Intuitions about Cases and Principles: A Defense of Reflective Equilibrium

by

Francisco Galan Tames

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Arts

Department of Philosophy

University of Alberta

Abstract

Ethical reasoning in normative and practical ethics is typically conducted by working back and forth between intuitions about both general moral principles (e.g., suffering is bad) and particular cases (e.g., that I must prevent a specific person from suffering). These two types of intuitions—which I refer to as *principle-intuitions* and *case-intuitions*—are fundamental tools that ethicists unavoidably use to answer ethical questions. However, principle-intuitions and case-intuitions often conflict: a principle such as lying is wrong could be inconsistent with the intuition that it is permissible to lie to the Nazis to save some refugees. What should one do when conflicts like these arise? Should one give more weight to the principle-intuition, to the case-intuition, or equally to both? Or could it be that there is, in fact, no justificatory relation between the two elements but a different one? For example, perhaps case-intuitions only serve to help us clarify or suggest plausible principles but not justify principles themselves?

Although the question of how should case-intuitions weight against principle-intuitions underlies almost any ethical deliberation and debate in normative and practical ethics, there is no universally-accepted answer to it. In other words, there is no unanimous agreement on what our ethical methodology should be. Thus, this thesis seeks to contribute to the literature by arguing in favor of one version of the most commonly-used method: reflective equilibrium. Among other things, reflective equilibrium holds that, all else being equal, case-intuitions carry the same epistemic weight as principle-intuitions.

The first part of this thesis (Chapters 1-2) offers an argument in favor of reflective equilibrium. This argument seeks to establish reflective equilibrium as the default methodology in a way that, I believe, has not been articulated in the literature, although it

arrives at similar conclusions as those of other supporters of this method. My claim is that we have reason to give credence to both case-intuitions and principle-intuitions because a) we should give credence to a judgment if it seems plausible to us, b) both cases and principles can seem plausible to us to an equal degree, and c) there is no reason at the outset of inquiry to think that case-intuitions are unreliable in a systematic way.

The second part of the thesis (Chapters 3-4) offers a reply to two objections that have been raised against reflective equilibrium. Chapter 3 addresses whether case-intuitions are systematically unreliable due to their evolutionary, cultural, social, and emotional influences. Chapter 4 examines whether case-intuitions are unreliable because they change in response to morally irrelevant features, such as the wording of a case, the force of habit, or the emotions we might experience while thinking about a case.

Acknowledgments

I want to express my most profound gratitude to Howard Nye, my academic and thesis supervisor. Howard was not only incredibly generous with his time—as a professor and as a supervisor—but also very patient along the way, encouraging me to think more carefully about each issue and offering valuable advice, support, and feedback. It was an absolute privilege to learn from him, and I couldn't have wished for a better mentor and inspiration.

I am also grateful to those who contributed in some way or another to this academic project. I want to thank Jennifer Welchman, for helpful comments on an earlier draft of this thesis, Ingo Brigandt, for his academic guidance and support, as well as my undergraduate professors Michael Baur, José Antonio Pardo, and Carlos Hernández, for encouraging me to be more careful in my thought and writing. My special thanks go to Peter Andes, a fellow graduate student, whose support, friendship, and comments made this process much more manageable. Also, this work would not have been possible without the generous financial support from the University of Alberta, from MITACS, and from the Mexican Ministry of Education.

Lastly, I want to thank my parents and family: all of this would not have been possible without all your love and support.

Table of Contents

Introduction

Whenever we engage in moral reasoning, there comes a moment when some of our intuitions conflict. We might believe, for instance, that people should never lie, but one day encounter a case in which we are led to think that perhaps lying is permissible on some occasions. Maybe we thought of someone who dishonestly told her grandparents that she liked their birthday gift or of someone who lied to a Nazi official to prevent a murder. Now we must do something if we want to resolve the tension. Our first option is to keep our original belief and maintain that people should never lie. Our other option is to give preference to our intuition about the cases by rejecting or modifying our initial belief. We might modify it and now be inclined to think that people should never lie *unless* it does not harm anyone (as with the grandparents) or prevents someone from dying (as in the Nazi situation). Regardless of what we believe is the right option, one thing is clear: our answer will depend to a large extent on our ethical methodology, namely, about how we think (or assume) intuitions about particular cases and general moral principles should weight-off in ethical inquiry.

Ethical methodology is relevant because it constitutes the basis by which we favor or disfavor theories in normative ethics and decide upon pressing issues in practical ethics. In normative ethics, for instance, one might support some theories over others depending on which intuitions one gives higher weight to. As an example, consider a case that seems to conflict with consequentialism: the case of a doctor who kidnaps and takes the organs of an innocent pedestrian as the only means of saving five patients in need of a transplant.[1] If we accept the intuition that it would be wrong for the doctor to do so, then it would deem objectionable a consequentialist principle such as *we should always do as much good as we can for everyone concerned*. However, if we disregard the intuition about the transplant case—say, by arguing that it is biased—then our consequentialist principle would remain unchallenged. The same goes for practical ethics, an area in which many arguments rely on the justificatory role of intuitions about cases. Judith Thomson, for example, famously used

[1] This case originally appeared in Judith Jarvis Thomson, "The Trolley Problem," *The Yale Law Journal* 94, no. 6 (May 1985): 1395–1415.

a thought experiment involving an unconscious violinist to argue in favor of the permissibility of abortion.[2] Likewise, Peter Singer's drowning child example has often been used as an analogy for the duties we have towards poor people in other countries.[3]

In this way, a philosophical strategy to object to an ethical view is to argue that an intuition on which the view is based is unreliable, and there are many examples of this argumentative strategy in the philosophical literature. Regarding the debate of how demanding should morality be, Brian Berkey contends that some positions that hold that "morality is not significantly more demanding than most of us ordinarily take it to be" are ultimately based on unreliable intuitions.[4] For that reason, Berkey argues that the intuition that morality should not be that demanding does not pose a problem for demanding ethical positions, such as the views that we have substantial obligations to aid the global poor or alleviate climate change. Similarly, Michael Huemer argues that the intuition typically triggered by the Repugnant Conclusion in population ethics is unreliable, as it is influenced by multiple distorting factors. Hence, Huemer embraces the 'Repugnant' Conclusion and calls it "one of the few genuine, nontrivial theorems of ethics discovered thus far."[5] One last

---

[2] The violinist thought experiment was first introduced in Judith Jarvis Thomson, "A Defense of Abortion," *Philosophy and Public Affairs* 1, no. 1 (Autumn 1971): 47–66. In this thought experiment, we are asked to imagine waking up plugged into an unconscious famous violinist whose kidneys have stopped working and whose blood type matches only that of us. We were kidnapped and plugged into him overnight because this was the only way to save the violinist. Now, if we decide to unplug our circulatory system from his, then the violinist would die. If we remain plugged, then we would only have to wait nine months to unplug in order to prevent any harm done to the violinist. Jarvis Thomson argues that our intuition leads us to think that it is permissible to unplug even though the violinist has a right to life. The case is then used as an analogy for the permissibility of abortion.

[3] The drowning child example is described as follows: "If I am walking past a shallow pond and see a child drowning in it, I ought to wade in and pull the child out. This will mean getting my clothes muddy, but this is insignificant, while the death of the child would presumably be a very bad thing." (Peter Singer, "Famine, Affluence, and Morality," *Philosophy & Public Affairs* 1, no. 3 (Spring 1972): 229–43). Singer, however, in accordance with his metaethics, has said that this thought experiment merely serves to illustrate—not justify—the principles he presents in that same article (See Peter Singer, *Practical Ethics* (New York: Cambridge University Press, 1993), pp. 169-173.

[4] See Brian Berkey, "The Demandingness of Morality: Toward a Reflective Equilibrium," *Philosophical Studies*, no. 173 (2016): 3015–35, p. 3020.

[5] The debate about the Repugnant Conclusion, in what is called "population ethics," centers around the Impartial Total Principle, which states that, *If other things are equal, the best outcome is the one in which there would be the greatest net sum of happiness minus misery*. Against it, Derek Parfit pointed out that this principle implies what he called the Repugnant Conclusion: "For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better, even though its members have lives that are barely worth living" (Derek Parfit, *Reasons and Persons* (Oxford: Claredon Press, 1984), pp. 387-388). For Huemer's discussion of it, see Michael Huemer, "In Defence of Repugnance," *Mind* 117, no. 468 (October 2008).

example comes from Lazari-Radek and Singer, who, in a recent introductory book, defend classical utilitarianism by arguing that typical objections to it are based on unreliable case-intuitions.[6]

The previous examples show that questions about ethical methodology matter, because understanding how intuitions about particular cases and general moral principles should be weighted in ethical inquiry will help us reach more sensible conclusions on normative and practical matters. For as it often happens when concrete evidence or principles in any discipline clash, it helps to move up the justificatory ladder and try to find the further theoretical assumptions that might help us clarify and resolve the tensions.[7]

In answering the central question of this thesis, I will put aside metaethical debates about the metaphysical or semantical aspects of moral judgments. I take it that regardless of which position we adopt in these debates, our defence of our moral theory will be similar, and it will not be drastically altered by which metaethical position we favor. If one uses certain intuitions to ground a theory, for instance, it is irrelevant whether these are seen as expressing truths, beliefs, or attitudes, since either way we would be appealing to the same intuitions.[8] Moreover, I will be focused on ethical methodology in a normative sense, rather than from a practical and psychological standpoint (e.g., coming up with strategies to think more clearly about moral issues), and at the level of individual beliefs, rather than at the level of public debate (e.g., ways in which we ought to reason as a society or as a community).

In what follows, I explain some aspects of my project. In the first two sections, I define three technical concepts that I will be using throughout the thesis. In the third

---

[6] See Katarzyna Lazari-Radek and Peter Singer, *Utilitarianism: A Very Short Introduction* (Oxford: Oxford University Press, 2017).

[7] There are many examples in various disciplines of why moving up the justificatory ladder is helpful. In science, more fundamental principles can be revised if there are unsolved conflicts between observations and less fundamental principles. A clear of example is quantum physics, a field which abandoned what seemed to be highly plausible principles—e. g. that things cannot be in two locations at the time—to make sense of new observations at a sub-atomic level. Another example comes from legal practice, in which one interpretation of the law might be favored over another depending on how it aligns with more general theoretical principles (See Ronald Dworkin, "In Praise of Theory," *Arizona State Law Journal* 353, no. 29 (1997), pp. 356-357).

[8] Hurka makes a similar remark about the relation between normative methodology and whether one is a non-naturalist or a non-cognitivist. See Thomas Hurka, *British Ethical Theorists from Sidgwick to Ewing* (Oxford: Oxford University Press, 2014), p. 107.

section, I give a synopsis of the rest of the thesis and of the debate around the issues pertaining this thesis.

i. Terminology: case-intuitions and principle-intuitions

Let us refer to intuitions about particular cases as *case-intuitions* (also called "case-based intuitions,"[9] "concrete intuitions,"[10] or "practical intuitions" [11]). We can understand a case-intuition as an intuition that supports a judgment with a particular level of generality, that is, a judgment about a particular action or situation and that describes the what, where, when, why, how, by whom, or to whom of a contemplated action.[12] For example, *that I must prevent a specific person from dying* or *that a specific CEO should not get paid considerably more than other employees* constitute case-intuitions. In general, these are the intuitions that philosophers most often refer to when they discuss intuitions in ethics and the ones that are typically triggered by philosophical thought experiments.

Also, let us also refer to intuitions about general moral principles as *principle-intuitions* (also called "theoretical intuitions"[13] or "abstract theoretical or mid-level intuitions" [14]). These intuitions refer to the intuitions that support a judgment with a more abstract level of generality, that is, a judgment which encompasses various actions or situations. For example, the intuitions elicited by the judgments *Suffering is bad* or *All human beings are born free and equal in dignity and rights* constitute principle-intuitions.

There is, of course, no clear-cut line between case-intuitions and principle-intuitions, since one can always expand or narrow the generality of the judgments we examine. For example, with regard to case-intuitions, one could present a scenario in which, instead of preventing one death, we could prevent two, twenty, a hundred deaths.

---

[9] Brian Berkey, "Climate Change, Moral Intuitions, and Moral Demandingness," *Philosophy and Public Issues* 4, no. 2 (2014): 157–89 p. 161.
[10] Michael Huemer, "Revisionary Intuitionism," *Social Philosophy and Policy* 25, no. 1 (2008): 368–92, p. 383.
[11] Joakin Sandberg and Niklas Juth, "Ethics and Intuitions: A Reply to Singer," *The Journal of Ethics* 15 (2011): 209–26, p. 213-215.
[12] Here I follow to the definition provided by Sandberg and Juth, p. 213-215, and of Albert W. Musschenga, "Empirical Ethics, Context-Sensitivity, and Contextualism," *Journal of Medicine and Philosophy* 30 (2005): 467–90.
[13] Sandberg and Juth, "Ethics and Intuitions: A Reply to Singer", pp. 213-215.
[14] Huemer, "Revisionary Intuitionism", p. 383.

Likewise, we can always narrow the generality of a principle and say, "All human beings, *in Latin America*, are born free and equal in dignity and rights." If we continue to do this, there will be a point where it is unclear whether we are talking about a case-intuition or a principle-intuition. Therefore, perhaps it is best to understand these intuitions as part of a continuum of varying *degrees* of generality, with each type of intuitions being at one end of the continuum.

ii. Terminology: intuitions

So far, I have been using the concept of *intuition* loosely, but it is also worth discussing what it refers to in order to avoid the risk of having a mere verbal dispute once we delve into more specific issues. In the philosophical literature, there is disagreement about what exactly an intuition is.[15] On one account, an intuition is a special type of *belief*, one that is firmly held, not derived from other beliefs, and justified by our understanding of it (e. g., 2 is less than 3).[16] On another account, intuitions are a conscious, *distinctive phenomenology or mental state* in which a proposition seems true or untrue (e.g., the mental state that arises in reaction to Peter Singer's drowning child example).[17] On a third account, the term *intuition* does not actually have a clear application or even something that falls within it.[18] Due to this plurality of views—and the further disagreements about the specific set of characteristics that intuitions have—some have even said that it seems hard "to isolate a single feature of intuition that elicits unanimous agreement."[19]

I believe, however, that this disagreement does not extend entirely to questions about ethical methodology once we think about how we define concepts and the role of

---

[15] See Joel Pust, "Intuition," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Summer 2017 (Metaphysics Research Lab, Stanford University, 2017), https://plato.stanford.edu/archives/sum2017/entries/intuition/.

[16] See Robert Audi, *The Good in the Right: A Theory of Intuition and Intrinsic Value* (Princeton, NJ: Princeton University Press, 2004), p. 33-36; Ernest Sosa, "Minimal Intuition," in *Rethinking Intuition*, ed. M. DePaul and W. Ramsey (Lanham, MD: Roman and Littlefield, 1998), 201–40.

[17] See Huemer, "Revisionary Intuitionism", p. 370-371; George Bealer, "A Priori Knowledge and the Scope of Philosophy," *Philosophical Studies* 81 (1996): 121–42, p. 123.

[18] See Timothy Williamson, *The Philosophy of Philosophy* (Oxford: Oxford University Press, 2007); Jennifer Nado, "Why Intuition?," *Philosophy and Phenomenological Research* 86 (2011): 15–41; Herman Cappelen, *Philosophy Without Intuitions* (Oxford: Oxford University Press, 2012).

[19] Jennifer Nado, "The Intuition Deniers," *Philosophical Studies* 173, no. 3 (2016): 781–800, p. 783.

intuitions in ethics. Let us review briefly two methods for defining concepts: conceptual analysis and, what I call, the normative method. I will argue that conceptual analysis, the traditional method for defining concepts, is inadequate for defining the term *intuition* and that an alternative approach, the normative method, is a better option.

Conceptual analysis. The standard method for answering questions of the form "What is X?" is known as conceptual analysis. As articulated by Frank Jackson, what we do in conceptual analysis is give initial preference to the description of a particular concept that aligns best with our ordinary conception of it. This is not to say that the preferred definition will always be the closest one to our ordinary conception—as there can be other outweighing considerations—but that it will have at least *prima facie* value.[20] Our ordinary conception, in turn, is revealed by our evidence and intuitions about possible cases, either real-world cases or thought experiments.[21] As Jackson says: "We are seeking the hypothesis that best makes sense of their responses taking into account all the evidence."[22]

To illustrate this method, consider the way in which some authors support their definition of *intuition*. It is argued, for instance, that an intuition is distinct from a belief because it is possible to believe a proposition but have an opposing intuition about it or vice versa.[23] Paradoxes are presented as an illustration, because one can reason through a paradox and identify an implausible proposition but still find it intuitive, much like optical illusions do not lose their intuitive appeal even after careful examination. This way of arguing would fall within conceptual analysis because one is using intuitions about thought experiments—the paradox and optical illusion cases—to modify one's definition.

Yet, these thought experiments only have probative value as long as one accepts conceptual analysis. Otherwise, an objector could merely double-down and say that

---

[20] See Frank Jackson, "The Role of Conceptual Analysis," in *From Metaphysics to Ethics: A Defence of Conceptual Analysis* (Oxford: Claredon Press, 1998).
[21] Jackson, p. 31. It may be objected that Jackson's account of conceptual analysis as applied to the definition of *intuition* is circular, since we would be presupposing a notion of intuition (e. g. finding a proposition 'intuitive') to arrive at another notion of intuition. This would be correct unless there are, in fact, two notions being used: the revised notion that is suggested *after* going through the example and the pre-theoretical, ordinary notion that appears in the premises.
[22] Jackson, p. 36.
[23] Bealer, "A Priori Knowledge and the Scope of Philosophy", p. 123; Pust, "Intuition."

intuitions are still beliefs and that the thought experiments presented illustrate a different concept, one unrelated to the concept of *intuition*.

Conceptual analysis as applied to the case of intuitions, however, faces various problems which make it an inadequate method for our purposes. The first problem with conceptual analysis is that intuitions about ordinary concepts may not be shared or even consistent. There is are various studies that suggest that there are gender, individual, and cultural differences in our intuitions about cases.[24] If this is correct, then conceptual analysis will arguably provide little guidance in finding a common definition of *intuition*, because it might not help us to solve disputes if people have different original conceptions about this concept. Certainly, there have not been any studies about the concept of intuition in particular, and this is an empirical question, so it can turn either way. But one could be at least suspicious about the prospects of finding common ground by using this method: so far, it hasn't.

