (altruism) (Batson 2014)

Table 1. How different types of moral emotions can favor certain appraisals and action tendencies (i.e., motivations).

5.2 From moral caring to automatic action

Thus far, I have shown how caring about moral standards can produce motivations to act. At this point, we have all the ingredients to explain how action can be automatic and motivated by moral attitudes.

The explanation is straightforward. A situation that involves some moral standard triggers a motivational state in a subject who cares about the standard. As argued, the motivational state can be understood as an emotional episode comprising an appraisal of

the situation and a certain action tendency. The motivational state, in turn, causes the subject's automatic response to the situation (Figure 5).

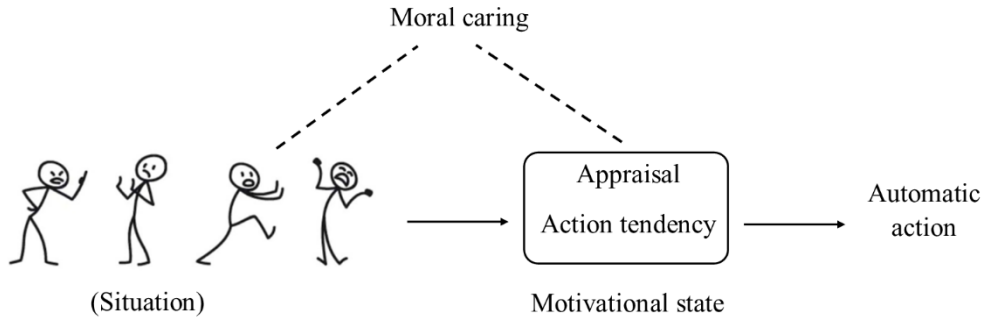

Figure 5. The caring-based explanation of automatic action.

In this explanation, moral caring is the mental disposition that links the morally relevant situation with the occurrence of the motivational state. Indeed, moral caring can be understood as a subjective disposition to generate a set of motivational states from a range of situations relevant to a moral standard.[46] A morally relevant situation triggers a corresponding motivational state, under suitable circumstances, in light of a caring attitude attributed to the agent. For example, suppose Mark observes some people throwing chemical waste into a lake; Mark feels angry about the situation and Mark's anger motivates him to protest. Mark's emotional reaction and behavior are intelligible *because Mark cares about living in a healthy environment*.

Note that caring is a disposition to produce motivations and, only indirectly, automatic actions. Caring does not directly cause actions for two reasons. First, the mediation of a proper motivational state is crucial to distinguish an action motivated by a standard from an action merely in accordance with a standard. Second, it is possible that a caring attitude manifests itself by the occurrence of an emotional state without generating a corresponding action. Therefore, moral caring always disposes the agents to perform automatic actions through the mediation of relevant motivational states.

As stated, automatic actions can conflict with an agent's deliberative beliefs and intentions. A caring-based theory of moral motivation can account for these common

[46] This is only another way to rephrase the definition of caring as emotional disposition, given the intrinsic connection between emotions and motivations.

cases by understanding them as mismatches between caring and moral beliefs. Let us consider some examples.

The first class of automatic actions that conflict with deliberative beliefs comprises those cases in which an agent does morally bad things, although she endorses good moral principles. For instance, currently, in Western countries, the majority of people tend to endorse anti-racist beliefs; however, discrimination and violence toward blacks are still widespread. This kind of racism is typically motivated by subtle feelings of distrust and fear toward black people rather than by explicit racist attitudes. Dovidio and Gaertner called this phenomenon "aversive racism":

> Because of current cultural values, most whites [in the US] have strong convictions concerning fairness, justice, and racial equality. However, because of a range of normal cognitive, motivational, and sociocultural processes that promote intergroup biases, most whites also develop some negative feelings toward or beliefs about blacks, of which they are unaware or which they try to dissociate from their nonprejudiced self-images. These negative feelings that aversive racists have toward blacks do not reflect open hostility or hatred. Instead, aversive racists' reactions may involve discomfort, uneasiness, disgust, and sometimes fear. (Dovidio and Gaertner 2004, 4)

According to my interpretation, aversive racists do not hold two inconsistent beliefs (that racism is wrong and that racism is right), but they possess conflicting types of attitudes. Aversive racists care about the preservation of the race, although they believe that racism is wrong. Racist caring activates with the sight of black people and manifests itself through aversive feelings toward blacks. Feelings, in turn, motivate racist behavior and discrimination.

Aversive racism is just an example among many kinds of implicit social biases that orient moral conduct independently of, and sometimes in tension with, moral beliefs. These automatic processes tend to predict moral behavior even more effectively than explicit beliefs (Frankish 2016). This fact underscores, once again, the importance of distinguishing between the moral standards that a person cares about and the moral standards she endorses.

In certain cases, automatic behavior responds to good reasons, despite bad moral beliefs. Formerly, I have mentioned the character of Rupert Cadell, who condemns the episode of gratuitous murder although he believes it is morally legitimate, and the case

of Huckleberry Finn, who helps Jim escape slavery although he believes that slavery is right. Arpaly has called such cases "inadvertent virtue" (Arpaly 2003, 8). In contrast to people moved by implicit biases, inadvertently virtuous people are motivated to act by caring toward a correct standard while having wrong moral beliefs. Therefore, their conduct constitutes another case of mismatch among caring and deliberative attitudes.

Another morally relevant example of good caring in tension with bad beliefs is the case of "outlaw emotions", i.e., emotions in tension with a dominant system of belief (Silva forthcoming, Jaggar 1989). Emotions of such kind are prevalent in people in conditions of oppression:

> People who experience conventionally unacceptable, or what I call 'outlaw' emotions often are subordinated individuals who pay a disproportionately high price for maintaining the status quo. The social situation of such people makes them unable to experience the conventionally prescribed emotions: for instance, people of color are more likely to experience anger than amusement when a racist joke is recounted, and women subjected to male sexual banter are less likely to be flattered than uncomfortable or even afraid. (Jaggar 1989, 166)

The motivational power of outlaw emotions is particularly important for oppressed people in a social environment in which certain kinds of moral beliefs are not even conceivable. For instance, the anger and grief of an oppressed woman who cares about her own dignity can still track valid moral reasons and push her to rebel, although the woman is not sufficiently educated to understand the reasons through reasoning.

6. Virtues of the theory, objections, and replies

Before considering some possible objections, let me summarize the most important benefits of the account I have presented in the previous sections. The most important virtue of a caring-based account of moral motivation is that it offers a satisfying explanation of automatic moral behavior that stands in tension with deliberative attitudes. In the preceding section, I have shown how the caring-based account can accommodate the various cases of mismatch among automatic behavior and deliberative attitudes. Importantly, the explanations I have outlined do not require a causal connection with a former episode of reasoning or explicit moral education. Furthermore, the concept of caring goes beyond the notion of goal-dependent automaticity by ruling out instrumental and "coldly" pursued goals. Therefore, the caring-based account of moral motivation

seems to provide a better response to the automaticity challenge than rival accounts.

Despite the acknowledged advantages, my account of moral motivation might be contested by some recent theories of moral motivation. I have defined moral caring as the attitude of caring about certain standards. Nonetheless, some authors (Darwall 2006, Rozeboom 2017) object that genuine moral attitudes are not constituted by concern for standards and reasons but rather by attitudes of regard toward other individuals. According to this view, which Rozeboom has called the "individual-first view", genuine moral motivations stem only from caring about the individuals involved in one's action.

The individual-first view contrasts with a "reasons-first view", according to which moral motivation is based on intentions to act for moral reasons (Rozeboom 2017). Against this account (and in support of the individual-first view), Rozeboom considers the case of Huckleberry Finn (5). Huck's good moral deed cannot be explained by his regard for moral reasons, since he appears to be convinced to do the wrong thing. Rather, what makes Huck's action good is his sympathetic concern for his friend Jim. Therefore, in contrast with my account, Huck's conduct is moved by moral motivation, although he does not care about moral standards or reasons.

Concerning the case of Huck, my reading is slightly different. According to the account of caring I have defended, caring about a moral standard does not require the awareness of the standard or the reasons that justify it. What matters is that an agent displays behavior and some feelings consistent with the standard. Therefore, in my interpretation, Huck does care about some standards concerning friendship or human rights, even though he does not *believe* that these standards are correct.

Assumed that, there is room for considering both the individual-first view and the reasons-first view as compatible with my account of caring. Specifically, we can understand the two views as two distinct normative explanations of *why people should care about moral standards*. According to the individual-first view, an agent should care about moral standards because the standards are grounded in attitudes of concern for other individuals. In contrast, according to the reasons-first view, moral standards are basic: the existence of moral reasons suffices as a reason for why people ought to care about them. Nonetheless, both normative accounts are consistent with the claim that moral action stems from carings about standards, in light of the conception of caring I have adopted. Therefore, so understood, the individual-first view does not appear to be in tension with my account of moral caring.

According to the view I have adopted, the attitude of caring about a moral standard is constituted by an emotional disposition or sentiment. This means that my account seems to be committed to a form of *moral sentimentalism*, that is, the view according to which moral judgment and motivation are mainly based on emotion rather than reasoning. While dominant in the 2000s (cf. Haidt 2001, Prinz 2007), moral sentimentalism has been attacked by a recent rationalist revival (Sauer 2017, May 2018).

It is not possible to discuss here all the rationalist criticisms raised to moral sentimentalism. However, let me point out that the account of caring I have outlined can meet many important rationalist demands. First, according to my account, caring-based motivations entail the recognition of some salient facts as reasons. Thus, motivational states depending on caring are not "brute feelings" but comprise a cognitive component consisting in an automatic inference from a represented situation to a subjective standard. Second, my account is silent concerning whether sentiments are the ultimate source for assessing the correctness of feelings and emotions. It is consistent with my account of caring that emotions and sentiments can be assessed by normative standards resulting from reasoning. Third, the claim that emotions can be influenced, shaped, regulated, and improved by reasoning is largely compatible with my view. Therefore, my sentimentalist account could respond to some crucial rationalist complaints.

One of the most important rationalist tenets I reject is the intellectualist claim that moral motivations depend on reasoning, and by reasoning, I mean conscious deliberation. However, not every rationalist account endorses such a narrow notion of reasoning. Those rationalists (cf. Horgan and Timmons 2007, May 2018) that understand reasoning more broadly as unconscious inference probably agree with my point.

Finally, a possible weakness of my account could lie in the fact that it depends on controversial assumptions in the philosophy of action and emotion such as a factive conception of reasons and a cognitivist view of emotions. If these two claims are false, the account I have defended cannot be true. In response to this doubt, which cannot be fully addressed here, I can only say that the two claims are not so contentious, even though not universally acknowledged. Indeed, several authors agree with the claim that reasons are constitutively independent of reasoning, and the fact that emotions do have a cognitive component is quite accepted in the empirical and philosophical research.