Secondly, conceptual analysis seems to give us a descriptive answer, not a normative one. In other words, even if we discover our original conception of a concept, we still lack a further reason to endorse it over other definitions. The fact that we happen to have a certain original conception of a concept does not show that this is the best way to understand it. If our original conception really had epistemic priority, we could simply conduct polls to discover people's original conception—as Jackson actually suggests—but, admittedly, this would only yield an empirical fact about how people use the concept, not about how we ought to use it. Are there additional reasons which might give normative force to our original conceptions, besides it simply being the original one?

A tentative reason might be that our original conceptions provide a starting place for discussion. This is expressed in an informal remark often given in defense of conceptual analysis: "We have to start somewhere. Don't we?" However, this reason would still not grant any normative force to our original conception, only practical usefulness. To give an analogy, if someone learned philosophy by reading Plato, this does not mean that he or she

---

[24] See Jonathan M. Weinberg, Shaun Nichols, and Stephen Stitch, "Normativity and Epistemic Intuitions," *Philosophical Topics* 29, no. 1 (Spring and Fall 2001); Joshua Knobe and Nichols Shaun, eds., *Experimental Philosophy* (New York: Oxford University Press, 2008); Wesley Buckwalter and Stephen Stich, "Gender and Philosophical Intuition," in *Experimental Philosophy, Volume 2*, ed. Knobe Joshua and Nichols (Oxford: Oxford University Press, 2014), 307–46.

should give preference to Plato's views, just because he or she had to start doing philosophy from reading Plato. It merely means that reading Plato is a convenient way to start learning about philosophy. An original conception, also, would not be necessary, as we could always agree to start from somewhere else.

Another possible reason in favoring our original conception might be that it helps us communicate about the same topics. Jackson says, for example, that it is only by conceptual analysis that we "define our subject—or, rather, only that way do we define our subject as the subject we folk suppose is up for discussion."[25] This suggests that it is impossible to discuss any topic without relating it to our original conceptions; yet, we could do this by merely being explicit about what we will be discussing. I could say, for instance, that I will be addressing the concept of intuition understood as a paranormal mental phenomenon, without thereby making any reference to our original conception. One could object that this is not what we usually mean by *intuition*, but this begs the question, since we would be presupposing that our original conception has a privileged status and therefore that my definition does not align with it. Admittedly, I might be considered an irrelevant interlocutor, since I would not be discussing a way of understanding *intuition* others have any interest in. But there is no denying that I would be defining my subject, even if it is in a counter-intuitive way.

Much more could be said about the method of conceptual analysis, but let us turn now to an alternative approach for defining concepts: the normative method (also referred to as the "analytical" or "ameliorative" approach[26]). This method might not necessarily be the one we want to use throughout philosophy whenever we want to define a concept, but it might be better suited to help us define *intuition* for the purposes of ethical methodology.

The normative method. In the normative method, we begin by asking what cognitive or practical task do concepts (or should concepts) enable us to accomplish and whether they are effective tools to do so. Instead of asking "What is X," we begin by asking

---

[25] Jackson, "The Role of Conceptual Analysis", p. 42.

[26] See Sally Haslanger, "Gender and Race: (What) Are They? (What) Do We Want Them to Be?," *Nous* 34, no. 1 (2000): 31–55, p. 33-34; Sally Haslanger, "What Are We Talking about? The Semantics and Politics of Social Kinds," *Hypatia* 20 (2005): 10–26; Sally Haslanger, "What Good Are Our Intuitions?," *Aristotelian Society Supplementary Volume* 80 (2006): 89–118.

"What we want X to be?" Our definitions are then constructed according to the general purposes that we have in a particular area and not by some underlying, essential characteristic of them.[27] As Gupta mentions, "different definitions do not all have the same goal: the boundary commission may aim to achieve precision; the Supreme Court, fairness […]. The standards by which definitions are judged are thus liable to vary from case to case."[28] This makes our definition stipulative, but limited by the goals that we have, which can include, ideally, anything that we agree upon. In the case of intuitions, instead of trying to figure out the essence of intuitions, we can begin by asking what is the point of having them at all or why do we care about them in the first place.

Insofar as one of the goals of having a methodology in ethics is to be able to justify some of our judgments adequately,[29] intuitions seem to matter because they provide one special type of justification: *non-inferential justification*. In other words, intuitions are able to justify other judgments, but their justification is not grounded on other premises. This is one of the central tasks that intuitions help us accomplish in ethical methodology, and it follows to some degree G. E. Moore's usage of the term: "When I call such propositions 'Intuitions,' I merely to assert that they are incapable of proof; I imply nothing whatever as to the origin of our cognition of them."[30] Other authors also seem to agree on this point. Pust, for instance, mentions that it is more critical whether our notion of 'intuition' captures relevant epistemological joints—such as non-inferentiality—rather than our linguistic usage.[31] Chalmers also raises doubts about the specific phenomenology or basis of intuitions but concedes that non-inferentiality about intuitions is of central importance.[32]

Indeed, consider what would happen if we abandoned non-inferentiality in our definition of intuitions. For those who think that intuitions are beliefs, abandoning non-

---

[27] Haslanger, "Gender and Race: (What) Are They? (What) Do We Want Them to Be?", p. 33.
[28] Anil Gupta, "Definitions," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Summer 2015 (Metaphysics Research Lab, Stanford University, 2015).
[29] One could say that the point of having a methodology of not to justify our judgements but to find truth. But, arguably, in order to achieve truth we need to distinguish between those judgements that are justified and those that are not, so even if justification is not an ultimate goal, the following considerations still aply.
[30] G. E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1903), Preface ¶ 6.
[31] See Pust, "Intuition."
[32] See David J. Chalmers, "Intuitions in Philosophy: A Minimal Defense," *Philosophical Studies* 171 (2014): 535–44, p. 536.

inferentiality would mean that most beliefs count as intuitions, since many of our beliefs are justified inferentially (that is, by appeal to other beliefs), and this would overextend our definition. And for those who think that intuitions are mental states, abandoning non-inferentiality would make it hard to distinguish between having an intuition and the mere process of reasoning, and this seems to encompass too much. Of course, under the normative method, it would be entirely possible to define *intuition* in either of these ways, but we would need to evaluate whether this would help us in any way to achieve our methodological tasks.

A second characteristic of intuitions is that the non-inferential justification they provide is *gradual*, in the sense that intuitions can go from providing a weak justification to a strong, almost self-evident one. This is clear from our everyday and philosophical experience. Some judgments seem compelling to a certain extent, and there can be others which are way more compelling, slightly more compelling, slightly less compelling, or not compelling at all. Additionally, the degree to which we accept intuitions triggered by thought experiments in philosophy seems to vary. Consider that the Repugnant Conclusion in population ethics seems totally unacceptable to many,[33] whereas Judith Thomson's violinist example might elicit mixed responses.[34]

A third claim that can be made about intuitions is that, in principle, they need not have a specific set of psychological features. To see why this is the case, consider whether it would be important if underlying psychological features about intuitions turned out to be one way or another. Suppose it turned out that three different areas of the brain produced three different mental phenomena that we had associated with the term *intuition*. Would this be of any relevance? As long as these mental phenomena still provided non-inferential justification, I believe it would not matter whether we called them all *intuitions*. Insofar as they helped us to justify our moral principles, it seems unnecessary to create new concepts to refer to them just because they originated from different places. Granted, the cerebral origin of these phenomena might correlate with the *degree* to which they provide

---

[33] As formulated by Derek Parfit, the Repugnant Conclusion is that "for any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better even though its members have lives that are barely worth living." Parfit, *Reasons and Persons*, p. 388.

[34] Thomson, "A Defense of Abortion."

justification. For instance, we might decide to confer higher credence to intuitions generated in the prefrontal cortex than to those originated in the limbic system because the former might be—for whatever reason—more reliable than the latter. If this were the case, then it might be reasonable to create new concepts to refer to them. Nevertheless, we would still be ultimately basing our classification on the non-inferential justification that these mental phenomena provide, and the psychological particularities would only have instrumental value. Therefore, it is not central to our definition whether the term *intuition* refers to a single, assorted, or vague psychological kind—as some authors argue[35]—or whether our philosophical definition aligns with the ones offered in the psychological literature.[36] Under the normative method, these psychological definitions are not central to ethics, since ethicists and scientists have different goals in mind (the former might be interested in justification, whereas the latter are interested, among other things, in simplicity and empirical adequacy).

With this said, I propose that we understand intuitions as mental states or dispositions which, at the outset of inquiry, provide gradual and non-inferential justification to judgments. This definition will suffice for our purposes, even if more components could be added to it. For example, one could object that the definition is not sufficiently discriminating, for it would include sensory perceptions as intuitions. However, if we consider that our goal in ethical methodology is to justify moral judgments, this would not be a problem, since sensory perceptions are not used to justify moral judgments in ethics. Moreover, it is important to emphasize the terms "at the outset of inquiry" in the previous definition to avoid taking sides on some methodological disputes. Someone might hold the view that case-intuitions do not ultimately provide non-inferential justification—even though he or she might concede that they can be compelling—and, therefore, not even refer to case-intuitions as *intuitions*. To avoid this problem, we say that intuitions provide

---

[35] See Williamson, *The Philosophy of Philosophy*; Nado, "Why Intuition?"

[36] The psychologists Jonathan Haidt and Joshua Greene, for example, understand intuitions as a mental processes that occur quickly, effortlessly, and automatically and that form a judgement with an affective valence. See Jonathan Haidt, "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review* 108, no. 4 (2001): 814–34, p. 818; Joshua D. Greene, "From Neural 'Is' to Moral 'Ought': What Are the Moral Implications of Neuroscientific Moral Psychology?," *Nature Reviews Neuroscience* 4 (2003): 847–50, p. 848.

non-inferential justification *at the outset of inquiry*, to leave open the possibility that they might ultimately not.

With these considerations in place, we can now proceed to give an overview of the topics that will be addressed throughout this thesis.

iii. Overview of the debate

When we ask the question of how should case-intuitions and principle-intuitions weigh-off in ethical inquiry, there are four general positions we can adopt. The first position is skepticism about moral intuitions in general.[37] The skeptic will reject the initial question altogether by saying that both case-intuitions and principle-intuitions have, actually, no weight at all. For the purposes of this thesis, I will not address the skeptical view, however, although the argument I offer in Part I still gives a reason to reject it.

The second position we can adopt is to say that case-intuitions have more epistemic weight than principle-intuitions. Typically, this position takes case-intuitions as the only valid data or premises from which we should think about morality. Although there is not a specific label to refer to this view, three philosophical positions sympathetic to this idea are *anti-theory*, *particularism*, and *virtue ethics*.[38] Anti-theory can be defined as the view that normative theory—the set of systematic principles aimed at guiding our behavior—is unnecessary, theoretically impossible, or undesirable.[39] Particularism can be thought of as a type of anti-theory, as it is the view that morality does not depend on a set of general principles and that features that count in favor of an action in one context can count against, or not at all, in another.[40] Although neither of these two positions explicitly make a

---

[37] Although he is not skeptical of ethics in general, Brandt is skeptical about moral intuitions. See R. Brandt, *A Theory of the Good and the Right* (Oxford: Claredon Press, 1979).

[38] Thomas Hurka mentions that the ethical theorists Carritt and Erwing endorsed the view that the most reliable intuitions were the ones about particular cases (Hurka, *British Ethical Theorists from Sidgwick to Ewing*, p. 122).

[39] See Stanley G. Clarke, "Anti-Theory in Ethics," *American Philosophical Quarterly* 24, no. 3 (1987): 237–44, p. 237.

[40] Although particularism can be cast in a number of ways, here, I am following Frykholm's reconstruction in Frykholm Erin, "A Humean Particularist Virtue Ethics," *Philosophical Studies* 172 (2015): 2171–91. For more discussion, see Jonathan Dancy, "Moral Particularism," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Fall 2013; and Jonathan Dancy, *Ethics Without Principles* (Oxford: Claredon Press, 2004).

claim about the moral weight of intuitions, they do suggest that principle-intuitions do not carry much weight and that case-intuitions do. Additionally, virtue ethical theories might also be said to be committed to a similar methodological view, although it is unclear whether all of them do or only some of them. Nonetheless, as with the previous case of general skepticism, in this thesis I will not consider the objections that anti-theorists, particularists, or virtue ethicists might provide against using moral principles in ethical reasoning. Still, the argument I will offer gives reasons to reject the claim that case-intuitions have a higher epistemic weight than principle-intuitions.

The third position we can adopt, and the one that will be central in my discussion throughout this thesis, is to say that principle-intuitions have more epistemic weight than case-intuitions. Those who subscribe to this idea are usually referred to as foundationalists,[41] and they try to conduct moral inquiry by identifying plausible, non-inferentially-based, or self-evident moral principles and construct their theories based on them. Some authors who use and defend this approach include Henry Sidgwick, G. E. Moore, Peter Singer, and Howard Nye.[42] The two objections I will consider in Part II come, precisely, from the foundationalist camp.

The last position we can adopt is to say that, all things being equal, case-intuitions carry the same epistemic weight as principle-intuitions. In other words, if a case-intuition conflicts with a principle-intuition, this provides evidence against accepting the principle-intuition and vice versa. This is the position known as reflective equilibrium, and it is the one I will defend in this thesis, although there are various ways in which versions of reflective equilibrium can differ.

---

[41] Although I use this label to follow traditional usage, I believe it is not an entirely accurate term. As I discuss in section 1.3, in epistemology, foundationalism refers to the view that "all knowledge or justified belief rest ultimately on a foundation of non-inferential knowledge or justified belief" (see Ali Hasan and Richard Fumerton, "Foundationalist Theories of Epistemic Justification," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Winter 2016). In ethics, this is taken to imply that those who favor principle-intuitions are foundationalists, since they start from a set of "ethical axioms" and work from there. However, one could also be said to be a "foundationalist" if one based his or her ethical theory merely on case-intuitions, as these would now constitute the foundational, non-inferential knowledge.

[42] For an account of Moore's and Sidgwick's position, see Hurka, *British Ethical Theorists from Sidgwick to Ewing*120-122. For Singer's position, see Katarzyna de Lazari-Radek and Peter Singer, *The Point of View of the Universe: Sidgwick and Contemporary Ethics* (Oxford: Oxford University Press, 2014). For Nye's account, see Howard Nye, "Directly Plausible Principles," in *The Palgrave Handbook of Philosophical Methods*, ed. Chris Daly (UK: Palgrave Macmillan, 2015).

I should note that foundationalism and reflective equilibrium can sometimes be indistinguishable in practice. For instance, a reflective equilibriumists might dismiss certain case-intuitions if he or she thinks they are unreliable, just as a foundationalist might.[43] However, the difference between the two positions is that in reflective equilibrium we cannot dismiss, in principle, all case-intuitions; we have to examine on a case-by-case basis to see whether a case-intuition is unreliable. In contrast, foundationalists could dismiss all case-intuitions, as they think that there is something systematically, intrinsically wrong about conferring epistemic weight to case-intuitions, even if they have other uses (see section 4.3). Additionally, a foundationalist might agree, in principle, with the claim that *all things being equal, case-intuitions carry the same epistemic weight as principle-intuitions*; yet, the foundationalist might argue that, in practice, things cannot be held equal, as there is something systematically, inevitably flawed about using case-intuitions.

The goal of this thesis is to provide a defense of reflective equilibrium: the view that case-intuitions and principle-intuitions have equal epistemic status. My defense will consist of two parts. Part I offers a positive argument in favor of this method which seeks to answer the question, *why reflective equilibrium?* Here, the aim is not only to defend reflective equilibrium but to establish reflective equilibrium as the default ethical methodology. Chapter 1 explains what different authors have understood by reflective equilibrium and what are its central characteristics. Chapter 2 presents a positive argument in favor of it, which, I believe, has not been articulated in the literature, although it arrives at almost the same conclusions as those of other supporters of reflective equilibrium. Roughly put, my argument is that we have reason to give credence to both case-intuitions and principle-intuitions because a) we should give credence to a judgment if it seems plausible to us, b) both cases and principles can seem plausible to us to an equal degree, and c) there is no reason at the outset of inquiry to think that case-intuitions are unreliable in a systematic way.

Part II of the thesis (Chapters 3-4) can be thought of as a further defense of the premise that there are no reasons at the outset of inquiry to systematically dismiss case-

[43] See Norman Daniels, "Reflective Equilibrium," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Winter 2016 (Metaphysics Research Lab, Stanford University, 2016).

intuitions. This is because, as I mentioned, a foundationalist might agree with most of the premises of my argument, but still claim that things cannot be held equal with regard to case-intuitions, as there is something systematically, inevitably flawed about case-intuitions. Therefore, Part II offers a reply to two major objections that foundationalists have raised against reflective equilibrium.[44] Chapter 3 addresses whether case-intuitions are unreliable because they can be evolutionary, cultural, social, and emotionally influenced. Chapter 4 examines whether case-intuitions are unreliable because they change in response to morally irrelevant features, such as the wording of a case, the force of habit, or the emotions we might experience while thinking about a case.

Without further ado, let us delve into these issues.

[44] For a discussion of other objections, see Kenneth Walden, "In Defense of Reflective Equilibrium," *Philosophical Studies* 2, no. 166 (2013): 243–56.

PART I: Reflective Equilibrium as the Default Methodology

Chapter 1 — What is Reflective Equilibrium?

1.1 Rawls's account in *A Theory of Justice*

John Rawls was the first to introduce the term *reflective equilibrium*, and it is illustrative to understand the context in which it is used. In Chapter I of his book *A Theory of Justice*, Rawls discusses—among other things—what is the best method to determine the proper conception of justice.[45] He proposes that we use a thought experiment: the original position. This is a hypothetical situation in which one is asked to imagine what principles of justice one would choose for society if one did not know "his place in society, his class position or social status, […] his fortune in the distribution of natural assets and abilities, his intelligence, strength, and the like."[46] By using this thought experiment, Rawls thought that we would be better positioned to put our personal biases or interests aside and reach a fairer set of principles and rules for society, ones which gave proper consideration to all persons.

But Rawls pointed out a potential problem with this approach: how are we to decide what the best formulation of the original position is? This thought experiment can be formulated as if people are selfish or altruistic, rational or irrational, risk-averse or risk-lovers, and since one specific set of restrictions will probably influence the conception of justice we end up choosing, we have to be careful about which conditions we stipulate. In order to decide, Rawls proposes two different but complementary steps. One is to think about those conditions that would elicit broad agreement, regardless of how trivial they might be. For instance, it seems reasonable to assume "that it should be impossible to tailor principles to the circumstances of one's own case"[47] and that everyone in the original position is equal in the sense that all "can make proposals, submit reasons for their acceptance."[48] A second way to proceed is to think whether the principles that result from a particular formulation of the original position match the ideas of justice in which we

[45] *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971), pp. 10-19.
[46] Rawls, p. 11.
[47] Rawls, p. 16.
[48] Rawls, p. 17.

already have high confidence in. For example, religious intolerance and racial discrimination are presumably unjust, so an appropriate formulation of the original position should not endorse intolerance or discrimination; if it does, we are justified in revising our stipulations of the thought experiment.

Rawls, in sum, instructs us to proceed in the following way. Come up with widely accepted constraints for the original position that most people would endorse; if they are too trivial, then try finding less trivial ones. Simultaneously, think about the judgments that people have high confidence in and see if they follow from our version of the original position; if they are too specific, then try finding more fitting ones. This should allow us to find principles both acceptable enough, so that people would endorse them, and non-trivial enough, so that they help us guide our actions. In case of conflict between a reasonable constraint and a plausible idea about justice, we can either revise the constraint or the idea, depending on what we think is best. After doing this a sufficient number of times, with thoroughness and carefulness, we should arrive at a point which "expresses reasonable conditions and yields principles which match our considered judgments duly pruned and adjusted."[49] This is the point of *reflective equilibrium*, which is also the name used to describe the procedure itself.

Reflective equilibrium, then, can be thought of as a method in which "justification rests upon the entire conception and how it fits in with and organizes our considered judgments."[50] As Rawls acknowledges, this structure of justification is similar to Quine's holism and metaphor of the "web of belief," in which a judgment is acceptable partly by how well it fits a body of theory.[51] It is also similar to the way Goodman justifies inductive logic:

> The basic task in justifying an inductive inference is to show that it corresponds to the general rules of induction. […] But how is the validity of the rules to be determined? [...] Their validity depends upon accordance with the particular deductive inferences we actually make and sanction. If a rule yields unacceptable inferences, we drop it as invalid [...] This looks flagrantly circular [...] But this circle

[49] Rawls, p. 18.
[50] Rawls, p. 507.
[51] See Rawls, p. 507; Peter Hylton, "Willard van Orman Quine," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Spring 2018 (Metaphysics Research Lab, Stanford University, 2018).

> is a virtuous one. […] The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences; and in the agreement achieved lies the only justification needed for either.[52]

In other words, justification does not depend on some set of foundational, self-evident or indubitable beliefs, but on how our beliefs interconnect with each other. More specifically, justification depends on *the degree of plausibility* that judgments have and *their degree of consistency with other judgments*. Thus, in reflective equilibrium, the moral theory we will prefer will be the one with the consistent set of judgments that, on the whole, seems most plausible to us.

In finding the set of judgments that seems most plausible, we take into account cases (in our terminology, case-intuitions), principles (in our terminology, principle-intuitions), and non-moral judgments (e.g., constraints like parsimony or consistency). Non-moral judgments can refer not only to empirical judgments in general but also to theories in other areas of philosophy or other disciplines that might bear relevance to moral theories. For instance, an account of personal identity over time will probably influence questions of moral responsibility and compensations of burdens, such as when one has to determine whether the person who performed a crime or was burdened is the same as the one who is now attributed responsibility or compensated.[53] Also, theories in the social sciences regarding what kind of biases we are prone to might inform our process of deliberation in ethics. These non-moral theories, in turn, must have some degree of independent justification, so that they do not merely serve as *ad hoc* devices or accidental generalizations to conveniently prefer certain moral judgments over others.[54]

Moreover, in finding the set of judgments that seems most plausible as a whole, one might revise *any* judgment, depending on its direct degree of plausibility and consistency with other judgments. There will be some judgments that are more central in our theories— in our "web of beliefs"—or to which we assign higher credence, and we might be more reluctant to abandon them; but they still can be revised:

---

[52] Nelson Goodman, *Fact, Fiction and Forecast* (Cambridge, MA: Harvard University Press, 1955), pp. 61-62.
[53] See David W. Shoemaker, "Utilitarianism and Personal Identity," *The Journal of Value Inquiry* 33 (1999): 183–99.
[54] See Norman Daniels, *Justice and Justification: Reflective Equilibrium in Theory and Practice* (Cambridge: Cambridge University Press, 1996), pp. 48-50.

> […] likely candidates for necessary moral truths are the conditions imposed on the adoption of principles; but actually it seems best to regard these conditions simply as reasonable stipulations to be assessed eventually by the whole theory to which they belong.[55]

A problem that Rawls noticed is that if one were to include only those judgments which we already accept, then our moral theory would merely be an expression of the *status quo*: a descriptive investigation of what people happen to believe. This is why Rawls points out that we should aim to include not only "those descriptions which more or less match one's existing judgments," but also "all possible descriptions to which one might plausibly conform one's judgments together with all relevant philosophical arguments for them."[56] In this way, reflective equilibrium is intended to be revisionary because it forces us to consider judgments and arguments which we might not initially agree with or think about.[57] This might involve "principles and theoretical constructions which go much beyond the norms and standards cited in everyday life."[58] Rawls refers to this approach as *wide reflective equilibrium*, which considers all plausible judgments and their consequences, in contrast to *narrow reflective equilibrium*, which only takes into account certain judgments.[59]

Additionally, to make reflective equilibrium even more capable of revision, Rawls explains that in this process of back-and-forth we can exclude certain judgments that might contaminate the process:

---

[55] John Rawls, "The Independence of Moral Theory," *Proceedings and Addresses of the American Philosophical Association* 48 (1975 1974): 5–22, p. 506.
[56] Rawls, *A Theory of Justice*, p. 43.
[57] Rawls recognizes that it is impossible in practice to consider all judgments and arguments, and is even doubtful that the set of possible judgments and philosophical arguments is well-defined. But this is intended as an heuristic and ideal goal, which is why Rawls suggests that use as a starting point the most established traditions in moral philosophy.
[58] Perhaps Rawls is exaggerating here, as in reflective equilibrium it is possible that we arrive both at a very counterintuitive moral theory or at a common-sensical one. The only point, then, is that we try to engage as many judgments and arguments as we can and see where this leads us.
[59] Rawls, "The Independence of Moral Theory", p. 8. Normal Daniels characterizes *wide reflective equilibrium* as the method which aims to include non-moral judgements as the set of considered judgments, in contrast to *narrow reflective equilibrium*, which only takes into account moral cases and principles: Norman Daniels, "Wide Reflective Equilibrium and Theory Acceptance in Ethics," *The Journal of Philosophy* 76, no. 5 (1979): 256–82, p. 259.

> …[considered judgments] enter as those judgments in which our moral capacities are most likely to be displayed without distortion. Thus in deciding which of our judgments to take into account we may reasonably select some and exclude others. For example, we can discard those judgments made with hesitation, or in which we have little confidence […] [or] those given when we are upset or frightened, or when we stand to gain one way or the other.[60]

In other words, in reflective equilibrium, while we aim to consider as many judgments as we can, it is advisable to exclude some of them, as they might contaminate the deliberation process. These judgments, according to Rawls, include a) those in which we have little confidence, b) those made under unreliable emotional states (e.g., being upset or frightened), and c) those which are influenced by personal interests (e.g., when we have something to gain). I take it that Rawls refers to both particular and general judgments, since any of these three previous conditions could apply to principles as well, although one could certainly argue that they apply more frequently to particular judgments.

Lastly, Rawls mentions one additional condition in which we might decide to exclude some judgments:

> An allowance must be made for the likelihood that considered judgments are no doubt subject to certain irregularities and distortions despite the fact that they are rendered under favorable circumstances. […] A person [...] may well revise his judgments to conform to its principles even though the theory does not fit his existing judgments exactly. He is especially likely to do this *if he can find an explanation for the deviations which undermines his confidence in his original judgments* [emphasis is mine].[61]

In this passage, Rawls is saying that even when a judgment is made under ideal conditions, if we find an explanation which undermines our confidence in the judgment, then we might exclude it from the deliberation process. Rawls does not elaborate on what exactly constitutes an "explanation of a judgment," but the idea sounds similar to the strategy that some foundationalists sometimes appeal to. On the one hand, one could explain a judgment by pointing out that its plausibility is *derivative*, that is, that it is dependent on the plausibility of a more fundamental judgment. For example, one could explain the duty to

---

[60] Rawls, *A Theory of Justice*, p. 42.
[61] Rawls, p. 42.

keep promises in terms of the duty not to cause harm, arguing that breaking a promise is wrong merely because it causes a harm. On the other hand, one could also explain a judgment by pointing out its *origin*. For instance, the British philosopher Rashdall Hastings thought that the belief in retributive punishment should weaken once we understand that it is based on an instinct for vengeance.[62] Another example comes from Peter Singer, who, in two recent works,[63] uses evolutionary theory to explain the origin of various case-intuitions and, in turn, undermine the confidence we have in them.[64] In this way, one could explain a judgment by pointing out its origin, although the explanation has to really undermine our confidence in the judgment: part of the origin of my belief in non-consequentialism, for instance, might be that I have read Kant, but this fact *per se* does not seem to undermine my confidence in non-consequentialism.

One could debate, of course, if Rawls even had any of these two interpretations in mind. Both of them seem plausible strategies to dismiss judgments, but they certainly depend on additional considerations. The latter one, explaining a judgment in terms of its unreliable origin, depends on whether this strategy is valid or not since one could argue that it is a genetic fallacy (see Chapter 3). Many foundationalists use this strategy in their argumentation, and I would not see why reflective equilibrium could not, in principle, incorporate it, as long as it filters *some* judgments only. The former one, explaining a judgment in terms of a more fundamental one, depends on of whether one successfully can show that the plausibility of a judgment is derivative. I believe, yet, that most philosophers would not object to this strategy by itself, but only to whether it is successful or not in a particular issue.

---

[62] See Hurka, *British Ethical Theorists from Sidgwick to Ewing*, p. 115.
[63] See Lazari-Radek and Singer, *The Point of View of the Universe*; Lazari-Radek and Singer, *Utilitarianism: A Very Short Introduction*.
[64] For example, Singer hints that we should be suspicious of our belief in the sanctity of infant life if it is based on feelings about babies being small, cute, and helpless, which are, in turn, explained by evolution. Singer might have said this merely as a rhetorical strategy, and not as a normative claim, but, still, I believe one could draw this interpretation from what he says. See Singer, *Practical Ethics*, p. 170.

1.2 The characteristics of reflective equilibrium

Let us now articulate more systematically the characteristics of the method of reflective equilibrium that Rawls proposes:

> 1. *In reflective equilibrium, we aim to take into account all possible judgments—moral and non-moral—along with all relevant philosophical arguments*. This allows us to consider positions that we initially did not think of or agree with, and it also forces us to consider judgments regardless of their form, structure, or level of generality.
>
> 2. A*ll judgments—particular or general, moral or non-moral—are, in principle, open to revision*. In other words, we do not take some judgments as unshakeable beliefs or axioms, although they could have a comparable role if they remain highly plausible after reflection.
>
> 3. *Judgments are revised depending on how plausible they seem and how well they fit with other judgments*. In other words, it is the *degree of plausibility* of the judgment and its *degree of fit with other judgments* that determines whether we revise it or not.
>
> 4. *Certain judgments can be excluded from the deliberation process*, such as when a) we have little confidence in them, b) they are made under unreliable emotional states, c) they are influenced by personal interests, or d) one can provide an explanation which undermines our confidence in them (i.e., an explanation based on the judgment's origin or derivative plausibility).

These four characteristics are the ones that Rawls understands, in general, as constitutive of reflective equilibrium. However, if we want to be even more precise, the only characteristics that we really need are (1) and (3), since (2) and (4) seem to follow from the former. To see this, note that (1) asks us to consider all possible judgments and that (3) tells the criteria by which we ought to revise them. Therefore, with these premises only, we are allowed to revise any judgment, something which (2) also states. Moreover, (4) can be thought as a set of highly plausible judgments that we take into account in the process of reflective equilibrium itself. These judgments can be thought of as those highly plausible non-moral judgments that guide or constraint our deliberation, analogous to the

conditions for the adoption of principles that Rawls presents in his theory. I am also reluctant to include them as constitutive of reflective equilibrium because, in Part II, I will argue, contrary to Rawls, that the specific conditions expressed in (4) should not allow us to systematically exclude particular judgments from the outset.

Another thing to note is that I have articulated reflective equilibrium as presented in Rawls's *A Theory of Justice* (and later, in *The Independence of Moral Theory*), but in his earlier work *Outline of a Decision Procedure for Ethics*, the approach is somewhat different. In this later work, Rawls articulates a method to validate or invalidate moral rules when competing interests or conflicts exist,[65] but here he seems to adopt a more foundationalist, principle-based approach. Consider what he says about evaluating particular cases:

> […] what is the test of whether a judgment in a particular case is rational? The answer 'is that a judgment in a particular case is evidenced to be rational by showing that, given the facts and the conflicting interests of the case, the judgment is capable of being explicated by a justifiable principle (or set of principles). Thus if the explicit and conscious adoption of a justifiable principle (or set of principles) can be, or could have been, the ground of the judgment, or *if the judgment expresses that preference which justifiable principles would yield if applied to the case, then the judgment is rational* [emphasis is mine]. Clearly the justification of particular judgments, if the above is correct, depends upon the use of justifiable principles.[66]

In this passage, Rawls seems to be expressing a view that various foundationalists would readily endorse: that the epistemic weight of a case-intuition is partly determined by whether it can be subsumed under a plausible principle. This is, in fact, a position defended by Nye, who I consider to be a foundationalist (see section 4.3), since he argues in favor of a similar principle-based approach: "to expose our ethical intuitions about particular cases to maximal critical scrutiny, we must determine whether they can be justified by directly plausible principles."[67] However, this is not what we would say in reflective equilibrium. In this method, a particular case (or case-intuition) has plausibility by itself—as long as it passes the previously established conditions, of course—although it can be outweighed if it

---

[65] John Rawls, "Outline of a Decision Procedure for Ethics," *The Philosophical Review* 60, no. 2 (1951): 177–97, p. 177.
[66] Rawls, p. 187.
[67] Nye, "Directly Plausible Principles", p. 610.

does not fit other cases or principles. In contrast, what Rawls seems to be saying in the quoted passage is that the plausibility of the case-intuition depends itself on whether the principle under which it can be subsumed is plausible.

In *Outline for a Decision…*, Rawls also mentions that the reasonableness of a principle is partly tested by looking at whether it remains plausible even if it conflicts with other particular or general judgments and, even more importantly, if it alters our confidence in those judgments. For example, assume we condemn others for doing wrong actions while at the same time believe that people "should not be morally condemned for the possession of characteristics which would not have been otherwise even if [they] had so chosen."[68] If this later principle remains plausible despite the conflict with our attitude about condemning others, and even alters it, then, according to Rawls, this makes the principle even more reasonable. Yet, this is not what we would say in reflective equilibrium. Recall that in reflective equilibrium the degree of inconsistency with other judgments is a *pro tanto* reason to revise a judgment. So, if our principle conflicts with our attitudes about condemning others, this would be a reason against our principle, not in favor. In other words, whereas the foundationalist Rawls would say, "this principle remains plausible despite the conflict, and that is a reason to favor it," the reflective equilibriumist Rawls would say, "this principle remains plausible despite the conflict, but this is not a reason to favor it, apart from the plausibility that the principle itself has."

Perhaps I am misreading Rawls here, and he is merely articulating reflective equilibrium differently (in *A Theory of Justice,* he mentions that he is following the ideas expressed in *Outline of a Decision…*). At any rate, due to the possibility of confusion, I will center our account of reflective equilibrium based on the ideas expressed mainly in *A Theory of Justice*.

1.3 Is reflective equilibrium a coherentist method?

So far, I have mentioned that reflective equilibrium involves consistency between judgments, but a contentious issue is also whether reflective equilibrium involves a

[68] Rawls, "Outline of a Decision Procedure for Ethics", p. 188.

coherentist view of epistemic justification. Let us refer to coherentism as the view that the mere fact that a set of judgments fits with each other provides a reason to endorse that set.[69] Under a coherentist account, the confidence we have in two judgments might increase if these judgments align with each other in particular ways, such as when "a general principle gives a particular judgement a satisfying rationale […] or when a particular judgement that is attractive in itself would be true if a given principle were true."[70] Thus, the judgments reinforce each other, which adds to the direct plausibility that they might independently have.

Many people take reflective equilibrium as a type of coherentist theory. Musschenga claims that "central to any coherence theory of justification in ethics is the method of reflective equilibrium."[71] Norman Daniels—probably the author who has written the most about reflective equilibrium—mentions that "an acceptable coherence requires that our beliefs not only be consistent with each other (a weak requirement), but that some of these beliefs provide support or provide a best explanation for others."[72] This characterization, in turn, has allowed various authors to criticize reflective equilibrium. Arguably, if the previous characterization is correct, then objections to coherence theories of epistemic justification will naturally extend to reflective equilibrium. And I suspect that coherentism is also why some authors think that reflective equilibrium is not revisionary enough: presumably, if consistent judgments reinforce each other, then it seems very hard to challenge the moral *status quo*.

But is there evidence to think that Rawls thought of reflective equilibrium as a coherence method? Daniels cites Rawls's *A Theory of Justice* when discussing this issue, and granted, Rawls does employ the term *coherence* on some occasions:

---

[69] This is meant as a theory of justification, rather than as a theory of truth. As Olsson explains, "The former is a theory of what it means for a belief or a set of beliefs to be justified, or for a subject to be justified in holding the belief or set of beliefs. The latter is a theory of what it means for a belief or proposition to be true." See Erik Olsson, "Coherentist Theories of Epistemic Justification," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Spring 2014 (Metaphysics Research Lab, Stanford University, 2014).