7. Moral caring and sensitivity

Having defended an account of moral caring and its role in the explanation of automatic action, it is worth discussing the relationship between caring and *moral sensitivity*. Call morally sensitive the action that is automatic and based on good reasons (not just on reasons that seem good from the perspective of the agent). Accordingly, moral sensitivity is the ability of a subject to produce automatic actions on the basis of normative reasons. Does moral caring suffice to have moral sensitivity? In this section, I consider and reject this hypothesis.

My account of caring aims to explain moral motivation but is normatively neutral. As shown in the previous sections, it is possible to care about correct moral standards, as well as incorrect standards. Plausibly, to act for a good moral reason, one has to care about a correct moral standard. Therefore, morally sensitive action requires a caring attitude toward a correct standard. It is possible to fail to act sensitively not just in the case of a lack of moral caring but also for caring about an incorrect standard. Assumed that, the question at stake is whether caring about a correct standard is sufficient to possess moral sensitivity.

The hypothesis that caring suffices for moral sensitivity is tempting, yet intuitively unconvincing, or so I argue. An interesting test is constituted by the behavior of people who have just experienced a moral conversion. If sincere, people affected by moral conversions, by virtue of revelatory moments, can develop deep moral caring in a relatively short time. A nice literary example is the moral conversion of Ebenezer Scrooge in Charles Dickens' tale *A Christmas Carol*. The old miser Scrooge, after receiving the visit of the ghost of his dead business partner and the ghosts of Christmas Past, Present, and Yet to Come, is transformed into a good man: he donates his money to charity, he takes care of a young ill boy, and he treats people with kindness and generosity. The tale seems to suggest that a deep change of heart is sufficient to become a virtuous person. However, I find this view quite unrealistic. We should consider that morality is inevitably intertwined with the knowledge of the particular, which requires experience and practice. Knowing what the right thing to do is in specific contexts entails familiarity with and competence regarding the social environment in which one acts beyond an appropriate attitude. People moved by sincere moral concern but without the appropriate competences might manifest naïve moral behavior. Therefore, in my view, sensitive actions require some degree of acquired competence, in addition to a correct motivation.

Moral sensitivity cannot be reducible to moral caring. As I will argue in the next chapter, moral sensitivity has a crucial motivational component, which is constituted by a caring attitude toward some standards. Nonetheless, moral sensitivity entails an epistemic component constituted by competence related to some domain of moral behavior. To clarify this point, it is helpful to consider the differences in moral behavior shown by some psychopaths and by some people affected by autism spectrum disorders (ASDs). Arguably, keeping distinct mere moral caring from sensitivity is important to account for the different moral conduct of psychopaths and people with ASDs.

As stated in Chapter 2, psychopathy is a mental disorder that affects the emotional, interpersonal, and behavioral components of a subject (Blair, Mitchell and Blair 2005, 7). The psychopathic individual tends to be characterized by impulsivity, conduct problems, and a callous-unemotional interpersonal style (8). Regardless of whether this is explained by dysfunction of the amygdala (Blair, Mitchell and Blair 2005) or by attentional deficits (Hamilton, Racer and Newman 2015), it is quite uncontentious that psychopaths' capacity to develop sentiments of regard or concern toward moral reasons is impaired. For this reason, it is plausible to assume that some people affected by psychopathy lack moral caring. The deficit in moral caring explains why the behavior of some psychopaths tends to be immoral and manifests a lack of regard before moral situations.

According to the literature, people affected by ASDs tend to be by some difficulties in social and communicative development, as well as by narrow interests and repetitive behavior (Baron-Cohen 2008). Like psychopaths, people with ASD have deficits in empathy. However, since their emotional dysfunction is less serious, ASD patients can develop some concern for moral standards (Kennett 2002). Some people affected by ASD, if put in conditions to understand the feelings and behaviors of other people, can manifest an emotional vulnerability toward morally relevant situations. Nevertheless, by virtue of their impaired social capacities, people with ASD manifest difficulties in applying their moral caring in ordinary contexts. As a result, the moral behavior of people with ASD is rarely virtuous: they tend to be inflexible because of obsessive regard for rules rather than paying attention to others' discomfort and suffering (McGeer 2008). Whereas psychopathic behavior tends to be characterized by a lack of moral caring, the moral behavior manifested by those with ASD is characterized by a lack of the necessary competence to correctly exercise moral caring. In other words, people with ASD lack *moral sensitivity*, even though their good moral intentions are intact.

Therefore, their behaviors might constitute a good case for distinguishing moral sensitivity from mere caring.

To summarize, moral caring can be understood as the motivational basis of moral sensitivity. Although it is an essential condition, caring about a correct moral standard is not sufficient to develop moral sensitivity and, thus, to have correct moral intuitions in specific contexts. In addition to moral caring, moral sensitivity requires some degree of competence. As I have suggested, moral competences can be acquired through experience and practice in the social environment in which one acts.

Chapter 6

Moral sensitivity as skillful automaticity

1. **Introduction**

Moral sensitivity is the subjective disposition to perform good actions from automatic mental processes, such as emotions or intuitions. This chapter outlines an empirically informed account of moral sensitivity based on recent research on skills and expertise (Ericsson 2008, Kahneman and Klein 2009, Stichter 2018, Fridland 2021). According to the view I will defend, moral sensitivity is a set of moral skills that can be acquired through a special habituation process called *deliberate practice*.

My account of moral sensitivity aims to vindicate the Aristotelian insight according to which the virtuous person possesses some kind of "perceptual knowledge" that enables her to do what particular situations require (Aristotle 2004, 1109b 23, McDowell 1998). The concept of skill, I will argue, suffices to explain how it is possible to reliably act well without the mental effort of deliberation.

Since Aristotle (2004, 1140b), there has been a widespread philosophical skepticism about the identification of moral sensitivity with skills. In particular, many authors have pointed out that the role of motivation constitutes a decisive disanalogy between virtuous moral conduct and skillful performance (Zagzebski 1996, 106-116, Rees and Webber 2014, Small 2021). In reply to this widespread objection, I will appeal to the caring-based account of motivation I have developed in the preceding chapter, and I will show how the acquisition, exercise, and possession of skills depend on caring attitudes.

The chapter proceeds as follows. In Section 2, I show that automatic mental processes can lead to either competent or incompetent decisions, which demands a psychological explanation of how competence can be acquired. In the following section (3), I briefly outline how the notion of moral sensitivity is meant in the history of moral philosophy, and I clarify how it is understood in the present work. Then, in Section 4, I introduce the concepts of skills and expertise according to the recent literature. On this basis, in Section 5, I show how skills fit the moral domain and explain sensitivity.

Subsequently, I consider the main objections against the claim of moral sensitivity as skills, that is, the motivation problem (Section 6) and other objections (Section 7). Finally, I state some concluding remarks regarding the relationship between moral caring and sensitivity (Section 8).

2. The two facets of automaticity

In his influential work about decision-making, Gary Klein reports the rescue operation of an expert firefighter he has interviewed:

> It is a simple house fire in a one-story house in a residential neighborhood. The fire is in the back, in the kitchen area. The lieutenant leads his hose crew into the building, to the back, to spray water on the fire, but the fire just roars back at them. "Odd," he thinks. The water should have more of an impact. They try dousing it again, and get the same results. They retreat a few steps to regroup. *Then the lieutenant starts to feel as if something is not right. He doesn't have any clues; he just doesn't feel right about being in that house*, so he orders his men out of the building—a perfectly standard building with nothing out of the ordinary. As soon as his men leave the building, the floor where they had been standing collapses. Had they still been inside, they would have plunged into the fire below. (Klein 1998, Example 4.1, my italics)

From the story, it emerges that the lieutenant has no clear-cut consciousness of the reasons why the situation represents a danger; he does not know that the house has a basement or that the fire is coming from downstairs. However, the firefighter sees that there is "something wrong". The flames are not reacting as he expected—it is too hot for a small fire in the kitchen, and this much heat should be generated by a larger fire. Thus, the lieutenant orders his men to leave the building. As Klein suggests, the commander's right call in this complex case is not a matter of luck: his intuition of the situation results from the experience and accumulated competence that the firefighter has acquired over the years.

Consider now another case:

> In February 1999, around midnight, four plain-clothes police officers were searching a Bronx, New York, neighborhood for a rape suspect. They saw Amadou Diallo, a 22-year-old West African immigrant, standing in the doorway of his apartment building. According to the police, Diallo resembled the suspect they were tracking.

> When they ordered him not to move, Diallo reached into his pants pocket. Believing he was reaching for a gun, the police fired a total of 41 shots, 19 of which hit and killed Diallo. Diallo was in fact unarmed. All four officers were later acquitted of any wrongdoing in the case. (Correll, et al. 2002, 1314)

Here, the competence of police officers' conduct is more disputable than that of expert firefighters. The officers' reaction is hasty, impulsive, and probably influenced by an implicit racial bias.[47] In other words, their behavior was the opposite of what the situation required.

The relationship between automaticity and competence is complex. The two decisions just described are both based on automatic processes, relatively independent of conscious beliefs; however, the results are in opposition to one another. The rescue operation of the firefighter is successful and extremely skillful. In contrast, the police officers' conduct is unsuccessful and possibly incompetent. As Brownstein points out (2018, 4), automaticity has two sides: on the one hand, automatic processes can lead to competent and virtuous behavior, but on the other hand, their outcomes can be impulsive and biased.

In the preceding chapter, I addressed the descriptive part of the automaticity challenge by arguing for a caring-based account of moral motivation. I also argued that caring about a moral standard is not sufficient to reliably track normative reasons. To achieve this result, two further normative conditions must be obtained: first, the agent must care about a correct standard to be disposed to respond to *good* reasons; second, the agent must develop sufficient competence through experience to *reliably* track the relevant reasons in particular situations. Whenever these two conditions are met, the agent is said to be *sensitive*, relatedly, to some moral standard. Unlike caring, moral sensitivity is a normative term, which I define as the disposition to perform competent moral actions automatically, without the mental effort of deliberation.

---

[47] Correll and colleagues' empirical evidence for the so-called "shooter bias" supports this hypothesis: by using simplified videogames, the experimenters show that people tend to shoot more quickly and accurately at Afro-American targets than at white targets. Importantly, explicit endorsements of racial biases do not explain the empirical results; the mere *knowledge* of the stereotype suffices to manifest the shooter bias (Correll, et al. 2002, 1325).

In this chapter, I will leave aside the normative question of what standard an agent should care about.[48] Rather, I will focus on the psychological conditions that make automatic moral behavior sensitive. Such conditions are crucial to distinguish a competent automatic action from an impulsive or biased action. To this concern, I will outline an empirically informed account of moral sensitivity.

3. Moral sensitivity in philosophical tradition

The idea of moral sensitivity (or "sensibility") has a long history in moral philosophy. Aristotle, in the VI book of the *Nicomachean Ethics*, stresses the importance of the knowledge of "particular things" for moral wisdom. Aristotle also points out that performing good actions in particular situations requires a kind of perceptual knowledge (Aristotle 2004, 1109b 23).