[70] Hurka, *British Ethical Theorists from Sidgwick to Ewing*, p. 126-127.

[71] Musschenga states it this way: "Central to any coherence theory of justification in ethics is the method of reflective equilibrium" (Musschenga, "Empirical Ethics, Context-Sensitivity, and Contextualism," p. 480). Also, see Lazari-Radek and Singer, *The Point of View of the Universe*, pp. 96-98.

[72] Daniels, "Reflective Equilibrium"; see also Norman Daniels, *Justice and Justification: Reflective Equilibrium in Theory and Practice* (Cambridge: Cambridge University Press, 1996), p. 2.

> A conception of justice cannot be deduced from self-evident premises or conditions on principles; instead, its justification is a matter of the mutual support of many considerations, of everything fitting together into one *coherent* view [emphasis is mine].[73]

In another work in which he discusses reflective equilibrium, *The Independence of Moral Theory*, Rawls also uses the term: "One tries to see how people would fit their various convictions into one *coherent* scheme [emphasis is mine]."[74] However, it is not apparent what Rawls means by *coherence*, as the term can mean various things which are often not adequately distinguished. More importantly—and putting aside interpretation issues of Rawls's work—it is not evident whether coherentism should be an essential characteristic of reflective equilibrium, and even if it were, in which way.

Coherence can be understood in two general ways. On the one hand, one can understand *coherence* as consistency: the absence of explicit contradiction.[75] We can refer to this type of coherence as "weak coherence." Under this definition, reflective equilibrium would ask us to assign credence to judgments only based on their degree of plausibility and their consistency with other plausible judgments. On the other hand, one can understand *coherence* as a positive connection between judgments, one which makes them "hang together," "align," or "fit," thus going beyond the mere absence on contradiction. This positive connection makes judgments reinforce each other, which occurs, for example, when one judgment gives another a satisfying rationale or an explanation. Bonjour, for instance, stresses that this positive connection involves the number and strength of inferential relations—i. e., when a judgment serves as a premise of a justificatory argument for a further belief.[76] Daniels mentions that judgments must "provide support or provide a

---

[73] Rawls, *A Theory of Justice*, p. 19.
[74] Rawls, "The Independence of Moral Theory", p. 8.
[75] Bonjour, however, claims that this is a "serious and perennial mistake," as consistency is a minimal but not sufficient requirement for coherence. See Laurence Bonjour, *The Structure of Empirical Knowledge* (Cambridge, Massachusetts: Harvard University Press, 1985), p. 95.
[76] Bonjour summarizes his view as follows: Firstly, "the coherence of a system of beliefs is increased by the presence of inferential connections between its component beliefs and increased in proportion to the number and strength of such connections." Secondly, "the coherence of a system of beliefs is diminished to the extent to which it is divided into subsystems of beliefs which are relatively unconnected to each other by inferential connections." See Bonjour, p. 98.

best explanation for others," and he often draws analogies to theory-acceptance in philosophy of science, so one might assume that he could appeal to certain views on explanation or inference to the best explanation that philosophers of science have defended.[77]

It is important to point out that a commitment to strong coherence does not entail a commitment to a full-blown *coherentist theory of epistemic justification*. This refers to the view that judgments are *ultimately* justified by the way in which they hang together to produce a coherent set. Davidson puts this nicely: "[w]hat distinguishes a coherence theory is simply the claim that nothing can count as a reason for a belief except another belief."[78] The commitments to strong coherence and coherentism are different because one can accept that strong coherence of judgments has probative value without claiming that this is the *only* source by which judgments can be justified. In such case, one would have to endorse strong coherence plus foundationalism (not to be confused with foundationalism in moral epistemology[79]): the view that some beliefs are basic, that is, that they need not be justified by appeal to other beliefs. Different versions of foundationalism differ on the degree of justification that these basic beliefs are held to possess. On the strong version of foundationalism—which has been historically more common—these beliefs must be absolutely certain or indubitable. On a more moderate version of foundationalism, the basic beliefs need not be self-evident or indubitable, as Descartes intended; they simply need to provide some degree of justification.[80] In this sense, one could consider reflective equilibrium as a moderate foundationalist theory, because the fundamental building blocks of our moral theory would be plausible yet corrigible case-intuition and principle-intuitions, providing credence to judgments without appeal to any other judgments.

I should mention briefly that Daniels would probably disagree with the claim that reflective equilibrium involves foundationalism, for he mentions that "wide reflective

---

[77] See Daniels, *Justice and Justification: Reflective Equilibrium in Theory and Practice*, p. 1, 24-25.
[78] Donald Davidson, "A Coherence Theory of Knowledge and Truth," in *Truth and Interpretation*, ed. E. LePore (Oxford: Blackwell, 1986), 307–19, p. 156.
[79] Foundationalism in moral epistemology can be considered a type of foundationalist theory, in the epistemological sense, but not vice versa, since one version of reflective equilibrium could also be considered a foundationalist theory, as I will explain in the following paragraphs.
[80] Bonjour, *The Structure of Empirical Knowledge*, pp. 26-27.

equilibrium [...] is not a standard form of moral intuitionism because it is not foundationalist."[81] In his characterization of foundationalism,

> some set of moral beliefs is picked out as basic or self-warranting [...] Some claim self-evidence or incorrigibility; others innateness; others some form of causal reliability. [...] Some intuitionists want to treat principles as basic. Others begin with particular intuitions."[82]

Daniels, however, seems to be understanding foundationalism too narrowly. Firstly, he seems to be suggesting that foundationalism treats as basic either principle-intuitions or case-intuitions. This is incorrect because one could take both principles and cases as basic and still be considered a foundationalist, as this last view refers to a form of justifying knowledge and not to the level of generality of the judgments being justified. The second and more important assumption is excluding the possibility that one can take judgments as both basic *and* revisable. Granted, more traditional foundationalist accounts try to ground theories in a set of self-evident principles or unshakeable case-intuitions. Daniels might have in mind authors like Descartes or Sidgwick—and those inspired by them—who try to come up with a set of axioms from which to derive their theories. But, as Bonjour and Olsson explain, foundationalism can also be cast as considering basic beliefs that provide direct *and* defeasible support,[83] and "to do so it may not have to appeal to self-evidence, indubitability or certainty."[84] Huemer also mentions—although in a discussion of intuitionism—that foundationalism does not imply that judgments are incorrigible or infallible: we can hold some judgments as starting points, *prima facie* justified, but revise them in case tensions arise with other justified beliefs.[85] This seems to be what goes on in reflective equilibrium, as we consider the direct plausibility of principle and case-intuitions although not take it as incorrigible.

One could be tempted to defend Daniels by saying that we can simply stick to a full-blown coherentist theory of epistemic justification and exclude appeals to direct

---

[81] Daniels, *Justice and Justification: Reflective Equilibrium in Theory and Practice*, p. 83.
[82] Daniels, p. 26.
[83] Bonjour, *The Structure of Empirical Knowledge*, p. 26-27.
[84] Olsson, "Coherentist Theories of Epistemic Justification."
[85] Michael Huemer, *Ethical Intuitionism* (Basingstoke [England]; New York: Palgrave Macmillan, 2005), pp. 106-107.

plausibility. But there are various reasons against doing so. Firstly, some argue that coherence theories necessarily involve some sort of foundationalism. Huemer argues that some traditional examples of the amplificatory, probabilistic effect of coherence are parasitic on the initial foundational justification.[86] Olsson, for its part, mentions that "influential coherence theorists […] [assign] some beliefs that are close to experience a special role," and that "these theories may be more fruitfully classified as versions of *weak foundationalism* than as pure coherence theories."[87] Secondly, there have been various objections against full-blown coherentism itself. Olsson, for instance, mentions that it is unclear how the fact that as a system is internally-coherent would allow us to approach truth or reality (assuming this matters to us), especially if it does not give experience any fundamental role. He also says that if there are equally coherent yet incompatible systems, then it is unclear how we would have reason to think we are approaching reality or truth.[88] Lastly, Daniels himself concedes that direct plausibility has to be included as a criterion for conferring either higher or lower credence to judgments: "coherence considerations in the moral case may be evidential in just the way they are in the sciences, though to make the case persuasively, some account of the initial credibility of moral judgments is owed."[89] Ignoring direct plausibility would conflict with Daniels's claim about *wide reflective equilibrium* in which we can revise moral judgments if they fail "to cohere with other, more plausible background theories."[90] Most likely, the plausibility of these background theories cannot be grounded on how well they cohere with moral judgments; otherwise, as Daniels points out, reflective equilibrium would have no revisionary force, as it would simply seek to match our current moral beliefs. [91] In fact, we would get into a problematic regress that coherentist accounts often face: are the background theories plausible because they support certain moral judgments which, in turn, are plausible because they are supported by these background theories? In order to avoid this problem, background theories would have to be justified by their internal coherence, by how well the judgments of the theory

[86] Michael Huemer, "Probability and Coherence Justification," *The Southern Journal of Philosophy* 35 (1997).
[87] Olsson, "Coherentist Theories of Epistemic Justification."
[88] Olsson.
[89] Daniels, *Justice and Justification: Reflective Equilibrium in Theory and Practice*, p. 61.
[90] Daniels, p. 28.
[91] Daniels, pp. 22-23.

itself align with each other. But this runs into what Olsson calls the *alternative systems objection*: conceivably, there are equally coherent yet incompatible background theories, and each of them, being coherent, would be justified.[92] Which one, then, should we use to revise moral judgments? If this is the case, it is unclear how Daniels's reflective equilibrium can be as revisionary as he wants it to.

Now that we have rejected full-blown coherentism as a possibility, let us return now to the central question of this section, which type of coherence does reflective equilibrium involve, weak or strong? Three reasons lead me to think that we should not take sides on this dispute and simply leave the door open for both approaches (in other words, that we should consider as reflective equilibrium any version that includes *at least* weak coherence, that is, consistency).

The first reason is that there does not seem to be any definite reason to think that strong coherence has probative value in ethics. Most of the arguments given in favor of strong coherence as having probative value are based on an inadequate analogy with the empirical case. Perhaps the most cited example is from C. I. Lewis.[93] He presents the following scenario: if we were interviewing various unreliable witnesses, and it happened that they were not collaborating and some of them told the same story, it makes it probable that their story is correct, because it would be improbable if they agreed. However, a significant problem with this example is that it is about empirical beliefs, not about moral intuitions, and there is a strong disanalogy.[94] In Lewis's case, we are talking about the testimony of eyewitnesses, which is based on sensory perception. But it is unclear whether this would apply to moral intuitions, which are not based on sensory perception. Presumably, in an analogous case, would we have a moral detective trying to figure out what is good or bad, who considers the intuitions of some people (the witnesses), and gives preference to an unlikely yet coinciding moral judgment. Yet, in this scenario, Lewis's

---

[92] Olsson, "Coherentist Theories of Epistemic Justification."

[93] C. I. Lewis, *An Analysis of Knowledge and Valuation* (La Salle: Open Court, 1962), p. 346. Bonjour also cites this example to defend strong coherence. Bonjour, *The Structure of Empirical Knowledge*.

[94] Perhaps the most cited example is from C. I. Lewis, presented in his *An Analysis of Knowledge and Valuation* (La Salle: Open Court, 1962), p. 346. Lewis's example is the following: If we were interviewing various unreliable witnesses, and it happened that they were not collaborating and they told the same story, it makes it probable that the story is correct, because it would be improbable if they agreed. Bonjour, in fact, uses this example to defend coherentism, but, again, his defense is of coherence about empirical knowledge, as is explicit from the title of his book, *The Structure of Empirical Knowledge*.

example seems to lose its intuitive appeal, presumably because there is no connection with sensory perception. Also, Lewis's example is about beliefs among people, not about the beliefs of one individual. For it to be analogous to moral intuitions, his example would need to apply to the conflicting beliefs of one individual and how the mere fact that some cohere gives them any additional credence.

The same could be said of a similar case presented by Huemer in defense of strong coherence in ethics. Michael Huemer asks us to imagine that if a detective interviews six eyewitnesses of a robbery, and two of them coincide in that the license plate of the getaway car is X78 41A, this adds force to their claim, since it would be extremely unlikely for two eyewitnesses to have provided the same license plate.[95] In his example, coherence presumably allows the detective to be confident on a license plate number even when there were other four six conflicting eye testimonies. But again, the example relates to sensory perception, which is disanalogous with intuitions in ethical inquiry.

The second reason not to include strong coherence in reflective equilibrium is that the notion of "strong coherence" in moral inquiry has not been completely clarified (this is even true for "strong coherence" in general, as Bonjour mentions).[96] If we consider Daniel's account, in which coherence involves judgments *explaining* each other, it is somewhat unclear how are we to understand explanation, beyond one's intuitive grasp of the concept.[97] Explaining cannot make reference to predictive power—as it is sometimes thought of in science—since there is no clear parallel in ethics. Also, explaining cannot refer straight-forwardly to "giving a satisfying rationale to a judgment," as it is not clear why would this be different from standard reflective equilibrium. For example, if I had the case-intuition *it is wrong to kill my pet and eat it*, presumably, an explanation involves asking what makes the intuition true or what makes it wrong to kill my pet and eat it. This, in turn, involves competing principles such as *Killing a sentient being is wrong* or *Killing pets is condemned by society*, but it seems that the best explanation is simply the principle which seems most plausible: in this case, *Killing a sentient being is wrong*. It is unclear, therefore,

---

[95] Huemer, "Revisionary Intuitionism", pp. 379-380.
[96] Bonjour, *The Structure of Empirical Knowledge*, p. 94.
[97] See Daniels, "Reflective Equilibrium." Moreover, the notion of coherence is not fully explained in another major work of Daniels: *Justice and Justification: Reflective Equilibrium in Theory and Practice.*

why "explaining" a judgment and "choosing the most independently plausible principles" would be any different. (We could undoubtedly pursue these issues further, and have lengthier discussions on explanation, probability, and inference, but it seems to me that, for reasons of parsimony, the burden of proof is on those who favor strong coherence).

The last reason not to include strong coherence in reflective equilibrium is that choosing between weak or strong coherence will likely not lead to a serious ethical disagreement about what principles will be favored or what ought to be done in practice.[98] Arguably, one possible difference is that strong coherence would allow us to dismiss a plausible principle-intuition which conflicts with a coherent set of slightly-plausible case-intuitions. Consider one of the principles proposed by Peter Singer: *If it is in our power to prevent something bad from happening, without thereby sacrificing anything of comparable moral importance, we ought, morally, to do it*.[99] Once the implications of this principle have been drawn out, we see that it conflicts with various case-intuitions: it presumably requires us to donate the money we would otherwise have spent in luxury items, to give it to people in need in poor countries; to pursue career paths we would otherwise not have taken, in order to do as much good as possible; or to spend time we would otherwise have spent with family or friends helping people in need. How are we to assess Singer's principle? Someone who is not sympathetic to Singer's principle might prefer strong coherence, because then he or she could argue that Singer's principle-intuition does not cohere with the set of case-intuitions. In fact, these case-intuitions could be articulated in a coherent whole under an alternative principle: *morality does not, in a world like ours, generally require very large sacrifices, even from rather well off people*.[100] However, Singer's objector could achieve the same outcome with weak coherence only. He could have argued that the principle is logically inconsistent with the set of case-intuitions, whose individual credences, if added together, result in a higher degree of plausibility. In other words, by considering the individual strength that each case-intuition has, as well as the plausibility

---

[98] Brandt and Daniels disagree with this claim and say that if we only take coherence as consistency, this is too weak, but they do not explain why this is the case. (See Brandt, *A Theory of the Good and the Right*, p. 20; Daniels, "Reflective Equilibrium").
[99] Singer, "Famine, Affluence, and Morality", p. 231.
[100] This is related to the so-called demandingness objection. (See Berkey, "The Demandingness of Morality: Toward a Reflective Equilibrium", p. 3017).

of the principle that groups them, we could add up the epistemic weight each of each our case-intuitions to dismiss the plausible principle-intuition. (This is not an endorsement or rejection Singer's position; it is only meant to illustrate the practical similarities of strong and weak coherence).

Another possible practical difference of strong coherence is that it might allow us to be more confident on some claims than weak coherence would. Consider again Huemer's example, in which we are asked to imagine that if a detective interviews six eyewitnesses of a robbery, and two of them coincide in that the license plate of the getaway car is X78 41A, then this adds force to their claim, since it would be extremely unlikely for two eyewitnesses to have provided the same license plate.[101] Granted, the detective's inference could be explained by the fact that matching eye testimonies "reinforce each other," as Huemer says, but a more parsimonious explanation relates to the detective's reliance on other underlying assumptions and their consistency. One assumption is that it is highly unlikely that two eye witnesses testimonies about a license plate number will coincide, so, in case they do, the best explanation for the conflict is not random chance, lack of attention, or memory-retrieval problems—as is the case with the other conflicting testimonies—but that their testimonies are correct. Therefore, the detective could have perhaps achieved the same outcome without a commitment to strong coherence.

In summary, my claim is that strong coherence should be a discretionary feature of reflective equilibrium. There seem to be no positive arguments in adopting it, and even if strong coherence does ultimately have evidential value, it is not clear what it amounts to in ethics and how it makes a practical difference as to what principles will be favored in practice. For this reason, any version of reflective equilibrium that includes at least weak coherence ought to be considered a version of this method, although I leave the door open for those who want to include strong coherence as an additional feature in their methodology.

[101] Huemer, "Revisionary Intuitionism", pp. 379-380.

Chapter 2 — Why Reflective Equilibrium?

Now that we have examined what the central characteristics of reflective equilibrium are, we still need a reason to adopt this method over others. In this chapter, I will present an argument in favor of it. If it is sound, then it gives a reason to deem reflective equilibrium as the default ethical methodology. Consequently, the burden of proof would then be on those who give higher credence to case-intuitions or to principles-intuitions or even those skeptical of moral intuitions. The argument involves four premises:

> *Premise 1:* If a judgment seems plausible to an agent—in other words, if an agent has an intuition regarding that judgment—then, other things being equal, an agent should assign credence to that judgment.
>
> *Premise 2:* Other things being equal, an agent should assign credence to a judgment proportional to the degree of plausibility that the judgment has for the agent.
>
> *Premise 3:* Judgments of all levels of generality can seem plausible to an equal degree.
>
> *Premise 4:* At the outset of inquiry, there is no reason to think that all case-intuitions or principle-intuitions are unreliable in a systematic way.
>
> *Conclusion:* Other things being equal, an agent should assign credence to judgments of all levels of generality—case-intuitions and principle-intuitions—depending on the degree of plausibility that those judgments have for the agent.

This argument supports some central characteristics of reflective equilibrium that we mentioned in the last chapter. Firstly, it supports the idea that we can take into account any judgment, regardless of its level of generality (particular or general) and type (moral or non-moral). Secondly, it supports the idea that judgments can be revised depending on how plausible they seem. Thirdly, it implies that even if some intuitions turn out to be unreliable, not all case-intuitions or principle-intuitions are (in contrast, the foundationalist will say that all case-intuitions are unreliable in a systematic way). Lastly, although the argument does not support the idea judgments should be consistent with each

other, this consideration is commonplace in philosophy, so it seems unnecessary to include it as a further premise. In the following subsections, I defend each of the premises of my argument.

2.1 Subjective plausibility as justification

The first premise is *that if a judgment seems plausible to an agent—in other words, if an agent has an intuition regarding that judgment—then, other things being equal, an agent should assign credence to that judgment*.[102] One might think of this claim as following analytically from the way I defined *intuition*. In the Introduction, I proposed that we understand *intuition* as a mental state or disposition which, at the outset of inquiry, provides gradual and non-inferential justification to judgments. Thus, it would be analytically true to say that an intuition provides justification to a judgment and therefore that an agent should assign credence to it. This way of proceeding, however, looks trivial, as we could always propose a different definition for *intuition*.

We can defend the premise with two alternative arguments. The first argument is a *reductio*. If it were not the case that we had reason to give credence to judgments which seem true or plausible to us, then an agent would act in irrational ways. This is because an agent would be justified in holding a judgment which, overall, seems less plausible to him over one that seems more plausible to him. He would be justified in saying, "it seems to me that *X* is true, although I confer zero credence to this idea; in contrast, I confer 100% credence to *Y*, although it does not seem plausible to me." Arguably, the agent might do this if he had a further reason (*Z*) for holding the less plausible judgment (*Y)*. However, if the agent did, then presumably he accepts this further reason because it seems more plausible, so therefore the agent is still conferring credence to what seems plausible. Otherwise, it would be irrational for the agent to say, "I confer 100% credence to *Y* because of *Z*, although *Z*—just as *Y*—does not seem plausible to me."

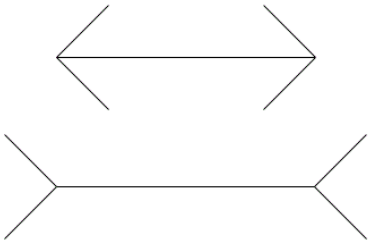
[102] This claim is similar to a principle introduced by Huemer called Phenomenal Conservatism, which states that "other things being equal, it is reasonable to assume that things are the way they appear." I formulated it slightly different just to avoid the use of the term *reasonable*. See Huemer, *Ethical Intuitionism*, p. 99.

The second argument is also a *reduction*, and it is that if we were to reject the first premise, then it seems that this would undermine a considerable part of ethical inquiry and, more generally, of philosophical methodology. In ethics, we mostly rely on intuitions about cases or principles to evaluate moral theories and practical issues. In fact, the alternative positions to reflective equilibrium that I have presented (foundationalism and anti-theory) do not deny that agents should assign credence to what seems plausible to them; they simply differ on which kinds of plausible judgments they think we should consider. And even if it is possible to minimize our appeal to intuition, it seems hard to have an ethical methodology that does not use them at all. More generally, in philosophical methodology, assigning credence to what seems plausible is a central feature, for example, in thought experiments, which rely on the intuitions they elicit and are often used to support or criticize a particular view. If someone working in one specific area of philosophy believes that agents should assign credence partly based on their intuitions about though experiments, then there is no initial reason to assume why this would not be permissible to do so in other areas of philosophy as well.

The underlying idea behind the two previous arguments is that a denial of the first premise is ultimately *self-defeating*. As Huemer mentions in defense of a principle similar in nature to the first premise, "All judgments are based upon how things seem to the judging […] Even the arguments of a philosophical skeptic who says we aren't justified in believing anything rest upon the skeptic's own beliefs, which are based upon what seems to the skeptic to be true." In other words, one cannot avoid appealing to what seems plausible, for a rejection of the first premise would be based on ideas which themselves seem plausible.

Huemer discusses a possible counterexample to the idea that we should assign credence to what seems plausible, which are optical illusions.[103] These illusions present to us a scenario which seems plausible but which we can come to see that it is not. Consider the Müller-Lyer illusion:

[103] Huemer, p. 100.



Initially, we might think that the bottom line is larger than the top one, but once we measure them, we come to see that they are of the same length. One might then be inclined to think that we should not always assign credence that what seems plausible. The problem, however, is that the revision we made is itself based on other beliefs which seem plausible: in this case, we thought that measuring was a better way to determine the length of the line than our glance at the lines. So, even in this scenario, there was no escape from assigning credence to what seems plausible—just a confirmation that assigning credence to judgments which seem plausible does not mean that we have to believe them, since they can be overruled by other plausible judgments.

A possible objection to the first premise comes from Brandt, who criticizes the appeal to intuitions in reflective equilibrium and—in reply to the second reason I offered in favor of my premise—proposes an alternative methodology. When criticizing reflective equilibrium, Brandt mentions the following:

> There is a problem here quite similar to that which faces the traditional coherence theory of justification of belief: that the theory claims that a more coherent system of beliefs is better justified than a less coherent one, but there is no reason to think this claim is true unless some of the beliefs are initially credible. […] No reason has been offered why we should think the initial credence levels, for a person, correspond to credibilities. The fact that a person has a firm normative conviction gives that belief a status no better than fiction. […] It is puzzling why an intuition—a normative conviction—should be supposed to be a test of anything.[104]

Putting aside the claim that reflective equilibrium is a coherentist method, a point that Brandt seems to be making here is that the fact that if I strongly believe something or think

[104] Brandt, *A Theory of the Good and the Right*, pp. 20-21.

something is highly plausible does not guarantee that my belief is true (and therefore Brandt's contrast with fiction). Brandt is asking for a positive argument as to why we should place credence on moral intuitions in the first place, and he does not think that reflective equilibrium by itself can provide it.[105]

The main problem with Brandt's objection is that it leads to global skepticism, one that is self-defeating for his own account. His objection to the use of moral intuitions is not only directed to reflective equilibrium but to any ethical methodology that uses intuitions, be it anti-theory, virtue ethics, foundationalism, or any other. In fact, in saying that "the fact that a person has a firm normative conviction gives that belief a status no better than fiction,"[106] his objection extends to any philosophical method that uses intuitions, unless one is able to show that the intuitions one relies on are, in some way, reliable. Establishing that certain intuitions are reliable, however, must be made by appeal to other normative convictions. As Huemer mentions, if we assume that nothing can be accepted until we first give a positive reason for trusting that kind of belief, this applies to any other source, be it perception, memory, reason, or anything else.[107]

For instance, Brandt's own theory claims that we should use *facts* to maximally scrutinize our moral views.[108] Presumably, facts (or non-moral observations) are credible because one can tell a causal story about how they originate, even if they can sometimes be affected by cognitive biases or perceptual illusions,[109] whereas moral intuitions do not seem to have this causal story. But the problem is that the criterion by which we decide to rely on facts or nonmoral observations is not itself based on a "fact" but on a normative conviction. The idea that providing a causal story makes an observation count as a "fact" cannot be grounded on a fact itself or it would be begging the question: it must be directly plausible or based on other normative convictions which are directly plausible. The same point applies for any other criterion one proposes for why non-moral observations or facts are reliable.

---

[105] Brandt also mentions other problems with using moral intuitions, but they are similar to the ones we will discuss in Part II, so I will put them aside for now
[106] Brandt, *A Theory of the Good and the Right*, p. 20.
[107] Huemer, *Ethical Intuitionism*, p. 107.
[108] See Brandt, *A Theory of the Good and the Right*, p. 13.
[109] A similar point is made by Daniels, who thinks this analogy is inappropriate, in *Justice and Justification: Reflective Equilibrium in Theory and Practice*, p. 31.

Objecting to the use of normative convictions would also be problematic if one thinks of areas such as logic, metaphysics, or epistemology, which have to appeal to intuitions and normative convictions but still we do not consider their body of knowledge as mere fictions. Why, then, would using intuitions or normative convictions in these areas be any different than using moral intuitions or normative convictions in ethics?

Therefore, Brandt's objection cannot be that normative convictions, *in general*, are not a test of anything, for this would undermine the criteria by which nonmoral observations are taken to be credible (and his own theory, more generally). He cannot escape the use or intuitions or normative convictions altogether; he can only hope to show that some of them are unreliable or problematic. His argument needs to be that moral intuitions *in particular* are not a test of anything. If this is the case, however, then we are not talking about an objection to the first premise, about plausibility as a reason to assign credence to a judgment. We would be in the domain of the objection made by foundationalist to the *other things being equal* clause, in saying that things are not equal with respect to moral intuitions, and this I shall address in Part II.

2.2 Degree of plausibility as degree of justification

In addition to conferring credence to judgments that seem plausible, the second premise establishes that *other things being equal, an agent should assign credence to a judgment proportional to the degree of plausibility that the judgment has for the agent*. This consideration seems straight-forward. If one thinks that credence should be assigned partly based on what seems plausible—as the first premise established—it is natural to extend this idea and say that if something seems more plausible, then one should assign higher credence to it. We can refer to the degree of plausibility of a judgment also as the *strength* of the intuition supporting that judgment.

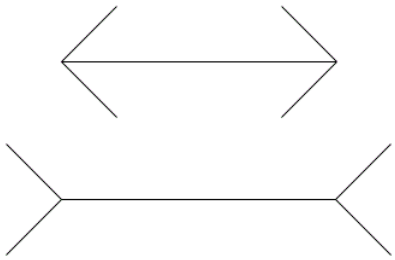
Degree of plausibility involves a spectrum, one in which judgments can go from seeing absurd to highly plausible,[110] and this is demonstrated by our everyday and

[110] Some intuitionists might say that this extreme of high plausibility is what is called self-evidence, although different accounts of what self-evidence might involve, see Philip Stratton-Lake, "Intuitionism in Ethics," in

philosophical experience. Some judgments seem compelling to a certain extent, but there can be others which are way more compelling, slightly more compelling, slightly less compelling, or not compelling at all. Additionally, we readily confer higher credence to ideas which seem highly plausible—it would be surprising not to do so. The principle of non-contradiction,[111] for instance, seems highly plausible, and even indubitable to the majority of people, so it seems natural to assign a high credence to it and to be reluctant to abandon it even if something contradicts it (*pace* the redundance).

Support for this premise can also be made on similar grounds than those used to defend the first premise. It would seem irrational for an agent to assign higher credence to what seems less plausible than other alternatives. If this were the case, an agent would be justified in saying "it seems to me that *X* is almost certainly true, although I confer less credence to it than to *Y*, which does not seem at all plausible." Presumably, if the agent did this, it is because of a further reason, *Z*, which is more plausible than *X*, but then, the agent would be giving higher credence to what seems plausible.

Consider again the Müller-Lyer illusion:



In this case, it might seem that we are not assigning credence to what seems initially plausible, which is that the bottom line is longer than the top one. But the fact is that, although we could assign some initial credence to it, we are assigning a higher credence to another idea—that measuring the lines or overlapping them is a more reliable way to determine their length than glancing at them—and that the plausibility of our initial idea

The Stanford Encyclopedia of Philosophy, ed. Edward N. Zalta, Winter 2016 (Metaphysics Research Lab, Stanford University, 2016).

[111] One simple and intuitive formulation of the principle is this: *p* and *not-p* cannot be both true at the same time.

might have changed based on other idea we considered. Thus, the degree to which something seems plausible might be overruled by other plausible judgments and it might also change once we consider other judgements. So, in the end, we are still assigning credence proportional to what seems plausible to us.

A possible objection to the second premise comes from Audi, who says that, although it is true that intuitions can provide either weak or strong evidential weight, we should only talk about intuitions when one firmly believes something. He writes,

> A mere inclination to believe is not an intuition; an intuition tends to be a "conviction." […]. Granted, some intuitions are easily overcome by doubts or counter-evidences, and certainly a proposition one is only inclined to believe may be or seem intuitive. Still, one does not have an intuition with that proposition as its content until one believes the proposition.[112]

We should distinguish, however, whether Audi's attack is aimed at the use of the word *intuition* or at the idea that semi-strong or weaker beliefs should not provide evidential weight. If Audi's idea is aimed at the former, at use of the word *intuition*, then this would leave unchallenged the second premise, as it would merely refer to a verbal dispute. His usage, nonetheless, would still remain problematic for additional reasons. First, there does not seem to be a clear-cut line between weak and strong intuitions, which, otherwise, might justify Audi's usage. Given that we can have mixed or semi-strong intuitions, it seems contrary to our experience with thought experiments and cases to say that they trigger either a weak or a strong intuition. Indeed, this makes it arbitrary to call someone's reaction an intuition if is strong enough while calling someone else's reaction an inclination if it is only semi-strong. More importantly, Audi's usage is not parsimonious, given that he would be introducing a distinction between intuitions and inclinations which plays no theoretical role in justification. If he concedes that plausibility is a reason to assign credence, be it weak or strong, it seems unnecessary to make a conceptual distinction between inclinations and intuitions.

If Audi's objection is to have any force, it must be stating, rather, that semi-strong or weaker beliefs should not provide evidential weight, only strong beliefs. One reason in

---

[112] Audi, *The Good in the Right: A Theory of Intuition and Intrinsic Value*, p. 34.

favor of this view might be that weaker intuitions do not seem to have a special role in substantive ethics, whereas strong intuitions do. Rawls, in fact, mentions that "we can discard those judgments made with hesitation, or in which we have little confidence."[113] Daniels makes a similar remark: "we begin by collecting the person's initial moral judgments and filter them to include only those of which he is relatively confident."[114] However, if the charge is that weaker intuitions have not played a special role in substantive ethics, this does not imply that they cannot. If they have not yet, it might be precisely because they are being excluded from the beginning; but if we accept reflective equilibrium, there is no reason why they would be. The objection must then be something besides saying that weak intuitions have no role to play, because, clearly, in reflective equilibrium they can. In fact, I suspect why Rawls and Daniels exclude these intuitions is not because they oppose them in principle—reflective equilibrium offers no reason why we should—but merely as a heuristic strategy: we save time and effort if we focus only on the most plausible judgments. Rawls hints at this when saying, "we cannot examine each of [possible description and philosophically relevant argument]. The most we can do is to study the conceptions of justice known to us through the tradition of moral philosophy and any further ones that occur to us."[115]

Audi offers another reason as to why we could give less credence to weak intuitions, which is that they are not clear enough:

> We might speak of intuitive inclinations as opposed to intuitions, and the former need not be denied some degree of evidential weight. But it would be less than that of intuitions proper: the data would be less clear, just as a view of an unexpected island in the fog is less clear than it would be in sunlight and provides less reason to alter one's map.[116]

Audi's suggestion here just seems amiss. It seems to me that I can correctly understand a judgment but still think that it has weak intuitive appeal. Consider the judgment *yellow things are good*. This is an extremely implausible judgment, but we can entirely understand

---

[113] Rawls, *A Theory of Justice*, p. 42.
[114] Daniels, *Justice and Justification: Reflective Equilibrium in Theory and Practice*, p. 22.
[115] Rawls, *A Theory of Justice*, p. 43.
[116] Audi, *The Good in the Right: A Theory of Intuition and Intrinsic Value*, p. 34.

what it means. Contrarily, I could have a vague principle which, nonetheless, can be highly plausible, such as the principle *we should aim for what is best for us*. This principle is vague because *what is best for us* could mean many things, but it still seems trivially plausible. Therefore, if we gives less credence to a weak intuition, then it cannot be because of its clarity, but because it is implausible.

2.3 Plausibility of judgments of all levels of generality to an equal degree

The third premise of my argument is that judgments of all levels of generality can seem plausible to an equal degree. This premise implies that the content of intuitions does not need to have a specific level of generality, and in this sense, both particular cases and general principles can trigger plausible intuitions. It also implies that, in principle, both particular cases and general principles can trigger intuitions that fall within the spectrum going from absurd to highly plausible.

The fact that judgments of all levels of generality can seem plausible gains support from our experience. On the one hand, particular cases can undoubtedly generate intuitions: any thought experiment or anecdote that elicits a moral response serves as an example. Even authors who dismiss case-intuitions in their methodologies do not deny that case-intuitions can seem plausible; what they deny is that this seeming is unreliable or that there are other more plausible intuitions that count against them (e.g., that they are systematically unreliable). Also, it seems true that case-intuitions can seem highly plausible and be met with almost unanimous agreement. Take the following example:

> *The Happiness Button.* Next week, a stranger will accidentally trip on an uneven sidewalk and break her kneecap. This will be extremely painful and will significantly reduce her happiness for several months. However, if you press a button (the happiness button) a little bit of magic will make her more attentive as she's walking along, and she won't break her kneecap. Will you push the button?[117]

[117] This is an example adapted from Joshua Greene, *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them.* (New York: The Penguin Press, 2013), pp. 190-191.

I believe most people would press the button, unless one is a psychopath or modifies the thought experiment to include additional assumptions such as that the person will benefit in some way from the pain experienced. But under normal circumstances, this case-intuition seems highly plausible.

On the other hand, principles also seem capable of triggering intuitions. There are plenty of examples of principles which trigger an intuition in favor of them:[118]

> p and not-p cannot be both true at the same time.
> If *A* is better than *B*, and *B* is better than *C*, then *A* is better than *C*.
> Enjoyment is better than suffering
> It is wrong to inflict enormous harm for relatively trivial benefits.