In the modern era, David Hume is one of the most influential advocates of "moral sense" as the perceptual capacity to distinguish good from evil (Hume 2007, SB470 23-29). In line with this idea, some contemporary Humean theorists have defended a view of moral sensitivity as a condition of the *existence* of moral properties (Prinz 2007, Lewis 1989, Wiggins 1987). According to this view, goodness, rightness, and wrongness do not exist without human sensitivity to feel appropriate emotions before particular events. However, it is not in this meaning that I understand moral sensitivity. Rather, by moral sensitivity, I mean a specific kind of practical knowledge, that is the competence required to recognize and respond to reasons in particular circumstances, without the mental effort of deliberation.

My conception of moral sensitivity is close to McDowell's theory of virtue. In his seminal article "Virtue and Reason", he takes kindness as an example to explain how sensitivity is at work in virtuous character:

> A kind person can be relied on to behave kindly when that is what the situation requires. Moreover, his reliably kind behaviour is not the outcome of a blind, non-rational habit or instinct […] Rather, that the situation requires a certain sort of behaviour is […] his reason for behaving in that way, on each of the relevant

[48] I will take for granted that *some* moral standards are correct. However, I will consider this point later (7).

> occasions. […] A kind person has a reliable sensitivity to a certain sort of requirement that situations impose on behaviour. The deliverances of a reliable sensitivity are cases of knowledge; and there are idioms according to which the sensitivity itself can appropriately be described as knowledge: a kind person knows what it is like to be confronted with a requirement of kindness. The sensitivity is, we might say, a sort of perceptual capacity. (McDowell 1998, 51)

In short, according to McDowell, moral sensitivity is the capacity to do what a particular situation requires, without the need to deliberate from moral principles, that is, in my terms, *automatically*. The sensitive person does not need to deliberate to act appropriately but knows the specific demands from a kind of reliable perception of the situation.

In the following sections, I will try to vindicate McDowell's insight, according to which it is possible to reliably act well without deliberation. Specifically, I will provide an account of moral sensitivity based on research on skills and expertise. My account is in line with some authors who appeal to the model of skill to explain virtue (Annas 2011, Stichter 2018, Fridland and Stichter 2020). However, my aim here is more modest: I do not intend to provide a theory of virtue, but rather of moral sensitivity, which is a necessary but insufficient condition to possess a virtuous character.[49]

4. Introducing skills

In recent decades, the concepts of skill and expertise have been subject to much attention in psychological (Ericsson, Hoffman, et al. 2018) and philosophical (Fridland and Pavese 2021) research. In particular, scholars have been interested in the automaticity of skillful performance in different domains, such as chess, sports, firefighting, or nursing.

In this section, I summarize the main results of the psychological and philosophical research on skillful performance. Specifically, I will focus on the process of skills acquisition (4.1) and the distinctive features of skillful automaticity (4.2).

4.1 Skills acquisition as deliberate practice

Complex performances, such as cooking, driving, or chess-playing, constitute *domains of*

[49] This leaves open the possibility to act virtuously in a deliberative mode.

expertise regulated by specific normative standards. Competent or excellent performances in a domain require the learning of a set of relevant practical abilities. For example, driving well requires the ability to start the car, change gears, turn left and right, etc. These abilities are called *skills*.[50]

It is largely documented that the acquisition (or learning) of skills requires a great amount of practice. However, it is also acknowledged that repeated practice alone is not sufficient. In fact, accumulated experience and successful performances in a task do not always go hand in hand (Ericsson 2018). Therefore, further conditions must occur to become skillful in a given domain.

A crucial condition to acquire a skill is the commitment of the subject toward the goals and standards that regulate the domain of the performance. Stichter has called such commitment "goal setting" (Stichter 2018, 11). For example, if one wants to learn to drive, one is committed to the goal of driving well and is aware of the standards that must be met to drive well. Once internalized, the normative standards enable the learner to evaluate her efforts to drive and improve accordingly. Importantly, goal setting in skill learning is not just desiring a general end but entails some degree of *planning*, i.e., defining the different subgoals and tasks necessary to achieve the end.

Another important condition concerns the subject's motivation to achieve the set goals and standards. Indeed, one who intends to acquire skills must exercise some performance *with the specific intention to improve*. In other words, improving must become part of the goals to which the learner is committed. Importantly, the will to improve cannot be a vague desire of "getting better" but a constant motivation to practice in various types of situations, the progression from simple tasks to more challenging tasks, and the implicit or explicit evaluation of the obtained results (Stichter 2018, 25).

A high level of performance cannot be achieved without being motivated to go beyond the first level of automaticity in a task. As Ericsson has documented:

> The key challenge for aspiring expert performers is to avoid the arrested development associated with automaticity. These individuals purposefully counteract tendencies toward automaticity by actively setting new goals and higher performance standards, which require them to increase speed, accuracy, and control

[50] As Kahneman (2011) points out, areas of expertise are not regulated by a single skill but entail a large collection of "miniskills". Therefore, it is preferable to refer to skills in the plural.

> over their actions […] The experts deliberately construct and seek out training situations to attain desired goals that exceed their current level of reliable performance. (Ericsson 2008, 991)

In short, the aspiring expert must be convinced that the process of skills acquisition is never ending, and one can always improve and refine the level of performance.

The last condition for acquiring expertise is the possibility of receiving useful feedback from the environment in which one learns. Feedback is crucial for the subject to understand whether her attempts of performance are going in the right or the wrong direction. However, not every environment allows the possibility to learn from feedback. There are "kind" and "wicked" learning structures (i.e., environments) according to the quality of feedback they provide (Hogarth 2001, 87-90).

In a kind learning structure, the subject receives accurate feedback. It is immediately clear when one acts well, as well as when one makes mistakes; in this way, it is possible to learn the appropriate lesson from experience. A kind learning structure is regular enough so that the subject can learn to recognize those cues that permit her to identify the nature of a situation on future occasions. Through time and practice, some patterns of recognition can be automatized, which favors the development of skilled intuitions (Kahneman and Klein 2009, Kahneman 2011). Chess playing, driving, or some professions, such as nursing and firefighting, are typical examples of kind environment (Kahneman 2011). These latter domains are sufficiently codified by certain rules that allow accurate predictions. As a result, an agent can become skillful by learning from prolonged practice.

In contrast, in a wicked learning structure, feedback can be misleading, and consequently, one tends to learn the wrong lesson from experience. Hogarth reports the example of an early twentieth-century physician who was thought to have infallible intuitions in diagnosing typhoid (Hogarth 2001, 85). The problem was that to make his diagnosis, the physician used to palpate the tongue of the patients without washing his hands before; consequently, what he deemed positive feedback was in fact irrelevant. Therefore, to acquire a skill, the quality of the feedback is crucial.

In sum, four conditions are required to learn a skill: (1) goal setting and planning, (2) strong motivation to improve, (3) receiving accurate feedback from experience, and (4) ample opportunity to practice. All of these conditions constitute what Ericsson has

called *deliberate practice* (Ericsson 2008, 2018). In deliberate practice, the quality of the practice matters just as much as the quantity (Stichter 2018, 24). Indeed, deliberate practice is not repetitive but *progressive*: it proceeds from simple to more demanding tasks, involves planning and evaluating the obtained level of performance and is never ending. Therefore, engaging in deliberate practice is crucial to reach a skillful and not repetitive nor mechanical performance in a domain of expertise.

Finally, two further aspects at play in deliberate practice need to be emphasized. The first is the role of metacognition, both declarative and procedural. In skill learning, procedural metacognition is crucial in internalizing the standard and favoring so a context-sensitive evaluation of the learner's efforts to improve. Explicit declarative metacognition, instead, is important to critically assess one's own level of performance and one's limitations.

The second important process at work in skill acquisition is the regulation of emotions. Throughout this research, I have emphasized the benefits of emotions in modulating attention and motivating behavior. However, it is widely accepted that emotions are not always helpful; for instance, emotions might be of the wrong intensity, duration, frequency, or type according to what a particular situation requires (Gross 2015, 4). Therefore, for various reasons, emotions can impede reaching some desired level of performance in many different domains. This means that acquiring a skill needs a process of emotion regulation.

Emotion regulation can be of two types (Gross 2015): one *downregulates* one's own emotions when one decreases the emotional engagement toward a certain object (e.g., try to calm oneself down when angry); in contrast, one *upregulates* the emotions when one increases one's own emotional engagement (e.g., firing oneself up before a big game). Skill learning can benefit from both types of emotion regulation, not just downregulation. For example, in sports, moderating anxiety and fear is important to provide good performance; nonetheless, at the same time, a high-level performance requires the right tension and anger. More generally, it is noteworthy that sustaining and enhancing emotional engagement is essential to maintain the right level of motivation to improve and exercise a skill. Therefore, emotions are not suppressed at all in skills learning, but rather *calibrated* to better accomplish the desired standards of performance.

4.2 Skillful automaticity

The role of skills in human conduct is to ensure a specific kind of behavior, which I call *skillful automaticity*. Such a way of acting requires a high level of expertise in a domain and, thus, many hours of deliberate practice. This is evident in sports, for example, in which athletes exhibit spontaneous successful performances as a result of enduring professional training.

Two main related features characterize skillful automaticity, unlike other automatic behaviors: *flexibility* and *control* (Fridland and Stichter 2020). The conduct of an expert performer (e.g., an expert driver, a good athlete, or a skillful firefighter) is *flexible* to the extent that it can achieve some desired goals in a variety of specific contexts. This is possible by virtue of developed attention to occasion-specific properties that suggest how a goal should be accomplished in a situation (Douskos 2019, 4319). Said otherwise, the expert performer knows how to pursue her goals flexibly, in a context-sensitive way, and typically without the effort of deliberation. Moreover, the skillful performance is *controlled*, although automatic. The type of control at play in skillful automaticity is not direct and simultaneous to action like the one in deliberative action. Rather, the skillful automatic action is controlled to the extent that it is dependent on prior plans and action schemas (*distal control*); moreover, if guided by a skill, an agent has the capacity to intervene to adjust her behavior or activate deliberation when required (*intervention control*).[51] This is possible by virtue of "control structures" (e.g., plans, action schemas, attentional and motor capacities) that an agent learns through deliberate practice (Fridland 2021). Control structures transform general intentions into successful actions in specific contexts. In this fashion, the agent can extend her control over her nondeliberative actions.

Flexibility and control are sufficient to distinguish skillful automaticity from other types of automatic conduct. The mere *habitual action*, for example, lacks sufficient attention toward the relevant situational features to achieve a goal flexibly. For instance, a driver who is guided by habit to follow a certain route to home will be inclined to follow that route, even though she knows that, on that day, a street she has to cross is blocked by a new building site. Neither are flexible *actions based on biases or stereotypes* since these latter are rough unprecise generalizations, which cannot be sensitive to specific features

[51] Here, as I argued in former chapters (1-2), metacognition is crucial to switch from the automatic to the deliberative mode of thinking.

of particular situations. The *impulsive action*, instead, lacks self-regulation to provide a response proportional to the specific demands of a situation; consequently, it escapes the distal control of the subject. Finally, some automatic actions are not skillful because they are *naïve*; in this case, the subject lacks sufficient experience to predict the outcomes of a particular situation. In sum, all these kinds of automatic actions (habitual, impulsive, biased, and naïve), for different reasons, lack flexibility and control.