All these principles seem not only plausible, but highly plausible, even if there are additional reasons as to why we would reject them.[119]

One possible objection to the third premise might be that, even if both types of intuitions can be plausible, maybe case-intuitions cannot be *as* plausible as some principle-intuitions—or vice versa—so it is false that they can seem plausible to an equal degree. Yet, this seems empirically false. The previous thought experiment of *The Happiness Button* serves as an example. Also, our everyday, political, and scientific awareness of how people change their moral views serves as a counterexample, because changes in moral beliefs usually occurs by appeal to moving narratives, shocking images, and personal faces and interactions.[120]

A second objection could be that, even if we accept that both cases and principles can seem plausible to an equal degree, perhaps some type of intuitions will be, in the end, consistently more plausible than their counterparts. In other words, while conceding that all intuitions can seem initially plausible, one could say that either case or principle-

---

[118] These examples are taken from Pust, "Intuition."
[119] Temkin, for example, argues against the second principle about transitivity. See L. Temkin, "Intransitivity and the Mere Addition Paradox," *Philosophy and Public Affairs* 16 (1987): 138–87.
[120] See, for example, Jonathan Haidt, *The Righteous Mind: Why Good People Are Divided by Politics and Religion* (New York: Pantheon Books, 2012). Moreover, at a social level, moral change seems to occur not by "pure" moral argumentation, as Anderson argues, but by "a variety of other ways of making interpersonal claims, including petitions, hearings, testimonials, election campaigns, voting, bargaining, litigation, demonstrations, strikes, disobedience, and rebellion." See Elizabeth Anderson, "The Social Epistemology of Morality," in *The Epistemic Life of Groups: Essays in the Epistemology of Collectives* (Oxford: Oxford University Press, 2016).

intuition will not remain plausible once we examine other possible judgments and philosophically relevant theories. It could be, for example, that principle-intuitions are, at the end of ethical inquiry, the ones that seem most plausible[121] or, contrarily, that case-intuitions are more plausible.[122] If this is correct, this would give us a reason to give higher epistemic weight to certain type of intuitions. However, whether case or principle-intuitions will remain plausible after reflection is something that remains to be determined. We cannot know in advance what will be the case, unless one gives a different argument as to why either of this type of intuitions will not remain plausible. Also, this objection is not strictly in conflict with reflective equilibrium. Determining whether case or principle-intuitions will remain plausible will involve a method, and this method will probably be some version of reflective equilibrium. So, the objection seems to pressupose reflective equilibrium, even if we ultimately reach a point where we give higher weight to a certain type of intuitions. As Rawls mentions,

> One's moral conception may turn out be based on self-evident first principles. The procedure of reflective equilibrium does not, by itself, exclude this possibility, however likely it may be. For in the course of achieving this state, it is possible that first principles should be formulated that seem so compelling that they lead us to revise all previous and subsequent judgments inconsistent with them.[123]

2.4 Reliability of case-intuitions and principle-intuitions

The fourth premise of the argument is perhaps the most contentious one: that *at the outset of inquiry, there is no reason to think that case-intuitions or principle-intuitions are unreliable in a systematic way*. Without this premise, a foundationalist might agree with the rest of my argument, but object that case-intuitions, as plausible as they might be, are not reliable. Thus, were it not for this premise, one could challenge the "other things being

---

[121] This seems to be the strategy that intuitionists like Sidgwick and Ross follow, because they claim that only some principles appear to be self-evident (a term which I interpret as the strongest possible degree of plausibility).
[122] This is suggested by Shelly Kagan in "Thinking about Cases," in *Moral Knowledge*, ed. Ellen Frankel Paul, Fred. D. Miller, and Jeffrey Paul (Cambridge: Cambridge University Press, 2001).
[123] Rawls, "The Independence of Moral Theory", p. 8.

equal" clauses of the first and second premises and say that, in practice, things can never remain equal for case-intuitions, as they are systematically unreliable. (Note that one could also raise the same worry about principle-intuitions, as anti-theorists and particularists have, but, as I mentioned in the Introduction, I will limit my discussion to the foundationalist objections against case-intuitions).

The fourth premise does not deny that *some* intuitions are unreliable. It could indeed be true that some case-intuitions or principle-intuitions are not trustworthy. Reflective equilibrium is not incompatible with this claim, since we could simply filter out the unreliable intuitions and consider only those that pass the test—as both Rawls and Daniels think we should—and still give equal consideration to both case-intuitions and principle-intuitions. Therefore, for premise four to be challenged, the foundationalist has to show that there is reason to think that *all* case-intuitions are unreliable, as this would justify consistently giving higher epistemic weight to principle-intuitions over case-intuitions. Peter Singer, one of the chief critics of reflective equilibrium, makes a remark along this line: "If the interpretation [of reflective equilibrium] is truly wide enough to countenance the rejection of all our ordinary moral beliefs, then I have no objection to it."[124]

The plausibility of the fourth premise relies on the fact that it seems legitimate to appeal to case-intuitions in an ethical discussion to support or undermine a position. In fact, it is hard to think of an ethical debate that does not involve appeals of this kind. If one presented us with a moral principle, it seems like a legitimate move to mention a counter-example to undermine it, an instance in which the principle would not apply. Similarly, it seems reasonable if a person supports a moral principle by presenting to us relevant instances in which the principle would adequately apply. Therefore, there seems to be no initial reason as to why we would distrust appeals of this kind.

Granted, one could argue that there are some situations in which case-intuitions are clearly unreliable. For example, if a slaveholder were debating the morality of slavery, one could be suspicious—beyond worries of insincerity—about the case-intuitions he might have, as his economic and social interests are at stake and he might not approach all claims

[124] Peter Singer, "Ethics and Intuitions," *The Journal of Ethics* 1, no. 9 (2005): 331–52, p. 347.

with an open mind. However, cases like these (which can be many) will only allow us to conclude that *some* case-intuitions are unreliable; they do not force us to conclude that *all* case-intuitions are unreliable. We need an argument of a different kind to make this claim.

What reason could there be for thinking that case-intuitions are ubiquitously unreliable? Foundationalists typically appeal to the untrustworthy origin or built-in bias of case-intuitions. As these challenges are the most common and perhaps the most serious ones, I will devote the next two chapters to address them. For the remainder of this section, let us consider another possible argument: that case-intuitions are unreliable not because of the intrinsic characteristics they might have, but because of their distribution among individuals. One could make the case that case-intuitions are unreliable because they are not universally shared (although "unreliable" might not be the best term but perhaps "inadequate" or "unhelpful"). In this sense, even if it is possible that judgments of all levels of generality can be plausible to an equal degree, there might not be overlap in the intuitions that people take as plausible. Presumably, this would be problematic because it seems unjustified for a moral theory to apply to those who cannot even hypothetically accept it, in which case, it would apply to those who do not share the same case-intuitions. The contractarian tradition often emphasizes a similar point: that for a moral theory to be justified, it must be able to be publicly recognized, possible to accept by those to whom it will apply.[125] And if it is true that case-intuitions are not universally shared, we would then have a reason to give lesser epistemic weight to them.

By itself, this objection poses not only a challenge to reflective equilibrium, but to any other method that uses intuitions, as it could be that other types of intuitions are also not universally shared. Then, for the previous objection to be successful without otherwise undermining alternative positions, one would need to add an additional premise: that principle-intuitions are those types of intuitions that are universally accessible. This claim, combined with the one presented in the previous paragraph, would indeed give the

[125] It is important to emphasize that for the moral theory to be problematic it has to apply to individuals who would not even hypothetically accept it, for one could think of a scenario in which we are deliberately working within certain framework or background assumptions. For example, a Christian community might find value in discussing their moral views and trying to build a theory even if their theory was not accessible to non-Christians. In this case, the theory would not be problematic if it is intended to apply only to the Christian community; if it were to be extended to non-Christians, then it could be considered problematic.

foundationalist a reason to give higher epistemic weight to principle-intuitions over case-intuitions.

However, I believe the universality argument against reflective equilibrium is ultimately unsuccessful, as it has numerous problems. Let us examine four of them. First and foremost, it is unclear whether some case-intuitions cannot be universally shared. For example, the thought experiment of *The Happiness Button*[126] that I presented in the previous section seems to me to be a strong candidate and one of which people would share a similar intuition. It may be objected that this case-intuition or any other cannot be universal because there will always be an individual who does not share it: a psychopath, for instance, could have a different reaction to *The Happiness Button*. But there are two possible replies to this objection. The first is that individuals might be confused, and they might not have thought carefully about the case-intuition. As the expression "being accessible" conveys, individuals do not need to currently accept the intuition; it only needs to be possible for them to find it plausible at some point, if they were to continue reflecting on it. This makes it possible that the case-intuition might be found plausible at some point, even if it is not currently shared. The second reply is that an intuition being "universally accessible" does not equal it being "accessible to each and every single individual." If this were so, it would require only one dissenting opinion—of a past, present, or future individual—to dismiss the intuition, and this is too restrictive even for the strongest principle-intuitions. To illustrate this, take the principle of non-contradiction, which I consider one of the strongest intuitions. [127] There surely is (or was or will be) an individual who does not find this principle plausible, maybe someone with the most severe form of schizophrenia. And we could expand this consideration beyond our species: the claim would also have to apply to any conscious creature in the universe with the capacity to think, if universality is to be taken in the strongest possible sense. So, given that this standard seems unachievable, I interpret that an intuition being "universally accessible"

---

[126] "Next week, a stranger will accidentally trip on an uneven sidewalk and break her kneecap. This will be extremely painful and will significantly reduce her happiness for several months. However, if you press a button (the happiness button) a little bit of magic will make her more attentive as she's walking along, and she won't break her kneecap. Will you push the button?" Adapted from Greene, *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them.*, pp. 190-191.

[127] In a simple formulation, the claim that p and not-p cannot be both true at the same time.

refers, instead, to the claim that it is accessible to a clear majority of individuals, even if there are some who do not share it. If this is correct, then it is not obvious, at least, that highly plausible case-intuitions would fare worse than principle-intuitions in terms of how widespread they can be.

Secondly, it is not even obvious that principle-intuitions are universally accessible, as the foundationalist might claim. Presumably, candidate universal principle-intuitions involve those abstract, theoretical judgments, such as the principle of non-contradiction. In the realm of ethics, candidates include what Huemer calls "formal intuitions," intuitions which impose formal constraints on ethical theories although they do not provide a positive or negative moral evaluation:

> If *x* is better than *y* and *y* is better than *z*, then *x* is better than *z*.
> If it is permissible to do *x*, and it is permissible to do *y* given that one does *x*, then it is permissible to do both *x* and *y*.

I am inclined to think that this kind of principle-intuitions might be universal and that most people would indeed accept them.[128] Of course, it only needs to be possible for them to find them plausible at some point, if they were to continue reflecting on them, as the expression "being accessible" conveys. However, although I concede that some principle-intuitions might be universal, it is unclear whether the same would be true for the vast majority of principle-intuitions, those unrelated to formal relations between propositions, actions, or attributes. After all, "formal intuitions," as plausible as they might be, might not be informative enough to justify more substantive conclusions in normative or practical ethics: it seems hard indeed to build a moral theory by appeal to transitivity and other highly abstract notions. Once we start addressing more substantive principle-intuitions—such as Sidgwick's maxim of benevolence ("each one is morally bound to regard the good of any other individual as much as his own"[129]) or some of Ross's *prima facie* duties (e.g. a

---

[128] Although it should be noted that Larry Temkin has mounted a strong critique of the transitivity principle. See L. S. Temkin, *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning* (New York: Oxford University Press, 2012).

[129] The complete formulation of the maxim is "each one is morally bound to regard the good of any other individual as much as his own, except in so far as he judges it to be less, when impartially viewed, or less certainly knowable or attainable by him. Lazari-Radek and Singer, *The Point of View of the Universe*, p. 119.

duty to keep our promises or to act to right a previous wrong we have done)—it is less clear the extent to which one can be confident about their universal character. Recall, these intuitions have to be accessible to a clear majority of individuals.

Thirdly, even if some principle-intuitions are universally accessible, probably some are not, and so we would have no reason to give systematic preference to principle-intuitions over case-intuitions. Given the sharp moral differences between individuals, communities, and cultures and the historical disagreement between moral philosophers, it seems likely that some principle-intuitions might not be universally accessible. In fact, I take it that foundationalists will not be concerned with defending that all principle-intuitions are universally accessible, but with finding those that might be accessible and plausible. Yet, if this is true, then there would no basis for giving systematic higher credence to *all* principle-intuitions over case-intuitions. It could well be that some principle-intuitions are more accessible, but some case-intuitions will undoubtedly be more accessible than the less accessible principle-intuitions. Thus, the decision of which intuitions to consider would have to be made on a case-by-case basis, depending on whether a certain intuition might be universally accessible or not. The level of generality would not be then what makes an intuition universal or not; rather, it would be the content of the intuition what makes the difference. Consequently, the fourth premise would remain unchallenged, as then we would have no reason at the outset of inquiry to systematically dismiss all case-intuitions.

Lastly, and similar to a point I made in the previous section, although it could well be that some principle-intutions remain universally plausible after reflection, this is not obvious at the outset of inquiry. Individuals still need to reflect, think carefully, and evaluate candidate principle-intuitions to see whether they remain ubiquitously plausible, and this will involve a version of the method of reflective equilibrium. Therefore, even if we ultimately realize that case-intuitions are not universally accessible and that some principle-intuitions are, we still cannot avoid considering case-intuitions to reach that point.

With this defense of the fourth premise, I conclude Part I, which presented a positive argument in favor of reflective equilibrium as the default methodology in ethics. There still remain some loose ends that we need to address, so let us turn now to Part II, which

examines two of the central arguments that foundationalist have presented against the systematic use of case-intuitions.

PART II: Challenges to Reflective Equilibrium

Chapter 3 — The Unreliable Sources of Case-intuitions

This chapter addresses perhaps the leading and most serious challenge that foundationalists have raised against reflective equilibrium: that some sources of case-intuitions are systematically unreliable. In the next Chapter, the fourth one, I will address an objection related to the stability of case-intuitions: the idea that our case-intuitions are unreliable because they change in identical or similar cases that only differ in some arguably morally irrelevant factors, such as their wording, the emotions we experience while thinking about them, or the mere passing of time. Both chapters are connected in a certain way, because arguments that appeal to the source of an intuition, as we will see, imply that what is problematic about case-intuitions is that they are arbitrary: culture and evolution could have turned out differently, leading us to hold different intuitions. For now, let us examine these two issues separately.

The challenge related to the source of case-intuitions comes in various forms, depending on which source exactly are we referring to. One claim is that the source of case-intuitions is an unreliable emotion or emotional process. The idea here is that case-intuitions are correlated with higher activity in parts of the brain associated with emotional processes, that they occur under certain emotional mental states (e.g., being angry), or that they are driven by strong emotional responses.[130] These facts, according to objectors, make case-intuitions unreliable. A second claim is that the source of case-intuitions is cultural indoctrination. The idea here is that case-intuitions are likely shaped by obsolete religious systems, outdated customs, habits, our parents, or our society. This, in turn, could lead us to suspect the reliability of case-intuitions.[131] One last claim is that the source of case-intuitions is an unreliable evolutionary process. The idea here is that case-intuitions are the result of adaptations throughout our evolutionary history which were useful for some

---

[130] Singer suggests various of these objections, although he sometimes intertwines them with other arguments. For instance, he mentions that emotionally-based responses are problematic because they are the result of evolution and because they are unstable across cases with varying morally irrelevant features. So, it sometimes unclear whether emotions are problematic by themselves or because of a further reason. See Singer, "Ethics and Intuitions." Greene also makes a similar point, and although his arguments are directed towards deontological theories, they can be extended to case-intuitions (as Singer does based on Greene's discussion). See Joshua Greene, "The Secret Joke of Kant's Soul," in *Moral Psychology, Vol. 3* (MIT Press, 2008).
[131] See Peter Singer, "Sidgwick and Reflective Equilibrium," *The Monist* 58, no. 3 (1974): 490–517, p. 516-517.

purpose. Thus, if case-intuitions are motivated by arbitrary goals set by evolution, they cannot offer a reliable guide.[132] As I mentioned before, any of these challenges needs to give us reason to suspect case-intuitions in a systematic way. It is not enough for them to show that only some of them are unreliable, for this would give us no reason to confer consistent and higher epistemic weight to principle-intuitions over case-intuitions, as the foundationalist wants. Rather, it would give reason to the reflective equilibriumist to filter some case-intuitions but keep his methodological assumptions intact.