Skillful automaticity has many benefits for the conduct of a person. Through the acquisition of skills, an agent is capable of making competent decisions under time pressure when it is not possible to deliberate. Moreover, the possession of skills frees up attention and cognitive resources to focus on more complex goals. This favors multitasking performances: for instance, the expert driver can change gears and make turns, while at the same time focusing on the route to reach the desired destination. Finally, spontaneity of action is another important valuable outcome of skills acquisition. In many domains, such as sports, excellent performances are not only successful but also spontaneous: the author does not need to think too much before executing the performance.

To summarize, rather than resulting from mere repetition, the acquisition of a skill is the outcome of a deliberate practice involving goal setting, planning, motivation to improve, and learning from feedback. The acquisition of some level of expertise favors a peculiar kind of automatic conduct, which I have defined as skillful automaticity. This latter is characterized by flexibility and control. Therefore, humans' capacity to learn skills explains how actions can be automatic, yet extremely competent and functional to some valuable goals.

5. Moral sensitivity and skills

In the previous chapter, I argued that having the right motivation toward a correct standard is not enough to act sensitively: the agent also needs experience and competence. As some authors have stressed, there are several connection points between moral virtues and skillful behavior (Annas 2011, Stichter 2018). On this basis, in the present section, I employ the model of skills outlined in the preceding section to provide an empirically informed account of moral sensitivity.

My claim is not one of analogy but identity: I argue that moral sensitivity relatedly to some standard *just is* a set of acquired skills relevant for that standard. This means that

there are *moral skills*, aside from nonmoral skills such as driving skills and sport skills. A skill is moral whenever it is regulated by a moral standard (e.g., being a good citizen, living in a healthy environment, being kind, etc.). Thus, according to the claim I shall defend, moral standards constitute domains of expertise, which require the learning of moral skills.

Acquiring moral sensitivity is relevantly comparable to the process of skills learning. Consider, for example, a person who wants to be a *good parent*. In this will, the person is committed to some normative standards of parenthood and feels motivated to achieve them. In the process of becoming a good parent, the agent will regulate her behavior to accomplish specific tasks, such as teaching social rules to the children, bringing them to school, helping with their homework, and taking care of them when they are ill. Most likely, in the first stages, the parent will stick to deliberative plans, and she needs to think before executing simple tasks. However, after a significant amount of experience, the agent may acquire sufficient sensitivity to recognize the children's needs by intuition. At this point, the parent's attention is free to focus on the more complex challenges that the growth of the children provides every day.

It is noteworthy that becoming a morally sensitive person, like skills acquisition, does not require passive habituation but deliberate practice, by which the agent structures her general aim into different strategic subgoals, proceeds from simple to complex tasks, and assesses her improvements. As in skill acquisition, feedback is crucial for the development of moral sensitivity (Stichter, 2018, 67). For instance, a sensitive parent should be attentive to whether her children look satisfied or unsatisfied to adjust the acquired routines to their needs. Therefore, skills acquisition and moral sensitivity learning are very close in this respect.

As in skillful behavior, metacognition is pivotal in moral sensitivity. In previous chapters (1-2), I have highlighted how calibrating the right level of confidence according to the context is crucial to have correct moral intuitions. If a situation is complex, a less confident moral intuition is necessary to activate deliberative processes. For example, in Section 2, I quoted the story of Amadou Diallo who was instinctively killed by the police, although he was unarmed. Perhaps if the police officers were less confident about their perception of the situation, the tragedy would have been prevented. Therefore, the morally sensitive person, like other kinds of expertise, has to feel, through the appropriate

metacognitive feelings of doubt, when to switch from the automatic to the deliberative mode of thinking.

Another important connection point between skills and moral sensitivity is the role of emotion regulation. As shown in the present research, moral emotions are crucial in modulating attention toward salient facts and providing motivational force to moral beliefs. However, it is acknowledged that emotions are not always beneficial for moral knowledge; for instance, they can track morally irrelevant properties or impede genuine reasoning. For these reasons, moral sensitivity does require regulation of emotions (plausibly, both down- and upregulation) to make them efficient in serving certain normative standards.[52]

Finally, it is noteworthy that the conduct of a morally sensitive agent shares the same features of flexibility and control with expert performers in other nonmoral domains. As McDowell has argued, the kind person, for example, is attentive to the needs of other people in particular situations and does kind deeds spontaneously, without the effort of deliberation. In other words, I would say, the kind person accomplishes some internalized moral standards flexibly, according to the context, and by exercising some distal and intervention control.

In sum, for all these reasons, it seems natural to identify the acquisition of moral sensitivity as a process of skills learning and the exercise of moral sensitivity as a skillful automatic performance. Perhaps this might be considered a "default position". However, since Aristotle (2004, 1140b), there has been a widespread philosophical skepticism about the identification of moral sensitivity with skills. Thus, the claim of moral sensitivity as a set of skills needs to be defended by some objections that point out some supposedly decisive differences between the moral domain and nonmoral domains of expertise.

6. The motivation problem and the role of caring in skills

As mentioned in the preceding chapter, it is widely acknowledged that good moral action must be based on good motivation. In other words, to act for moral reasons, and not just in accordance with them, the agent must be motivated by those reasons. This means that motivation toward correct ends plays a constitutive role in moral conduct.

---

[52] See Helion and Ochsner (2018) for a review of the role of emotion regulation in moral judgment.

Compared with moral behavior, motivation seems to play a less important role in nonmoral domains of performance, in which skillfulness appears to be independent of the motives for which one acts. For example, being indifferent or less than wholehearted toward the ends of a domain of performance does not count against one's being skillful. A popular example, mentioned by Woodcock (2021, 582), is the case of tennis player Andre Agassi, whose autobiography has confessed his hatred for the sport since his early career. However, this aspect does not seem to affect the fact that Agassi was a great tennis player. Moreover, one can be extremely skillful yet motivated by ends external to the domain of performance. For instance, an excellent doctor can be motivated by money rather than the goals of medicine.

In short, at first glance, motivation plays a more important role in moral conduct than in skillful behavior. According to several authors (e.g., Zagzebski 1996, 106-116, Small 2021, Rees and Webber 2014), this constitutes a decisive disanalogy between the moral and nonmoral domains of performance. Such disanalogy concerning motivation can be explained by different standpoints: according to an Aristotelian explanation, motivation is more important in ethics because in skills the end is extrinsic, while in moral action the end is internal; a Kantian, instead, could stress the fact that morality is based on categorical imperatives, whereas nonmoral domains of expertise are based on hypothetical imperatives; finally, reasons-first approaches to ethics could argue that motivation plays a different role to the extent that in ethics, reasons are central, while just successful outcomes matter for skillful performances. All these lines of argument converge on the claim that moral sensitivity importantly differs from skillful performances by virtue of the different weights that motivation has in the two domains. Call this the *motivation problem*.

Recently, Stichter (2016, 2018, 93-117) provided insightful replies to the motivation problem. In short, Stichter points out that the objection from motivation is based on a misleading, although widespread, conception of skills and expertise, according to which skills have just instrumental value. Rather, according to a more robust conception of practical expertise that he defends, which is close to the one outlined here, skills do have an internal connection with the ends of the domain of performance; accordingly, not just successful outcomes matter in skillful behavior but regard for reasons related to the domain of expertise as well (see also Birondo 2021). Therefore, Stichter argues that right motivations in skills are as important as in moral conduct.

To refine Stichter's reply to the motivation problem, on which I substantially agree, it is helpful to consider the concept of caring that I have developed in the preceding chapter. Caring, as argued, is a strong and enduring motivational attitude in relation to some object felt as important. Caring disposes the agent to have motivational states (i.e., emotions) toward situations related to the object of caring. On this basis, I argue that caring about the standards of a domain of performance plays an important role in the *acquisition*, *exercise*, and *possession* of skills. Importantly, if caring is constitutive of skillful performance, there is no disanalogy between moral sensitivity and skills concerning the role of motivation.

Consider skills learning. As argued, becoming skillful in a domain requires a constant motivation to practice and improve. To proceed from simple to more challenging tasks, the agent must be sufficiently emotionally vulnerable to the results she obtains; for instance, a bad result should motivate her to repeat the task, whereas the satisfaction of a good result should motivate her to move to a more challenging task. Surely, external incentives can contribute to motivating the learning process, but they are often not enough to reach an excellent level of performance. Indeed, excellent performers have developed deep *caring* about the standards and values of their discipline. Cases such as Agassi, who become skillful despite the hatred for the discipline, are possible but quite rare. Therefore, it is reasonable to state that a caring attitude toward the standards of a domain favors the acquisition of skills.

Caring-based motivation also seems important for a correct exercise of skills in many domains. Previously, I mentioned that emotions are important in nonmoral domains to modulate the agent's attention toward relevant situational properties, as well as to put the agent in the right tension to perform successfully. Caring about the standards can favor the occurrence of such emotional and motivational states. Indeed, caring-based emotions are often crucial to make a performance less mechanical and habitual. Of course, caring attitudes must be regulated to be more functional to the success of the performance. However, this is also the case in the moral domain.

If caring is relevant for the acquisition, maintenance, and exercise of skills, this means that there is causal dependence between caring and skills. More contentious is the claim that the *possession* of skills depends in part on caring. To argue that, one must show that caring about the standards of a discipline is relevant for assessing skillfulness in that

discipline. However, this does not seem to be the case: as mentioned, the skillfulness of a performance appears to depend only on its successful outcome.

As Stichter points out, motivations become relevant once we switch the evaluation from the single performance to the character of the performer (Stichter 2018, 105). A performer who is not motivated by the ends of the domain of performance can be criticized for not being reliably responsive to the reasons of the domain. For instance, a doctor who does not care about the ends of medicine but just about the money is not an ideal doctor, although she is successful in surgical operations. Such a doctor, for example, will tend to treat her patients as a means to her career, to recommend expensive but unnecessary medical procedures, or to avoid giving her best if underpaid; in short, she is less trustworthy than a doctor who cares about the principles of medicine. Similarly, a tennis player, such as Agassi, who does not care about his profession will tend to have an unprofessional lifestyle and thus can be criticized for not being faithful to the values of the sport. Therefore, caring about the ends of a domain (e.g., medicine or sport) counts as a reason to positively evaluate a performer in that domain; conversely, not caring about the ends of the domain counts as a reason against the goodness of the performer. This means, in other words, that caring is relevant for the possession of skills, although it is not all. There is a normative dependence between skills and caring, beyond a causal dependence.

Whether good outcomes matter more than good motivations is debated in moral philosophy. However, importantly, parallel debates seem to be at play in many nonmoral domains of expertise. Medicine is the perfect example: whether successfully healing patients is more important than being faithful to medical deontology is an open question. Likewise, not everyone agrees that successful results in sports are more important than playing for fun and reciprocal respect. This suggests that the weight of motivation might vary according to the domain of expertise and to the different views one has of the domain. Therefore, the disanalogy between moral and nonmoral domains of expertise concerning motivation does not obtain if we consider this aspect.

To summarize, the motivation problem challenges the view that identifies moral sensitivity with skills. However, the motivation problem dissolves to the extent that nonmoral skills depend on caring attitudes toward the internal ends of the domain of expertise. Specifically, caring is relevant for the acquisition, exercise, and possession of

skills. Like moral sensitivity, nonmoral skills require the right combination of caring and competence.

7. Other objections

Having defended the claim of moral sensitivity as skills from the motivation problem, in this section, I will consider some further objections.

A challenging objection points out that the moral domain cannot satisfy the conditions for developing expertise through deliberate practice (Alfano 2021, 551). Moral life, the objection goes, might be a wicked environment. First, there is no wide consensus on what the standards of good action are; thus, the correctness of a moral action can be assessed from many different perspectives. Second, even if one can agree on the normative standards, moral life does not appear sufficiently regular to provide accurate feedback. Indeed, moral feedback is slow: it can take days, weeks, or decades to know whether one's moral decision was the right one to make. Furthermore, moral feedback is not unequivocal: one often receives mixed feedback in response to many moral decisions. Given such suboptimal conditions, the objection concludes that moral skills cannot be developed.

In response to such objection, I offer three considerations. First, one must consider that, as mentioned, disagreement about the standards of good performance is also present in some nonmoral domains of expertise (e.g., sport, medicine). Second, it is possible to divide the moral domain into as many areas of expertise as one deems necessary. The more a domain of action is restricted, the easier the conditions for intuitive expertise are met. In a restricted domain of moral conduct, it is more likely for subjects to agree on normative standards; for instance, there is much more consensus on what counts as being a good parent than on what counts as being a good person. Moreover, the more restricted a domain is, the greater the domain is codifiable and feedback is accurate. Therefore, if one is skeptical about the regularity of moral life, one could still conceive moral sensitivity as domain-specific; for example, sensitivity in parenthood, in nursing ethics, in business ethics, etc. In contrast, according to a unified view, cross-domain moral skills

do exist, and it is possible to speak about *moral expertise* as a single domain of skills.[53] My account of moral sensitivity is neutral regarding these two options. Moreover, my account is also silent about how moral sensitivity is better learned, whether separately or within particular domains of action. Third, and finally, the role of artificial learning structures in favor of moral skills learning should not be underestimated. For example, in ethics classes, it is possible to simulate moral situations, discuss their possible solutions, and codify moral behavior. Thus, in sum, these three considerations provide some license for optimism concerning the feasibility of moral skills.

A radical objection could complain that my account of moral sensitivity depends on a controversial metaethical assumption such as the existence of correct moral standards. If there are no correct moral standards, as error theorists and emotivists suggest, no improvement in moral sensitivity can be assessed. Therefore, the existence of moral sensitivity seems to be committed to some objectivist metaethics. In response to this complaint, I can only admit that my account is not compatible with moral skepticism (as stated in the introduction of this research). Nevertheless, it is worth noting that the account of moral sensitivity I have defended is consistent with many different accounts of moral objectivity (naturalism, nonnaturalism, constructivism, expressivism).

8. Concluding remarks: the continuum from caring to sensitivity

The automaticity challenge, in its normative meaning, questions how actions can be morally sensitive though based on automatic mental processes. In this chapter, I have shown that agents can become morally sensitive by learning a set of skills related to some ethical domain. On this basis, I have argued that moral sensitivity regarding some standard is just a set of learned skills related to that standard. Like nonmoral skills, moral skills involve a deliberate practice including goal setting, planning, motivation to improve, and learning from feedback.

In the preceding chapter, I argued that caring about a correct moral standard does not suffice to reliably perform good actions in particular situations. To achieve the latter, moral sensitivity is required. While moral caring is a subjective disposition to have

[53] This view is tight up to the Aristotelian idea of *practical wisdom*. For a defense of the view of practical wisdom as ethical expertise, see De Caro et al. (2018). For a recent discussion about the psychological groundwork of practical wisdom, see De Caro and Vaccarezza (eds.) (2021).

appropriate motivational states, moral sensitivity disposes a subject to perform good actions from automatic processes. Skills enrich good caring by providing the necessary practical competence to develop insightful moral intuitions and, consequently, sensitive actions. Therefore, skills learning can be considered a type of habituation process that is normatively beneficial for caring by ensuring a reliable transition from motivations to successful actions. Importantly, throughout this chapter, I have emphasized, in contrast with the widespread view, that skills depend on caring attitudes from different standpoints. This means that the learning and exercise of skills are caring preserving.

Such considerations suggest that moral caring and sensitivity are not independent dispositions, but that there is a *continuum* between the two. On the one hand, caring is the motivational basis of moral sensitivity and puts the agent in a position to develop the relevant skills. On the other hand, moral skills regulate caring-based motivations by making them more controlled and less naïve or impulsive. Therefore, moral skills can be seen as a *rational development* of a caring attitude.

Conclusion

This dissertation has defended different claims. In the first part, I have argued for a metacognitive account of moral intuition, according to which moral intuitions are automatic mental states characterized by a substantial degree of confidence. As argued in chapter 2, this account clarifies the role of intuition in moral reasoning, at the interface between type 1 and type 2 processes. I have shown that the function of moral intuition is not just heuristic, but can be sensitive to counterevidence, thus favoring the activation of reflection. My account of moral reasoning contributes to the idea that automatic and reflective processes are not conflicting but tend to cooperate in the moral domain. How moral reasoning works at the social level constitutes an interesting future line of research in moral psychology. I have shown that studying the social dimension of moral reasoning can clarify how moral reasoning can override biased intuitions, thus favoring moral progress.

In the second part, I have addressed the most relevant empirical challenges to moral intuitionism, that is the view according to which accepting moral intuitions is epistemically justified. I have highlighted two possible strategies to defend moral intuitionism. First, intuitionists can appeal to the role of confidence in regulating the level of credibility that moral reasoners assign to moral intuition. In this way, moral reasoners can protect themselves from cognitive biases and irrelevant factors. Second, accepting moral intuitions might be legitimate in virtue of the conditions of limited cognitive resources under which moral reasoners must think. However, both arguments need further research to be developed. As concerns the first argument, it is unclear how intuitive confidence can be reliable, given the existence of several metacognitive biases. Therefore, intuitionists need to provide more robust evidence to vindicate the reliability of moral intuitions. As regards the argument from limited cognitive resources, intuitionists need to better clarify the relationship between intuitions and moral understanding.

In the third part, I have addressed the automaticity challenge to moral action. I have pointed out that the challenge comprises two different problems. The first one concerns how actions can be morally motivated by automatic mental processes. In response to such a problem, I have defended a caring-based account of moral motivation.

The second challenge demands an explanation of how moral action can be sensitive (i.e., competent) although based on automatic processes. For this concern, I have argued for an account of moral sensitivity based on the concept of skillful automaticity. An interesting topic I could not address but would deserve further attention is the question of responsibility for automatic actions. The topic is too big to be seriously discussed in the present work. However, my account of moral sensitivity could be the groundwork for a theory of moral responsibility for automatic actions, to the extent that the possibility to exert control over automatic actions constitutes an important requirement for attribution of responsibility.

Bibliography

Aharoni, E., W. Sinnott-Armstrong, and K.A. Kiehl. 2014. “What’s wrong? Moral understanding in psychopathic offenders.” *Journal of Research in Personality* 53: 175–181.

Alfano, M. 2021. “Comments on Stichter’s The Skillfulness of Virtue.” *Ethical Theory and Moral Practice* 24: 549-554.

Alter, A.L., and D.M. Oppenheimer. 2009. “Uniting the Tribes of Fluency to Form a Metacognitive Nation.” *Personality and Social Psychology Review* 13 (3): 219-235.

Andow, J. 2016. “Reliable but not home free? What framing effects mean for intuitions.” *Philosophical Psychology* 29 (6): 904-911.

Annas, J. 2011. *Intelligent Virtue*. Oxford: Oxford University Press.

Aristotle. 2004. *Nicomachean Ethics*. Edited by Roger Crisp. Cambridge: Cambridge University Press.

Arpaly, N. 2003. *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford: Oxford University Press.

Arpaly, N., and T. Schroeder. 2012. “Deliberation and Acting for Reasons.” *The Philosophical Review* 121 (2): 209-239.

Ask, K., and A. Pina. 2011. “On Being Angry and Punitive: How Anger Alters Perception of Criminal Intent.” *Social Psychological and Personality Science* 2 (5): 494-499.

Audi, R. 1985. “Rationalization and Rationality.” *Synthese* 65: 159-184.

—. 2004. *The Good in the Right: A Theory of Intuition and Intrinsic Value*. Princeton: Princeton University Press.

—. 2015. “Intuition and Its Place in Ethics.” *Journal of the American Philosophical Association* 1 (1): 57–77.

Bago, B., and W. De Neys. 2019. “The Intuitive Greater Good: Testing the Corrective Dual Process Model of Moral Cognition.” *Journal of Experimental Psychology: General* 148 (10): 1782-1801.

Bargh, J.A. 1992. “The Ecology of Automaticity: Toward Establishing the Conditions Needed to Produce Automatic Processing Effects.” *The American Journal of Psychology* 105 (2): 181-199.

Baron-Cohen, S. 2008. *Autism and Asperger Syndrome*. Oxford: Oxford University Press.

Batson, D. 2014. “Empathy-Induced Altruism and Motivation.” In *Empathy and Morality*, edited by H. Maibom, 41-59. Oxford: Oxford University Press.

Bedke, M.S. 2008. “What They Are, What They Are Not, and How They Justify.” *American Philosophical Quarterly* 45 (3): 253-269.

Benbaji, H. 2013. “How is Recalcitrant Emotion Possible?” *Australasian Journal of Philosophy* 91 (3): 577-599.

Bengson, J. 2013. “Experimental Attacks on Intuitions and Answers.” *Philosophy and Phenomenological Research* 86 (3): 495-532.

—. 2015. “The Intellectual Given.” Mind 124 (495): 707-760.

Bengson, J., T. Cuneo, and R. Shafer-Landau. 2020. “Trusting Moral Intuitions.” *Nous* 54 (4): 956-984.

Bialek, M., and W. De Neys. 2016. “Conflict Detection During Moral Decision-Making: Evidence.” *Journal of Cognitive Psychology* 28 (5): 631–639.

—. 2017. “Dual processes and moral conflict: Evidence for deontological reasoners’ intuitive utilitarian.” *Judgment and Decision Making* 12: 148-167.

Birondo, N. 2021. “Aristotle and Expertise: Ideas on the Skillfulness of Virtue.” Ethical Theory and Moral Practice 24: 599-609.