Arguments that appeal to the source of a judgment face two initial problems, which will be relevant for the remainder of the chapter. The first problem is that these arguments can be self-defeating. As Sidgwick points out, every belief has an antecedent source, including the premises of the argument that appeals to the source of judgments, so if pointing out the source of a judgment is enough to undermine it, then the premises of one's own argument would be undermined.[133] (Of course, one could try to argue that some of our beliefs have no antecedent causes, but this would be hard to defend). For example, in the case of evolution, if one argued that evolution is the source of some of our moral cognitions (e.g., case-intuitions) and that this makes them unreliable, one could extend this argument to any of our other cognitions, as they were also shaped by evolution, leaving us with distrust for the same cognitions we used to ground our argument.[134] To avoid this problem, the foundationalist has argue that not all sources are equally unreliable, thus stopping the regress that undermines all of our beliefs. More specifically, the foundationalist has to show both that some sources tend to produce less reliable beliefs, whereas others do not, and that some judgments are connected to these reliable sources, whereas others are not.[135]

Besides self-defeatingness, the second problem for arguments that appeal to the source of a judgment is the so-called genetic fallacy. This fallacy involves confusing the causal origins of a judgment with its justification.[136] One way in which this fallacy is often described is as confusing the contexts of discovery with the contexts of justification. For example, the way in which a physicist finds a theorem constitutes the context of discovery,

---

[132] See Singer; Singer, "Ethics and Intuitions"; Lazari-Radek and Singer, *The Point of View of the Universe*.
[133] Henry Sidgwick, *The Methods of Ethics*, 7th ed. (Palgrave Macmillan, 1901), pp. 212-213.
[134] "The Theory of Evolution in Its Application to Practice," *Mind* 1, no. 1 (1876): 52–67, p. 54.
[135] Sidgwick, *The Methods of Ethics*, p. 213.
[136] See Ted Honderich, *The Oxford Companion to Philosophy* (Oxford: Oxford University Press, 2005), p. 331.

but how the theorem is presented to the public—connecting facts and premises in a logical way—constitutes the context of justification.[137] Both Rawls and Sidgwick, who are in opposite methodological camps, agree with this idea expressed by the genetic fallacy, as well as many other philosophers and almost every single logic textbook. Rawls considers irrelevant the origin of an intuition for the purposes of reflective equilibrium:

> […] it is immaterial whether the judgments in its range are caused by the intuition of nonnatural ethical characteristics, or by the response of intentional feelings to directly experienced value qualities, or by emotional attitudes which may in turn have been caused by certain specifiable psychological and sociological determinants. Questions about the actual causes, while interesting, are irrelevant from the standpoint of the present method.[138]

Similarly, Sidgwick questions why determining the origin of a belief is a ground for undermining its justification:

> I cannot see how the mere ascertainment that certain apparently self-evident judgments have been caused in known and determinate ways, can be in itself a valid ground for distrusting this class of apparent cognitions.[139]

In general, the underlying problem of genetic arguments is taken to be that they are not deductively valid: even if the premises of the argument are true, this does not mean that the conclusion is false. As applied to moral judgments, we could say that even if it is true that our judgment is the result of emotions, culture, or evolution, this does not mean that the judgments is false, as there could be additional evidence—known or unknown to us—in support of it. For example, if someone believed in climate change merely because scientists said so, this does not mean that climate change is not real, as there might be additional evidence, unknown to him but known to the scientists, that climate change is real. Likewise, if someone believed that it is wrong to eat meat produced by factory farming methods, it seems irrelevant to the justification of this belief whether it is mostly explained by the person having watched YouTube videos of factory farming methods, as there can be

---

[137] Andrew C. Ward, "The Value of Genetic Fallacies," *Informal Logic* 30, no. 1 (2010): 1–33, p. 1-2.
[138] Rawls, "Outline of a Decision Procedure for Ethics", p. 185.
[139] This point is made by Sidgwick, *The Methods of Ethics*, p. 212.

additional reasons in support of the belief. Thus, the genetic fallacy seems to shift the burden of proof to those that seek to establish a relation between the causal origin of a judgment and its justification: an objector has to show why pointing out the origin of case-intuitions would undermine their reliability.

To be clear, there are authors who have thought and think that the process that led to forming a belief *is relevant* to the justification of the belief. Sigmund Freud, for example, attempted to undermine our belief in religion by showing its psychological origins. Jane Flax argued that our credence in some beliefs is unjustified if they originated from men embedded in a patriarchal society that institutionalizes disempowerment and alienation.[140] Peter Singer, as we will see, thinks that the facts that some of our judgments originate from emotions, cultural influences, or evolutionary adaptations constitutes a reason to doubt them. However, Andrew C. Ward mentions that contemporary epistemologists who are sympathetic to the idea that forming a belief matters to its justification defend a weaker form of it. It is not that the causal history of a belief *always* matters for its justification, says Ward; rather, the idea is that the causal history is only *sometimes* relevant to its epistemic status.[141] Conceding that a genetic argument is deductively invalid, Ward argues that genetic arguments can be, nonetheless, inductively valid, as they might give us reason to think that the conclusion is less probable.

Ward mentions two ways in which genetic arguments can be appropriate. Firstly, they can give us evidence to think that the premises one *would* use in supporting a claim are false. For example, if a person is naturally disposed to distrust republicans, we have reason to be suspicious of the claims that person would make, say, against environmental policies of republicans. Another example is when there is conflict of interest, such as when a pharmaceutical company publishes a study of its own product. In both cases, it is not that the person's and the company's claims are false, says Ward, but only that we have reason to be careful when assessing them. Secondly, a genetic argument reminds us that "our prejudices—whatever their cause or character—sometimes lead us to accept as true what isn't, and to unquestioningly accept ideas and practices that should give us pause."[142] In

[140] Ward, "The Value of Genetic Fallacies", p. 3.
[141] Ward, p. 4-5.
[142] Ward, p. 27.

this way, for example, physicians might be recommending a drug merely because it had a lot of marketing behind it. A genealogical critique, then, might uncover this fact and allow us to assess the evidence in favor of it—to determine whether the drug should continue being recommended or not.

As plausible as Ward's account might be, for the purposes of the foundationalist, however, it does not seem to give a compelling reason to systematically reject case-intuitions, even if we could make genetic arguments against them. Ward's conclusions suggest, at most, prudential reasons to distrust case-intuitions: reasons to think that it might be hard to think properly about particular cases, but not an epistemic reason to think that we should not consider them at all. Ward's ideas imply that we should not take for granted those ideas that are subject to genealogical critiques and find further reasons in support of them; but as long as we are careful when considering the important factors in a case, we would have no reason to distrust case-intuitions. I take it, then, that the deductive motivation underlying the genetic fallacy stands: it is invalid to reject a judgment on the basis of its origin, as there could be additional reasons that might justify it.

Let us then examine the challenges raised against case-intuitions and see whether they can successfully address the two previous problems posed by the genetic fallacy and by the self-defeatingness objection.

3.1 Emotional Sources

As we mentioned before, one argument against the systematic use of case-intuitions is that these intuitions tend to be based on emotions. In this way, one could claim that case-intuitions are unreliable because they are driven by emotions or emotional processes. To properly examine this argument, it is helpful to distinguish between the three separate claims that are being made. The first claim is an empirical one, and it is the claim that moral judgments, in general, can be influenced by emotions or emotional processes. The second claim is a normative claim, and it is the idea that this influence is problematic, and so judgments can be considered unreliable if they are made under the influence of strong emotions or emotional processes. Finally, the last claim is also an empirical claim, and it is the idea that case-intuitions are especially prone to emotional influences. Although all

these three claims need to be justified for the argument to work, I will limit my discussion to the second premise. This is because I take it that the empirical claims (the first and third ones) need to be assessed by looking at current studies and revised in light of new ones, so I do not believe there is much to say here: our discussion might become outdated as further information is gathered. Thus, the claim that interests me the most is the second premise, and, for the sake of argument, I will not dispute the other premises.

Let me briefly mention that there is indeed substantial empirical evidence which shows the influence that emotions can have on moral judgments. Schnall et. al., for example, conducted a study in which participants made more severe moral judgments while experiencing feelings of disgust than the control group which was not presented with the disgusting stimuli.[143] Also, Kahneman et. al. reported that people's desire to see a corporation punished for its behavior is determined partly by the amount of emotional outrage they experience as result of the corporation's behavior.[144] Moreover, the relation between moral judgments and emotions has also been established in an indirect way. Rather than study whether emotions directly cloud our moral judgments, it has been shown that certain moral judgments are correlated with activation in areas of the brain associated with emotion. Moll et. al. found, for instance, that the activation of areas of the brain varied depending on whether participants were presented with sentences with explicit moral content, with factual statements devoid of moral connotation, or disgusting non-moral sentences.[145] Also, Greene et. al. found that "personal" moral dilemmas engaged emotional processing to a greater extent than more "impersonal" moral dilemmas.[146] As there seems to be substantial empirical evidence of the connection between emotions and

---

[143] Simone Schnall et al., "Disgust as Embodied Moral Judgment," *Personality and Social Psychology Bulletin* 34, no. 8 (2008): 1096–1109.

[144] Daniel Kahneman, David Schkade, and Cass R. Sunstein, "Shared Outrage and Erratic Awards: The Psychology of Punitive Damages," *Journal of Risk and Uncertainty* 16 (1998): 49–86.

[145] Jorge Moll, Paul J. Eslinger, and Ricardo de Oliveira-Souza, "Frontopolar and Anterior Temporal Cortex Activation in a Moral Judgment Task: Preliminary Functional MRI Results in Normal Subjects," *Arq. Neuropsiquiatr* 59, no. 3 (2001): 657–64; Jorge Moll et al., "Functional Networks in Emotional Moral and Nonmoral Social Judgments," *NeuroImage* 16, no. 3 (2002): 696–703.

[146] "Personal" moral dilemmas included pushing a large man off a bridge to save five people, a case of stealing one person's organs in order to distribute them to five, and a case of throwing people off a sinking lifeboat. "Impersonal" moral dilemmas included diverting a train with would otherwise kill five people but at cost of killing one person, a case of keeping money found in a lost wallet, and a case of voting for a policy expected to cause more deaths than its alternative. Joshua Greene et al., "An fMRI Investigation of Emotional Engagement in Moral Judgment," *Science* 293, no. 5537 (2001): 2105–8.

moral judgments, the claim that interests me is whether it is problematic for a judgment to be driven by such emotions.

There are various ways in which the idea that emotions are problematic for moral judgment has been defended. One way is by appeal to its direct plausibility. Knobe and Nichols, for instance, take for granted that emotionally-charged mental states can render certain intuitions unreliable:

> Clearly, an intuition developed in a jealous rage is less trustworthy than one developed after calm and careful consideration. Thus, if our hypothetical philosopher discovers that her intuition about a case is driven by such distorting emotional reactions, this will and should affect how much she trusts the intuition.[147]

However, we cannot be content by merely justifying this premise by appeal to its direct plausibility. One reason to think otherwise is, as I mentioned before, that there is also great plausibility in the genetic fallacy, which directly conflicts with the second premise. Recall, the genetic fallacy is the idea that the context of discovery should not matter for the context of justification. To take Knobe and Nichols's scenario, suppose a boyfriend is in a jealous rage and develops the idea that couples should not cheat each other. In this case, it seems that his idea cannot be dismissed merely because it was produced in a jealous rage, as there might be additional reasons to think that it is justified. Accordingly, it seems that moral judgments should be assessed by the reasons we can give for or against them, and not by whether they are correlated with an emotional response or not.

A second way in which it has been defended that it is problematic for a judgment to be driven by emotions is by way of analogy. Huemer, for example, mentions that there is empirical evidence to think that emotions impair our factual judgment. Then, by way of analogy, Huemer mentions that there might be *prima facie* reason to assume that emotions impair our moral judgment.[148] However, this analogy relies on the assumption that emotions do cloud or impair our factual judgments, and this is not clear from an empirical standpoint. As noted by Mason—drawing the work of Antonio Damasio—subjects with lesions in emotional parts of their brain but that retain normal intelligence and semantic

---

[147] Knobe and Nichols, *Experimental Philosophy*, p. 8.
[148] Huemer, "Revisionary Intuitionism", p. 377-378.

knowledge have trouble assessing the risk of practical tasks, which suggests that emotions carry important information about the environment and are thus essential for good judgment.[149] Moreover, the second problem with Huemer's suggestion is that there is a fundamental disanalogy between factual and moral judgments. In the factual case, we have an independent way—e.g., sensory perception—to test whether a judgment is true or not; thus, we can indeed establish whether there is a correlation between experiencing a strong emotion and making true factual judgments. But we do not have a similar way to validate moral judgments in ethics. To say that strong emotions tend to bias or cloud our judgment, one needs to know beforehand which judgments are biased and which are not. Yet, how are we to distinguish between biased and unbiased judgments in the first place? If we say that unreliable judgments are simply those made under the influence of strong emotions, our argument would be blatantly circular. And if we propose an alternative method to determine the reliability of our judgments, then appealing to emotions would be unnecessary, for we could simply use our alternative method to directly test whether our judgments are reliable or not.

It may be objected that we can stop the regress before, without knowing beforehand which judgments are biased and establishing a correlation. We can attempt to do this by adopting a reliabilist epistemology. Roughly speaking, reliabilism is the view that emphasizes truth-promoting factors over the truth of the target proposition. Accounts of what these truth-promoting factors are vary widely, but the most discussed one is process reliabilism.[150] As articulated by Goldman, process reliabilism is the view that the process that lead to the formation of a belief matters for its justification.[151] What makes a process confer justifiedness or undermine it? According to Goldman, it is its tendency to produce beliefs that are true rather than false. Examples of reliable processes include perceptual processes, remembering, good reasoning, and introspection. Contrarily, examples of

---

[149] Kelby Mason, "Moral Psychology and Moral Intuition: A Pox On All Your Houses," *Australasian Journal of Philosophy* 89, no. 3 (2011): 441–58, p. 446.
[150] Alvin Goldman and Bob Beddor, "Reliabilist Epistemology," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Winter 2016 (Metaphysics Research Lab, Stanford University, 2016), https://plato.stanford.edu/archives/win2016/entries/reliabilism/.
[151] Alvin Goldman, "What Is Justified Belief?," in *Reliabilism and Contemporary Epistemology* (New York: Oxford University Press, 2012), 29–49.

unreliable processes include confused reasoning, wishful thinking, mere hunches or guesswork, hasty generalization and emotional attachment.[152]

I will not discuss the general merits or defects of Goldman's account here, so let us assume for the sake of argument that his account is roughly correct. What we are concerned with, however, is with whether his account applies *specifically* to the case of moral judgments driven by emotions. And it seems to me that it does not. In mentioning emotional attachment as a justification-undermining process, Goldman is probably thinking about factual beliefs. But, as I said, there is a strong disanalogy with moral judgments here. With factual issues, we can test our judgments by other means than by an appeal to the process itself: we can do this, for example, with sensory perception or by looking at the predictive power of our scientific theories. This is why we could determine whether a process has a tendency, a frequency to lead to true beliefs without falling into circularity. Yet, in the case of moral judgments, we do not have these resources. Thus, the only way to know whether a process has a tendency to produce true moral judgements would be to know beforehand which moral judgments are true or false, but if we do, as I said before, then there would be no need to think about emotionally-driven judgments at all.

Perhaps one could try to strengthen the analogy between factual and moral judgments by saying that moral deliberation relies partly on the accurate consideration of the facts of a case. Thus, one could say that strong emotions are problematic because they tend to lead us to ignore important factors. For example, people from the left or the right will tend to seek information that confirms their own views and ignore contradicting information, as research in moral psychology has shown.[153] Also, in a criminal trial, friends or family members are presumably not allowed to serve because they will tend to ignore important factors that play against their interests. However, as I mentioned before, it is not clear from an empirical standpoint that emotions impair factual judgment, as emotions sometimes carry important information about the environment and are thus essential for good judgment. But even if they do not, the argument does not show that it is impossible to

[152] Goldman, p. 37.
[153] Haidt, *The Righteous Mind: Why Good People Are Divided by Politics and Religion*.

consider the many factors of a case, only that it is hard to do so in practice *and* for some people. The problem would lie, then, in not considering various factors, rather than on the emotion itself. Claiming otherwise would be like saying that we should not engage in philosophical reasoning because people tend to commit fallacies. So, as long as we are careful when considering the critical factors in a case, we would have no reason to distrust case-intuitions.

Another way in which one might try to reply would be by pressing the analogy between predictive power in science and ethics, which might allow us to tell which judgments are correct or wrong without appeal to our method itself. One could do this by saying, for example, that predictive power in ethics refers to the tendency to produce judgments which have been historically adopted. Rawls hints at this is possibility in one passage:

> In general, a principle evidences its reasonableness by being able to resolve moral perplexities which existed at the time of its formulation and which will exist in the future. This test is somewhat analogous to a test which we impose upon an empirical theory: namely, its ability to foresee laws and facts hitherto unknown, and to explain facts and laws hitherto unexplainable.[154]

Some utilitarian-leaning authors have made an argument of this sort. It is said, for example, that utilitarian thinkers like Bentham or Mill were way ahead of their time, as they were advocates of laws protecting animals from cruelty, gay rights, better conditions for prisoners, environmental protection, separation of church and state, a better system relief for the poor, and women's rights in voting, owning property in marriage, and going to university.[155] Contrarily, authors like Kant defended views which we now consider laughable or wrong, such as that masturbation is wrong and that whites are superior to blacks.[156] Also, Greene makes a persuasive case that "deontological" type judgments, the ones that Kant relied on, are driven by emotion to a larger extent than "consequentialist"

---

[154] Rawls, "Outline of a Decision Procedure for Ethics", p. 188.
[155] Greene, *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them.*, p. 55; Lazari-Radek and Singer, *Utilitarianism: A Very Short Introduction*.
[156] Greene, *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them.*, p. 301.

type judgments, the ones that Bentham and Mill presumably relied on.[157] Consequently, one might conclude that emotionally-driven judgments have a greater tendency to lead to false beliefs, whereas non-emotionally-driven judgments tend to produce truer beliefs.[158]

Nonetheless, I believe this response would fail for the simple reason that we cannot assume that judgments which have been historically adopted are true, whereas those that have not been adopted are false. Assuming otherwise involves a problematic (almost Hegelian) account of inevitable progress. Also, even if there is such progress, it does not mean that it is linear—as we cannot assume that every single change that occurs is for the better—or that it is uniform—as there are substantial differences in the moral views between people around the globe.

So far, I have argued that the problematic relation between emotions and justification cannot be defended merely by its initial plausibility or by way of analogy with factual judgments. Let us examine now one last possible defense that comes from the work of Joshua Greene. In one of his papers, Greene et. al. says that empirical results are "descriptive rather than prescriptive" and that they do not show "any actions or judgments to be morally right or wrong."[159] In a later paper, however, Greene defends that empirical results can indeed have normative implications. Although Greene's argument is aimed at attacking deontological moral theories, here I am interested in his underlying reasons for dismissing certain moral judgments—reasons which could apply to case-intuitions as well—rather than on the specific claim that deontological judgments are unreliable.

Greene's argument seems to go as follows. Emotions and emotional processes are the best explanation for why we make certain moral judgments, namely, deontological ones. This is clear because, as shown in the empirical studies, there is a strong correlation between experiencing strong emotions and making deontological judgments—in other words, were not for the emotions experienced, participants would probably not make

---

[157] Greene, "The Secret Joke of Kant's Soul."

[158] To be fair to Greene, he does not claim that there are emotionless moral judgments. He does mention, however, that the emotions that the emotions that influence deontological or consequentialist judgments are different: "I am sympathetic to Hume's claim that all moral judgment (including consequentialist judgment) must have some emotional component. But I suspect that the kind of emotion that is essential to consequentialism is fundamentally different from the kind that is essential to deontology, the former functioning more like a currency and the latter functioning more like an alarm." Greene, p. 41.