Blair, J. 1995. “A cognitive developmental approach to morality: investigating the psychopath.” *Cognition* 57: 1-29.

Blair, J., D. Mitchell, and K. Blair. 2005. *The Psychopath: Emotion and the Brain*. Malden, MA: Blackwell.

Boyd, K. 2020. “Moral Understanding and Cooperative Testimony.” *Canadian Journal of Philosophy* 50 (1): 18-33.

Brady, M. 2009. “The irrationality of recalcitrant emotions.” *Philosophical Studies* 145: 413–430.

—. 2013. *Emotional Insight: The Epistemic Role of Emotional Experience*. Oxford: Oxford University Press.

Brownstein, M. 2018. *The Implicit Mind: Cognitive Architecture, the Self, and Ethics*. Oxford: Oxford University Press.

Callahan, L.F. 2018. “Moral Testimony: A Re-conceived Understanding Explanation.” *The Philosophical Quarterly* 68 (272): 437-459.

Cecchini, D. 2021. “Dual-Process Reflective Equilibrium: Rethinking the Interplay between Intuition and Reflection in Moral Reasoning.” *Philosophical Explorations* 24 (3): 295-311.

—. 2022. “Moral intuition, strength, and metacognition.” *Philosophical Psychology* 1-25. https://doi.org/10.1080/09515089.2022.2027356.

Chudnoff, E. 2013. *Intuition*. Oxford: Oxford University Press.

Cima, M., F. Tonnaer, and D.M. Hauser. 2010. “Psychopaths Know Right from Wrong but Don't Care.” *Scan* 5: 59-67.

Clavien, C., and C. FitzGerald. 2017. “The Evolution of Moral Intuitions and Their Feeling of Rightness.” In *The Routledge Handbook of Evolution and Philosophy*, edited by R. Joyce, 309-321. New York: Routledge.

Conway, P., and B. Gawronski. 2013. “Deontological and Utilitarian Inclinations in Moral Decision Making: A Process Dissociation Approach.” *Journal of Personality and Social Psychology* 104 (2): 216–235.

Correll, J., B. Park, M.C. Judd, and B. Wittenbrink. 2002. “The Police Officer’s Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals.” *Journal of Personality and Social Psychology* 83 (6): 1314–1329.

Cowan, R. 2013. “Clarifying Ethical Intuitionism.” *European Journal of Philosophy* 23 (4): 1097–1116.

Craigie, J. 2011. “Thinking and feeling: Moral deliberation in a dual-process framework.” *Philosophical Psychology* 24 (1): 53-71.

Cushman, F. 2013. “Action, Outcome, and Value: A Dual-System Framework for Morality.” *Personality and Social Psychology Review* 17 (3): 273–292.

—. 2020. “Rationalization is Rational.” *Behavioral and Brain Sciences* 43 (28): 1–59.

Cushman, F., L. Young, and M. Hauser. 2006. “The Role of Conscious Reasoning and Intuition in Moral Judgment.” *Psychological Science* 17: 1082-1089.

D’Cruz, J. 2015. “Rationalization, Evidence, and Pretense.” *Ratio* 28: 318-331.

Damasio, A. 1994. *Descartes’ Error*. New York: Putnam.

Damm, L. 2010. “Emotions and Moral Agency.” *Philosophical Explorations* 13 (3): 275–292.

Daniels, N. 2016. “Reflective Equilibrium.” In *The Stanford Encyclopedia of Philosophy*, edited by E.N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2020/entries/reflectiveequilibrium/.

Darwall, S. 2006. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA: Harvard University Press.

De Caro, M., and M. S. Vaccarezza (eds.). 2021. *Practical Wisdom. Philosophical and Psychological Perspectives*. Abingdon: Routledge.

De Caro, M., M. S. Vaccarezza, and A. Niccoli. 2018. “Phronesis as Ethical Expertise: Naturalism of Second Nature and the Unity of Virtue.” *The Journal of Value Inquiry* 52: 287–305.

De Neys, W. 2018. “Bias, Conflict, and Fast Logic: Towards a Hybrid Dual Process Future?” In *Dual Process Theory 2.0*, edited by W. De Neys, 47–65. Abingdon: Routledge.

—. 2020. “Rational rationalization and System 2.” *Behavioral and Brain Sciences* 43 (28): 22.

Decety, J., and S. Cacioppo. 2012. “The speed of morality: a high-density electrical neuroimaging study.” *Journal of Neurophysiology* 108: 3068–3072.

Deffendi, P., and C. Regeni. 2020. *Giulio fa cose*. Milano: Feltrinelli.

Demaree-Cotton, J. 2016. “Do Framing Effects Make Moral Intuitions Unreliable?” *Philosophical Psychology* 29 (1): 1-22.

Deonna, J., and F. Teroni. 2012. *The Emotions: A Philosophical Introduction*. Abingdon: Routledge.

Douskos, C. 2019. “The spontaneousness of skill and the impulsivity of habit.” *Synthese* 196: 4305–4328.

Dovidio, G.F., and S.L. Gaertner. 2004. “Aversive Racism.” In *Advances in Experimental Social Psychology, vol. 36*, edited by M.P. Zanna, 1-52. Amsterdam: Elsevier Academic Press.

Dunning, D. 2011. “The Dunning-Kruger Effect: On Being Ignorant of One's Own Ignorance.” In *Advances in Experimental Social Psychology, vol. 44*, edited by M.P. Zanna and J.M. Olson, 247-296. San Diego: Elsevier.

Efklides, A. 2006. “Metacognition and affect: What can metacognitive experiences tell us about the learning process?” *Educational Research Review* 1: 3-14.

Epstein, S. 2010. “Demystifying Intuition: What It Is, What It Does, and How It Does It.” *Psychological Inquiry* 21 (4): 295-312.

Epstein, S., R. Pacini, V. Denes-Raj, and H. Heier. 1996. “Individual Differences in Intuitive-Experiential and Analytical-Rational Thinking Styles.” *Journal of Personality and Social Psychology* 71 (2): 390-405.

Ericsson, K.A. 2008. “Deliberate Practice and Acquisition of Expert Performance: A General Overview.” *Academic emergency medicine* 15 (11): 988-994.

Ericsson, K.A. 2018. “The Differential Influence of Experience, Practice, and Deliberate Practice on the Development of Superior Individual Performance of Experts.” In *The Cambridge Handbook of Expertise and Expert Performance*, edited by K.A. Ericsson, R.R. Hoffman, A. Kozbelt and A.M. Williams, 685-705. Cambridge: Cambridge University Press.

Ericsson, K.A., R.R. Hoffman, A. Kozbelt, and A.M. Williams (eds.). 2018. *The Cambridge Handbook of Expertise and Expert Performance.* Cambridge: Cambridge University Press.

Evans, J. 2019. “Reflections on Reflection: the Nature and Function of Type 2 Processes in Dual-Process Theories of Reasoning.” *Thinking and Reasoning* 25 (4): 383–415.

Evans, J., and K. Stanovich. 2013. “Dual-Process Theories of Higher Cognition: Advancing the Debate.” *Perspectives on Psychological Science* 8 (3): 223-241.

Farsides, T., P. Sparks, and D. Jessop. 2018. “Self-reported Reasons for Moral Decisions.” *Thinking and Reasoning* 24 (1): 1-20.

Feinberg, M., R. Willer, O. Antonenko, and O. John. 2012. “Liberating Reason From the Passions: Overriding Intuitionist Moral Judgments Through Emotion Reappraisal.” *Psychological Science* 23 (7): 788-795.

Festinger, L. 1962. *A Theory of Cognitive Dissonance, vol. 2*. Stanford: Stanford University Press.

Fine, C. 2006. “Is the Emotional Dog Wagging its Rational Tail, or Chasing It?” *Philosophical Explorations* 9 (1): 83-98.

Fletcher, G. 2016. “Moral Testimony: Once More with Feeling.” In *Oxford Studies in Metaethics, vol. 11*, edited by R. Shafer-Landau, 45-73. Oxford: Oxford University Press.

Frankfurt, H. 1999. “On Caring.” In *Necessity, Volition, and Love*, 155-180. Cambridge: Cambridge University Press.

Frankish, K. 2016. “Playing Double: Implicit Bias, Dual Level and Self Control.” In *Implicit Bias and Philosophy, vol. 1*, edited by M. Brownstein and S. Saul, 23-46. Oxford: Oxford University Press.

Frederick, S. 2005. “Cognitive refl ection and decision making.” *The Journal of Economic* 19: 25–42.

Fridland, E. 2021. “The nature of skill: functions and control structures.” In *The Routledge Handbook of Philosophy of Skill and Expertise*, edited by E. Fridland and C. Pavese, 245-257. Abingdon: Routledge.

Fridland, E., and C. Pavese (eds.). 2021. *The Routledge Handbook of Philosophy of Skill and Expertise*. Abingdon: Routledge.

Fridland, E., and M. Stichter. 2020. “It just feels right: An account of expert intuition.” *Synthese* 199 (1-2): 1327-1346.

Frijda, N.H. 2007. *The Laws of Emotion*. New York: Routledge.

Gailliot, M., N. Mead, and R. Baumeister. 2008. “Self-Regulation.” In *Handbook of Personality, 3rd edition*, by O. P. John, R. W. Robins and L. A (eds.) Pervin, 472-491. New York: The Guilford Press.

Gangemi, A., S. Bourgeois-Gironde, and F. Mancini. 2015. “Feelings of error in reasoning—in search of a phenomenon.” *Thinking and Reasoning* 21 (4): 383-396.

Gantman, A.P., and J.J. van Bavel. 2015. “Moral Perception.” *Trends in Cognitive Science* 19 (11): 631-633.

Gigerenzer, G. 2008. “Moral Intuition = Fast and Frugal Heuristics?” In *Moral Psychology, vol. 2*, edited by W. Sinnott-Armstrong, 1-26. Cambridge, MA: MIT Press.

Glenn, A., A. Raine, R.A. Schug, L. Young, and M. Hauser. 2009. “Increased DLPFC Activity During Moral Decision-Making in Psychopathy.” *Molecular Psychiatry* 14: 909–911.

Goupil, L., and S. Kouider. 2019. “Developing a Reflective Mind: From Core Metacognition to Explicit Self-Reflection.” *Current Directions in Psychological Science* 28 (4): 403-408.

Greene, J. 2008. “The Secret Joke of Kant’s Soul.” In Moral Psychology vol. 3, edited by W. Sinnott-Armstrong, 35–80. Cambridge, MA: The MIT Press.

—. 2013. *Moral Tribes: Emotions, Reason, and The Gap Between Us and Them*. New York: The Penguin Press.

—. 2014. “Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics.” *Ethics* 124: 695-726.

Greene, J., B. Sommerville, L. Nystrom, J. Darley, and J. Cohen. 2001. “An fMRI Investigation of Emotional Engagement in Moral Judgment.” *Science* 293 (5537): 2105-2108.

Greene, J., F. Cushman, L. Stewart, K. Lowenberg, L. Nystrom, and J. Cohen. 2009. “Pushing Moral Buttons: The Interaction Between Personal Force and Intention in Moral Judgment.” *Cognition* 111 (3): 364–371.

Greene, J., S. Morelli, K. Lowenberg, L. Nystrom, and J. Cohen. 2008. “Cognitive Load Selectively Interferes with Utilitarian Moral Judgment.” *Cognition* 107: 1144–1154.

Gross, J.J. 2015. “Emotion Regulation: Current Status and Future Prospects.” *Psychological Inquiry* 26 (1): 1-26.

Gürcay, B., and J. Baron. 2017. “Challenges for the Sequential Two-system Model of Reasoning.” *Thinking and Reasoning* 23 (1): 49-80.

Haidt, J. 2001. “The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment.” *Psychological Review* 108 (4): 814-834.

—. 2003. “The Moral Emotions.” In *Handbook of Affective Sciences*, edited by R. J. Davidson, K.R. Scherer and H.H. Goldsmith, 852-870. Oxford: Oxford University Press.

Hamilton, R.K.B., K.H. Racer, and J.P. Newman. 2015. “Impaired Integration in Psychopathy: A Unified Theory of Psychopathic Dysfunction.” *Psychological Review* 122 (4): 770-791.

Hare, R.D. 1993. *Without Conscience: The Disturbing World of the Psychopaths among us*. New York: Pocket Books.

Hauser, M., F. Cushman, L. Young, K. Jin, and J. Mikhail. 2007. “A Dissociation Between Moral Judgments and Justifications.” *Mind and Language* 22 (1): 1-21.

Helion, C., and D. Pizarro. 2015. “Beyond Dual-Processes: The Interplay of Reason and Emotion in Moral Judgment.” In *Springer Handbook for Neuroethics*, edited by J. Clausen and N. Levy, 109-125. Springer Reference.

Helion, C., and K.N. Ochsner. 2018. “The Role of Emotion Regulation in Moral Judgment.” *Neuroethics* 11: 297-308.

Heyes, C., D. Bang, N. Shea, C.D. Frith, and S.M. Fleming. 2020. “Knowing Ourselves Together: The Cultural Origins of Metacognition.” *Trends in Cognitive Sciences* 24 (5): 349-362.

Hills, A. 2009. “Moral Testimony and Moral Epistemology.” *Ethics* 120: 94-127.

—. 2016. “Understanding Why.” *Nous* 50 (4): 661-688.

Hogarth, R. 2001. *Educating Intuition*. Chicago: University of Chicago Press.

Horgan, T., and M. Timmons. 2007. “Morphological Rationalism and the Psychology of Moral Judgment.” *Ethical Theory and Moral Practice* 10: 279–295.

Howard, N.R. 2018. “Sentimentalism about Moral Understanding.” *Ethical Theory and Moral Practice* 21: 1065-1078.

Howell, R.J. 2014. “Google Morals, Virtue, and the Asymmetry of Deference.” *Nous* 48 (3): 389–415.

Huebner, B., S. Dwyer, and M. Hauser. 2008. “The role of emotion in moral psychology.” *Trends in Cognitive Science* 13 (1): 1-6.

Huemer, M. 2005. *Ethical Intuitionism*. HoundmiUs, Basingstoke: Palgrave Macmillan.

Hume, D. 2007. *A Treatise of Human Nature*. Edited by D.F. Norton. Oxford: Clarendon Press.

Jaggar, J.M. 1989. “Love and knowledge: Emotion in feminist epistemology.” *Inquiry* 32(2): 151-176.

Jaworska, A. 2007. “Caring and Internality.” *Philosophy and Phenomenological Research* 74 (3): 529-568.

Johnson, D.D.P., and J.H. Fowler. 2011. “The evolution of overconfidence.” *Nature* 477: 317-320.

Kahane, G., K. Wiech, N. Shackel, M. Farias, J. Savulescu, and I. Tracey. 2012. “The Neural Basis of Intuitive and Counterintuitive Moral Judgment.” *Scan* 7: 393-402.

Kahneman, D. 2011. *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kahneman, D., and G. Klein. 2009. "Conditions for Intuitive Expertise: A Failure to Disagree." *American Psychologist* 64 (6): 515-526.

Kahneman, D., P. Slovic, and A. Tversky (eds.). 1982. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Kauppinen, A. 2013. "A Humean theory of moral intuition." *Canadian Journal of Philosophy* 43 (3): 360–381.

Kelly, D. 2011. *Yuck! The Nature and Moral Significance of Disgust*. Cambridge, MA: The MIT Press.

Kelly, T., and S. McGrath. 2010. "Is Reflective Equilibrium Enough?" *Philosophical Perspectives* 24: 325-359.

Kennett, J. 2002. "Autism, Empathy and Moral Agency." *Philosophical Quarterly* 52 (208): 340-357.

Kennett, J. 2006. "Do Psychopaths Really Threaten Moral Rationalism?" *Philosophical Explorations* 9 (1): 69–82.

Kennett, J., and C. Fine. 2008. "Internalism and the Evidence from Psychopaths and 'Acquired Sociopaths'." In *Moral Psychology, vol. 3* edited by W., edited by W. Sinnott-Armstrong, 173–190. Cambridge, MA: The MIT Press.

Ketelaar, T., and W. Tung Au. 2003. "The effects of feelings of guilt on the behaviour of uncooperative individuals in repeated social bargaining games: An affect-as-information interpretation of the role of emotion in social interaction." *Cognition and Emotion* 17 (3): 429-453.

Klein, G. 1998. *Sources of Power: How People Make Decisions*. Cambridge, MA: The MIT Press.

Koenigs, M., L., Adolphs, R. Young, D. Tranel, F. Cushman, M. Hauser, and A. Damasio. 2007. "Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgments." *Nature* 446: 908-911.

Koenigs, M., M. Kruepke, J. Zeier, and J.P. Newman. 2012. "Utilitarian Moral Judgment in Psychopathy." *Scan* 7: 708–714.

Kohlberg, L. 1981. *The Philosophy of Moral Development. Essays on Moral Development, vol. 1*. San Francisco: Harper and Row.

Koop, G. 2013. "An Assessment of the Temporal Dynamics of Moral Decisions." *Judgment and Decision Making* 8 (5): 527–539.

Koriat, A. 2007. “Metacognition and Consciousness.” In *The Cambridge Handbook of Consciousness*, edited by P.D. Zelazo, M. Moscovitch and E. Thompson, 289-325. Oxford: Oxford University Press.

Korsgaard, C. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.

Landy, J.F., and G.P. Geoffrey. 2015. “Evidence, Does Incidental Disgust Amplify Moral Judgment? A Meta-Analytic Review of Experimental.” *Perspectives on Psychological Science* 10 (4): 518-536.

Lewis, D. 1989. “Dispositional Theories of Value, II.” *Proceedings of the Aristotelian Society, supplementary volumes* 63: 113-137.

Lewis, M. 2020a. “The New Puzzle of Moral Deference.” Canadian Journal of Philosophy 50(4): 460-476.

—. 2020b. “A Defense of the Very Idea of Moral Deference Pessimism.” *Philosophical Studies* 177: 2323-2340.

Liao, S.M. 2008. “A defense of intuitions.” *Philosophical Studies* 140: 247–262.

Maiese, M. 2014. “Moral Cognition, Affect, and Psychopathy.” *Philosophical Psychology* 27 (6): 807-828.

Maio, G.R., G. Haddock, and B. Verplanken, . 2019. *The Psychology of Attitudes and Attitude Change*. Los Angeles: Sage.

Mantel, S. 2018. *Determined by Reasons: A Competence Account of Acting for a Normative Reason*. New York: Routledge.

Mata, A. 2019. “Social Metacognition in Moral Judgment: Decisional Conflict Promotes Perspective Taking.” *Journal of Personality and Social Psychology* 117 (6): 1061–1082.

May, J. 2018. *Regard for Reason in the Moral Mind*. Oxford: Oxford University Press.

McDowell, J. 1998. “Virtue and Reason.” In *Mind, Value, and Reality*, 50-76. Cambridge, MA: Harvard University Press.

McGeer, V. 2008. “Varieties of Moral Agency: Lessons from Autism (and Psychopathy).” In *Moral Psychology, vol. 3*: *The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, by edited W. Sinnott-Armstrong, 227-258. Cambridge, MA: The MIT Press.

McGrath, S. 2020. “Reflective Equilibrium, Its Virtues and Its Limits.” In *Moral Knowledge*, 11-58. Oxford: Oxford University Press.

Mercier, H., and D. Sperber. 2017. *The Enigma of Reason*. Cambridge, MA: Harvard University Press.

Metcalfe, J. 2008. "Evolution of Metacognition." In *Handbook of Memory and Metamemory*, edited by J. Dunlosky and R. Bjork, 29-46. New York: Taylor and Francis.

Moore, G. E. 1903. *Principia Ethica*. 1993. Edited by T. Baldwin. Cambridge: Cambridge University Press.

Moors, A. 2016. "Automaticity: Componential, Causal, and Mechanistic Explanations." *Annual Review of Psychology* 67: 263-287.

Narvaez, D., and D. Lapsley. 2005. "The psychological foundations of everyday morality and moral expertise." In *Character Psychology and Character Education*, edited by D. Lapsley and C. Power, 140-165. Notre Dame, IN: University of Notre Dame Press.

Nichols, S. 2004. *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford: Oxford University Press.

Nichols, S., and R. Mallon. 2006. "Moral dilemmas and moral rules." *Cognition* 100: 530–542.

Paxton, J., L. Ungar, and J. Greene. 2012. "Reflection and Reasoning in Moral Judgment." *Cognitive Science* 36: 163-177.

Pizarro, D., and P. Bloom. 2003. "The Intelligence of the Moral Intuitions: Comment on Haidt." *Psychological Review* 110 (1): 193-196.

Prinz, J. 2007. The Emotional Construction of Morals. Oxford: Oxford University Press.

—. 2015. "Naturalizing Metaethics." *Open Mind* 30 (T): 1-27.

Pritchard, D. 2010. "Knowledge and Understanding." In *The Nature and Value of Knowledge: Three Investigations*, by D. Pritchard, A. Millar and A. Haddock, 1-88. Oxford: Oxford University Press.

Proust, J. 2013. *The Philosophy of Metacognition: Mental Agency and Self-Awareness*. Oxford: Oxford University Press.

Railton, P. 2014. "The Affective Dog and Its Rational Tale: Intuition and Attunement." *Ethics* 124: 813-859.

Rees, C.F., and J. Webber. 2014. “Automaticity in virtuous action.” In *The Philosophy and Psychology of Character and Happiness*, edited by N. Snow and F. Trivigno, 75-90. Abingdon: Routledge.

Rehren, P., and W. Sinnott-Armstrong. 2021. “Moral framing effects within subjects.” *Philosophical Psychology* 34 (5): 611–636.

Riaz, A. 2015. “Moral Understanding and Knowledge.” *Philosophical Studies* 172: 113-128.

Rini, R. 2016. “Debunking debunking: a regress challenge for psychological threats to moral judgment.” *Philosophical Studies* 173: 675–697.

Rosas, A., and D. Aguilar-Pardo. 2019. “Extreme Time-Pressure Reveals Utilitarian Intuitions in Sacrificial Dilemmas.” *Thinking and Reasoning* 26 (4): 534–551.

Ross, W. D. 1930. *The Right and the Good*. 2002. Edited by P. Stratton-Lake. Oxford: Clarendon Press.

Royzman, E.B., K. Kim, and R.F. Leeman. 2015. “The curious tale of Julie and Mark: Unraveling the moral dumbfounding effect.” *Judgment and Decision Making* 10 (4): 296-313.

Rozeboom, G. 2017. “The Motives for Moral Credit.” *Journal of Ethics and Social Philosophy* 11(3): 1-29.

Sauer, H. 2012. “Educated intuitions. Automaticity and rationality in moral judgment.” *Philosophical Explorations* 15 (3): 255-275.

—. 2017. Moral Judgments as Educated Intuitions. Cambridge, MA: MIT Press.

—. 2018. *Debunking Arguments in Ethics*. Cambridge: Cambridge University Press.

—. 2019. *Moral Thinking, Fast and Slow*. Abingdon: Routledge.

Scanlon, T.M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Scherer, K.R., and A. Moors. 2019. “The Emotion Process: Event Appraisal and Component Differentiation.” *Annual Reviews* 70: 719–745.

Schnall, S., J. Benton, and S. Harvey. 2008. “With a Clean Conscience: Cleanliness Reduces the Severity of Moral Judgments.” *Psychological Science* 19 (12): 1219-1222.

Schnall, S., J. Haidt, G. Clore, and A. Jordan. 2008. “Disgust as Embodied Moral Judgment.” *Personality and Social Psychology Bulletin* 34 (8): 1096-1109.

Schwitzgebel, E., and F. Cushman. 2012. “Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers.” *Mind and Language* 27 (2): 135-153.

Segall, M.H., D.T. Campbell, and M.J. Hesrskovits. 1963. “Cultural Differences in the Perception of Geometric Illusions.” *Science* 139: 767–771.

Seidman, J. 2016. “The unity of caring and the rationality of emotion.” *Philosophical Studies* 173: 2785-280.

Seidman, J. 2009. “Valuing and Caring.” *Theoria* 75: 272–303.

Seligman, M.E.P., and M. Kahana. 2009. “Unpacking Intuition: A Conjecture.” *Perspectives on Psychological Science* 4 (4): 399-402.

Shafer-Landau, R. 2008. “Defending Ethical Intuitionism.” In *Moral Psychology, vol. 2: The Cognitive Science of Morality: Intuition and Diversity,* edited by W. Sinnott-Armstrong, 83-96. Cambridge, MA: The MIT Press.

Shoemaker, D. 2003. “Caring, Identification, and Agency.” *Ethics* 114: 88-118.

Sidgwick, H. 1874. *The Methods of Ethics*. 7th edition. London: Macmillan.

Silva, L.L. forthcoming. “The Epistemic Role of Outlaw Emotions.” *Ergo* 1-39.

Simon, H.A. 1992. “What is an 'Explanation' of Behavior?” *Psychological Science* 3 (3): 150-161.

Sinnott-Armstrong, W. 2008. “Framing Moral Intuitions.” In *Moral Psychology, vol. 2,* edited by W. Sinnott-Armstrong, 47-76. Cambridge, MA: The MIT Press.

Sinnott-Armstrong, W., L. Young, and F. Cushman. 2010. “Moral Intuitions as Heuristics.” In *The moral psychology handbook,* edited by M. Doris, 246-272. Oxford: Oxford University Press.

Sliwa, P. 2012. “In Defense of Moral Testimony.” *Philosophical Studies* 158: 175-195.

—. 2017. “Moral Understanding as Knowing Right from Wrong.” *Ethics* 127: 521–552.

Slovic, P., and D. Västfjäll. 2010. “Affect, Moral Intuition, and Risk.” *Psychological Inquiry* 21 (4): 387-398.

Small, W. 2021. “The Intelligence of Virtue and Skill.” *The Journal of Value Inquiry* 55: 229-249.

Snow, N. 2006. “Habitual Virtuous Actions and Automaticity.” *Ethical Theory and Moral Practice* 9 (5): 545-561.

—. 2010. *Virtue as Social Intelligence: An Empirically Grounded Theory*. New York: Routledge.

Stanovich, K.E. 2018. “Miserliness in Human Cognition: The Interaction of Detection, Override and Mindware.” *Thinking and Reasoning* 24 (4): 423-444.

Sterelny, K., and B. Fraser. 2016. “Evolution and Moral Realism.” *British Journal of the Philosophy of Science* 68 (4): 981-1006.

Stichter, M. 2016. “Practical Skills and Practical Wisdom in Virtue.” *Australasian Journal of Philosophy* 94 (3): 435-448.

—. 2018. *The Skillfulness of Virtue. Cambridge*: Cambridge University Press.

Stratton-Lake, P., ed. 2002. *Ethical Intuitionism: Re-Evaluations*. Oxford: Oxford University Press.

Sturgeon, N.L. 2002. “Ethical Intuitionism and Ethical Naturalism.” In *Ethical Intuitionism: Re-evaluations,* edited by P. Stratton-Lake, 184-211. Oxford: Clarendon Press.

Summers, J.S. 2017. “Post hoc ergo propter hoc: some benefits of rationalization.” *Philosophical Explorations* 20 (sup19): 21-36.

Sunstein, C. 2005. “Moral Heuristics.” *Behavioural and Brain Sciences* 28: 531-573.

Suter, R. S., and R. Hertwig. 2011. “Time and Moral Judgment.” *Cognition* 119: 454–458.

Tangney, J.P., and R.L. Dearing. 2002. *Shame and Guilt*. New York: The Guilford Press.

Thompson, V., and K. Morsanyi. 2012. “Analytic thinking: do you feel like it?” *Mind and Society* 11: 93–105.

Thompson, V., J.P. Turner, and G. Pennycook. 2011. “Intuition, Reason and Metacognition.” *Cognitive Psychology* 63: 107-140.

Thomson, J.J. 1985. “The Trolley Problem.” *The Yale Law Journal* 94 (6): 1395–1415.

Timmons, M. 1999. *Morality without Foundations*. Oxford: Oxford University Press.

Tinghög, G., D. Andersson, C. Bonn, M. Johannesson, Kirchler M., Koppel, L., and D. Västfjäll. 2016. “Intuition and Moral Decision Making– The Effect of Time Pressure and Cognitive Load on Moral Judgment and Altruistic Behavior.” *PLoS ONE* 11 (10): 1–19.

Trémolière, B., and J. Bonnefon. 2014. “Efficient Kill–Save Ratios Ease Up the Cognitive Demands on Counterintuitive Moral Utilitarianism.” *Personality and Social Psychology Bulletin* 40 (7): 923–930.

Trémolière, B., De Neys, W., and J. Bonnefon. 2012. “Mortality Salience and Morality: Thinking About Death Makes People Less Utilitarian.” *Cognition* 124: 379–384.

Turiel, E. 1983. *The Development of Social Knowledge: Morality and Convention*. Cambridge: Cambridge University Press.

Tybur, J.M., D. Lieberman, R. Kurzban, and P. DeScioli. 2013. “Disgust: Evolved Function and Structure.” *Psychological Review* 120 (1): 65–84.

Ugazio, G., C. Lamm, and T. Singer. 2012. “The Role of Emotions for Moral Judgments Depends on the Type of Emotion and Moral Scenario.” *Emotion* 12(3): 579-590.

van Overschelde, J. 2008. “Metacognition: Knowing About Knowing.” In *Handbook of Memory and Metamemory,* edited by J. Dunlosky and R. Bjork, 47-71. New York: Taylor and Francis.

van Prooijen, J. 2021. “Overconfidence in Radical Politics.” In *The Psychology of Populism*, edited by J.P. Forgas, W.D. Crano and K. Fiedler, 143-157. New York: Routledge.

Västfjäll, D., P. Slovic, M. Mayorga, and E. Peters. 2014. “Compassion Fade: Affect and Charity Are Greatest for a Single Child in Need.” *PLOS One* 9 (6): e100115.

Väyrynen, P. 2008. “Some Good and Bad News for Ethical Intuitionism.” *The Philosophical Quarterly* 58 (232): 489-511.

Vega, S., A. Mata, M.B. Ferreira, and A.R. Vaz. 2020. “Metacognition in Moral Decisions: Judgment Extremity and Feeling of Rightness in Moral Intuitions.” *Thinking and Reasoning* 20 (2): 215-244.

Vuilleumier, P. 2005. “How brains beware: neural mechanisms of emotional attention.” *Trends in Cognitive Sciences* 9 (12): 585-594.

Ward, S.J., and L.A. King. 2018. “Individual Differences in Reliance on Intuition Predict Harsher Moral Judgments.” *Journal of Personality and Social Psychology* 114 (5): 825–849.

Weinberg, J.M. 2007. “How to Challenge Intuitions Empirically Without Risking Skepticism.” *Midwest Studies in Philosophy* 21: 318-343.

Wheatley, T., and J. Haidt. 2005. “Hypnotic Disgust Makes Moral Judgments More Severe.” *Psychological Science* 16: 780-784.

Wiegman, A., Y. Okan, and J. Nagel. 2012. “Order effects in moral judgment.” *Philosophical Psychology* 25 (6): 813-836.

Wiggins, D. 1987. “A Sensible Subjectivism.” In *Needs, Values, Truth: Essays in the Philosophy of Value*, 185–214. Oxford: Blackwell.

Woodcock, S. 2021. “Thinking the Right Way (at the Right Time) about Virtues and Skills.” *Ethical Theory and Moral Practice* 24: 577–586.

Wright, J.C. 2010. “On intuitional stability: The clear, the strong, and the paradigmatic.” *Cognition* 115: 491–503.

—. 2013. “Tracking instability in our philosophical judgments: Is it intuitive?” *Philosophical Psychology* 26 (4): 485–501.

—. 2016. “Intuitional Stability.” In *A Companion to Experimental Philosophy*, edited by J. Sytsma and W. Buckwalter, 568-577. Malden, MA: Wiley.

Young, L., A. Bechara, D. Tranel, H. Damasio, M. Hauser, and A. Damasio. 2010. “Damage to Ventromedial Prefrontal Cortex Impairs Judgment of Harmful Intent.” *Neuron* 65: 845-851.

Young, L., M. Koenigs, M. Kruepke, and J.P. Newman. 2012. “Psychopathy Increases Perceived Moral Permissibility of Accidents.” *Journal of Abnormal Psychology* 121(3): 659–667.

Zagzebski, L. 1996. *Virtues of the mind*. Cambridge: Cambridge University Press.

Zamzow, J.L., and S. Nichols. 2009. “Variations in Ethical Intuitions.” *Philosophical Issues* 19: 368-388.