[159] Greene et al., p. 2107.

certain deontological judgments.[160] Additionally, these emotions we experience when making deontological judgments are best explained by evolution. For instance, in the case of retribution, it seems that "the emotions that drive us to punish are blunt biological instruments […] Thus, it seems that as an evolutionary matter of fact, we have a taste for retribution […] because retributive dispositions are an efficient way of inducing behavior that allows individuals living in social groups to more effectively spread their genes."[161] Finally, Greene claims that although these considerations do not automatically render deontological judgments as wrong or misguided, they do pose one major problem for achieving moral truth. If our emotion-driven intuitions were "shaped by morally irrelevant factors having to do with the constraints and circumstances of our evolutionary history,"[162] then it would be a great coincidence if they "correspond to some independent, rationally discoverable moral truth."[163]

I should say that there is also a slightly different way to interpret Greene's basis for dismissing emotionally-driven judgments. At one point in his discussion, Greene mentions that "our distinctively deontological moral intuitions […] reflect the influence of morally irrelevant factors and are therefore unlikely to track the moral truth."[164] Here, we also have a connection to moral truth, but the appeal to evolution is absent (although Greene's discussion happens in the context of evolution and biology, and by a "morally irrelevant factor" he might be referring to evolution). But I believe one could make a slightly different claim. One could say that experiencing a certain emotion is morally irrelevant, and so it should not make a difference in our response to a case if everything else is held equal. For example, in Schnall's study, participants gave different responses to various scenarios, going from less severe to more severe moral judgments, when everything else was held equal except for the feelings of disgust that some participants experienced.[165] However, if experiencing a feeling of disgust is morally irrelevant, then it seems that we should distrust the judgment that results from experiencing this feeling.

---

[160] Greene, "The Secret Joke of Kant's Soul", p. 67-68.
[161] Greene, p. 71.
[162] Greene, p. 75.
[163] Greene, p. 72.
[164] Greene, p. 70.
[165] Schnall et al., "Disgust as Embodied Moral Judgment."

If these reconstructions of Greene's argument are charitable—the second one is not, I should clarify, as it is simply one way in which *I think* Greene could present his argument—we can see that emotions are only *derivatively* problematic. The first argument makes it clear that the fundamental reason as to why emotions are problematic is because they were shaped by evolution. Then, we are talking about an evolutionary debunking argument. The second argument makes it clear that the reason as to why emotions are problematic is because of they make us vary our responses to cases which presumably only vary with regard to morally irrelevant features. Then, we are talking the *instability* of our judgments being problematic. These two arguments require more extensive examination, so I refer the reader to section 3.3 and Chapter 4, respectively, which will address the evolutionary-debunking argument and the stability argument. This is all I should say in response to Greene and in closure of this section, and I am satisfied if I have presented a convincing case as to why an appeal to emotions cannot be justified by its direct plausibility or by way of analogy and as to why the argument is only derivatively problematic.

3.2 Cultural sources

Another argument that is sometimes made against case-intuitions is that they originate from or are influenced by obsolete religious systems, outdated customs, habits, our parents, or our society. Brandt, for example, doubts intuitions on the ground that "our normative beliefs are strongly affected by the particular cultural tradition which nurtured us, and would be different if we had been in a learning situation with different parents, teachers or peers."[166] Singer also suggests this idea in an earlier paper, when criticizing reflective equilibrium:

> Why should we not rather make the opposite assumption, that all the particular moral judgments we intuitively make are likely to be derived from discarded religious systems, from warped views of sex and bodily functions, or from customs necessary for the survival of the group in social and economic circumstances that now lie in the distant past? In which case, it would be best to forget all about our

[166] Brandt, *A Theory of the Good and the Right*, p. 21.

> particular moral judgments, and start again from as near as we can get to self-evident moral axioms.[167]

In this passage, Singer is denying that we should start moral inquiry by taking into account our moral judgments in general, as Rawls wants us to do. He is rejecting the idea that it is obvious to start by doing this, as there is a plausible reason to distrust particular judgments (case-intuitions), namely, that they are influenced by outdated customs or values. In other words, we might interpret this passage as saying that particular judgments are influenced by outdated customs and values (an empirical claim) and that this influence is problematic (a normative claim).

Why would the influence of outdated customs and values be problematic? One option to say that is a directly plausible idea and that it requires no further justification (a similar idea to the one we discussed in the previous sections). However, cultural relativists will argue the opposite: that being influenced by culture is not a problem, but a strength, as moral judgments are correct precisely because they are held by a culture. Moreover, I believe this argument is unsuccessful for additional reasons, even if we assume for the sake of argument that our moral views are indeed influenced by cultural or social influences. The problem is that by way of direct plausibility we cannot satisfactorily address the two problems I mentioned at the beginning of the chapter: the genetic fallacy and the self-defeatingness objection.

Firstly, thinking that we should dismiss culturally or socially influenced judgments is indirectly self-defeating. Arguably, every judgment is influenced to some extent by culture and society. Therefore, the very same idea that we should distrust culturally-influenced judgments is culturally-influenced to some extent. Even if one denies this or some other judgments are not culturally or socially influenced, at the might at some point if they are plausible enough; when that happens, then we would have reason to distrust it. The same applies if we distrust a judgment for it having become habitual. Singer's views are probably habitual to him by now, so if his argument is correct, then he would be justified in dismissing his own view, which would be self-defeating.

---

[167] Singer, "Sidgwick and Reflective Equilibrium", p. 516.

Secondly, it seems that there should be no connection between a judgment being culturally or socially influenced and its justification. Singer argues, for instance, that our views on abortion, suicide, and voluntary euthanasia are influenced by Christian tradition; but, arguably, many other moral views we hold are as well. The Golden Rule, for example ("do to others what you want them to do to you") expresses a concern for other people that the majority of us would probably agree with, and it is mentioned by Jesus in the Bible. However, it seems that we should not dismiss the Golden Rule merely on the basis that it is part of the Christian tradition. The reasonable thing to do would be to assess it by its own merits, just as we should do with any other view in Christianity. We could then then decide whether we want to keep it or not. The same could be said of Singer's views. His views are greatly incluenced by the utilitarian tradition of moral philosophy, but it seems that we should not dismiss Singer's claims merely because they were influenced by authors like Bentham, Mill, Sidgwick, or R. M. Hare, but assess them on their own merits.

Why, then, can genetic arguments be so rhetorically effective and often convincing on their own right? I believe the more plausible story is that a genetic argument gives us evidence to think that we are prejudiced and, more importantly, that the premises one *would* use in supporting a claim are false. An example illustrates this point. In his discussion of the permissibility of abortion, Singer is charged with the claim that his view implies that infanticide is permissible in many cases. Singer responds by saying that many societies have considered infanticide acceptable and even obligatory in some cases.[168] What is Singer's point? If it were true that we should decrease confidence in culturally inherited views, Singer's response would undermine itself, as he would be defending his view by saying that it is culturally inherited. But the more charitable and likely interpretation is that Singer is trying to make us see that we should not take for granted our intuition that infanticide is wrong: that the judgment *infanticide is wrong* might be a candidate for which we need additional justification which we might not be able to find. Yet, we cannot dismiss at the outset the possibility this idea—we should examine it on its own merits.

---

[168] Peter Singer, "A Response," in *Singer and His Critics* (Oxford, UK; Malden, Mass.: Blackwell Publishers, 1999), 269–335.

We have seen that direct plausibility might not be sufficient to ground a genetic argument based on cultural influences. Another way to defend it is suggested by Singer in a passage:

> Particular views have been inculcated into us by parents, teachers and society from childhood. Many of them we act upon every day—telling the truth, not stealing when we have the opportunity to do so, and so on. These judgments sink deep, and become habitual. […] If it is then pointed out to us that this fundamental moral principle is incompatible with some of the particular moral judgments we are accustomed to making, and that therefore we must either reject the fundamental principle, or else abandon our particular judgments, surely the odds are stacked against the fundamental principle. Most of us are familiar with fingering guilt feelings that occur when we do something that we are quite certain is right, but which we once thought to be wrong. These feelings make us reluctant to abandon particular moral views we hold, but they in no way justify these views.[169]

Singer's idea seems to be that cultural or social influence is problematic because it will bias us in favor of whatever moral views we inherited or learned about. Thus, even if a plausible principle were presented to us, we would likely not accept it because of the force of habit and history, or the cognitive dissonance we might experience.[170] Singer emphasizes again this idea about the force of history and habit in a later article:

> On abortion, suicide, and voluntary euthanasia, for instance, we may think as we do because we have grown up in a society that was, for nearly 2000 years, dominated by the Christian religion. We may no longer believe in Christianity as a moral authority, but we may find it difficult to rid ourselves of moral intuitions shaped by our parents and our teachers, who were either themselves believers, or were shaped by others who were.[171]

---

[169] Singer, "Sidgwick and Reflective Equilibrium", p. 516-517.
[170] In psychology, cognitive dissonance refers to the mental discomfort experienced by a person who simultaneously holds two or more contradictory beliefs, ideas, or values. Brant also emphasises it as an obstacle to reflective equilibrium: "People may just dislike an incoherent set of moral attitudes, just as they dislike harbouring incoherent beliefs—they may dislike this form of cognitive dissonance."Brandt, *A Theory of the Good and the Right*, p. 21.
[171] Singer, "Ethics and Intuitions", p. 345.

In sum, Singer is claiming that cultural and social influences (e.g., those of Christianity) predispose us to accept certain moral views, namely, those particular judgments that happen to be common or habitual to us.

I believe this line of argument is also unsuccessful because of a consideration I have mentioned before: it is unclear whether Singer's argument poses an epistemological problem for case-intuitions or, rather, a practical obstacle. Singer's argument does not show that it is impossible to consider alternative views, only that it is hard to do so. So, even if it is sound, it would give us a reason to find creative and ingenious solutions to avoid our practical biases, but not an epistemic reason to dismiss case-intuitions. Singer's argument is the equivalent of saying that the scientific method is inadequate because scientists, in practice, have an inclination to support their own theories. But why would this matter for what epistemic status should the scientific method have? Thus, the practical obstacles we might face for constructing a moral theory seem irrelevant for the justification of the theory.

Admittedly, if we only considered the cases and principles that occurred to us or that we wanted to consider, then revision would surely be difficult. However, in reflective equilibrium we seek the systematization of our intuitions about as many judgments as we can—both non-moral judgments and moral judgments—not only the ones that we currently have or that uncritically occurred to us. In this sense, I follow Rawls when he says that we have to consider all possible descriptions (or, in our terminology, all possible intuitions about cases and principles).[172] By doing this, the racist, the sexist, and the elitist are then forced to consider other ethical theories or plausible alternatives to their views which they have not consider yet. Therefore, even if Singer is right and it is true that we will be initially inclined to accept what is habitual to us, this only gives us reason to be careful in trying to consider as many alternative views as we can.

[172] Rawls, *A Theory of Justice*, p. 43.

3.3 Evolutionary sources

Let us now address one last genetic argument that is often made against case-intuitions: that they are originated in or are influenced by unreliable evolutionary processes (hereafter, I will sometimes refer to this idea as the evolutionary debunking argument). This section ties in together with what we concluded in section 3.1 regarding Joshua Greene's argument against emotions. In it, I argued that emotions are only derivatively problematic, as Greene suggests that the problem with emotions is that they are explained by evolution, a process which would be highly unlikely to track moral truth. Thus, this section responds both to Greene and to other authors who defend evolutionary debunking as a way to dismiss case-intuitions.

My discussion in this section will center around the account of evolutionary debunking that has perhaps received the most attention, the one of Lazari-Radek and Peter Singer. Their argument goes as follows. Lazari-Radek and Singer begin by drawing from an argument by Sharon Street, although they depart from it later on. Street claims that evaluative attitudes—the ones that are present in moral judgments—are largely shaped by evolutionary forces. If this is correct, Street says, we then face a dilemma: either we think that these evolutionary forces led to the selection of beings who do not hold objectively true evaluative attitudes or of beings that do hold these true evaluative attitudes. If our answer is negative—that is, if evolution did not lead us to hold moral truths—then most of our evaluative attitudes are unjustified. It could certainly happen by mere chance that our evaluative attitudes actually correspond to the objectively true ones, but, Street says—in similar fashion to Greene—this would be highly unlikely: it would be the equivalent of sailing to Bermuda, letting our boat's course be determined by the winds and tides, and reaching Bermuda.[173] Now, if our answer to the dilemma is positive—that is, if evolution did lead us to hold moral truths—then we would have to say that the presence of our evaluative attitudes is best explained by the fact that they are true, but this is contrary to our theories of evolution. According to scientific understanding, the presence of our

[173] Sharon Street, "A Darwinian Dilemma for Realist Theories of Value," *Philosophical Studies* 127 (2006): 109–66; referenced in Lazari-Radek and Singer, *The Point of View of the Universe*, p. 179-185.

evaluative attitudes is best explained, instead, by their conduciveness to survival, to reproductive success, and to offspring survival.[174]

Let alone, this argument would be problematic for any ethical methodology that relies on moral realism, not just for reflective equilibrium. Assuming that it is unlikely that evolutionarily-shaped moral intuitions track moral truth and that this is problematic, Street's argument would undermine equally all moral intuitions, general or particular. This sort of skepticism, nonetheless, does not interest me for the purposes of this thesis; my only interest is in whether foundationalists (and, to a lesser extent, anti-theorists) could give any reason to prefer their methodology over reflective equilibrium. Consequently, I take it that a foundationalist challenge against case-intuitions is unsuccessful if it undermines all moral intuitions. If a foundationalist wants to use Street's evolutionary debunking, then he or she has the burden of proof in showing that not all intuitions are subject to debunking, namely, that principle-intuitions are immune to this type of criticism, whereas case-intuitions are not.

This is precisely what Lazari-Radek and Singer attempt to do. These authors dodge part of Street's dilemma by arguing that there is a difference between particular judgments and more general ones. According to Lazari-Radek and Singer, whereas particular moral judgments are indeed best explained by evolutionary forces—and so they are unjustifiably held—some more general moral judgments are not. One example is Sidgwick's axiom of universal benevolence: *each one is morally bound to regard the good of any other individual as much as his own, except in so far as he judges it to be less, when impartially viewed, or less certainly knowable or attainable by him*. According to Lazari-Radek and Singer, this judgment goes against evolutionary forces because it does not favor individual or kin survival. Although some degree of altruism towards other individuals can be explained by evolution, this is not true for universal altruism towards all sentient beings, and Lazari-Radek and Singer cite various evolutionary biologists who agree with this idea.[175]

---

[174] This is evidenced by the variety of evaluative attitudes of different species and how they fit with their reproductive goals: social insects have a stronger orientation towards the welfare of the community, male lions kill offspring that are not their own, and queen bees strive to kill their fertile daughters. See Street, "A Darwinian Dilemma for Realist Theories of Value", p. 125-135.

[175] Lazari-Radek and Singer, *The Point of View of the Universe*, p. 182-187.

As noted by Nye, it would not be enough to show that a moral judgment is not influenced by evolution to justify it. This would only lead us to think that the judgment is a *spandrel* of evolution, that is, the result of an evolutionarily-shaped process that did not contribute directly to the goals of evolution, but this does not mean that the spandrel is correct or truth-tracking. In fact, Nye correctly points out that the problematic aspect about a debunking explanation lies not in a judgment being influenced by evolution, but in it not tracking truth. This becomes clear when considering sensory perception, which is primarily influenced by evolution but we still consider it reliable because there is a story of how it tracks truth. [176] So, in a similar vein to Street's argument, we could say that without any additional justification, it would be highly coincidental if an unintended consequence of an evolutionary processes led us closer to truth.

To address this problem, Lazari-Radek and Singer make a further claim: they argue that moral judgments like Sidgwick's axiom are indeed spandrels, but what makes them special is that they are spandrels of our capacity for reason, which, although evolutionarily-influenced, is truth-tracking.[177] The authors mention that it is not only plausible to think that reason does track truth—as it seems to have been necessary to survive in the world—but that reason's spandrels do as well. Evidence for this are abstract mathematical truths and some normative epistemic beliefs, which, presumably, are spandrels of reason—as they are not strictly necessary for survival—but also seem to be true.

Whatever its merits, Lazari-Radek and Singer's argument has various problems that it needs to address for it to work, and it seems to me that it is not able to. Let us examine some of them. One initial problem with the debunking argument is that it presupposes moral realism—as Joshua Greene's argument also does—which is hardly an uncontroversial philosophical position. Moral realism can be understood as the view that there are moral properties, that some things have those properties, and that these properties do not depend on the psychological attitude or response of observers.[178] In the argument, the main reason why we would trust, say, Sidgwick's axiom of benevolence is

---

[176] Nye, "Directly Plausible Principles", p. 622-623.
[177] Huemer takes a similar stance when defending the reliability of ethical intuitions: "On my account, the correct explanation for why we have ethical intuition will refer to whatever is the explanation for why we have reason and intelligence in general." Huemer, *Ethical Intuitionism*, p. 218.
[178] Huemer, p. 4-6.

because it presumably corresponds to a moral property in the world; otherwise, we would have no reason to prefer this moral judgment over others. But if moralism realism is needed as a premise, then the argument will only convince those who already endorse this view. Rawls would certainly not fall within this group, as he was willing to put these metaethical debates and even suggested that it would be wise to suspend consideration about this issue, given how problematic it has been.[179] If pressed for an alternative position, we could adopt, for instance, a constructivist account of moral truth, according to which moral truth is just whatever comes out of a process of thorough reflection. We could not then claim that evolutionarily-influenced judgments lead us farther from the truth, for the truth would simply be what results of subjecting those same judgments to thorough reflection. Therefore, in addition to their evolutionary debunking argument, Lazari-Radek and Singer would need to give us an argument in favor of moral realism, which is certainly not an easy position to defend.

Secondly, it is not clear what "reason" involves and whether it is sufficient to ground Sidgwick's axiom of benevolence.[180] Although Lazari-Radek and Singer dedicate a chapter to discuss whether reason by itself can ground moral judgments and motivate action,[181] they never provide a detailed account of "reason." This term, however, is used in various ways, and with regard to such a wide range of topics, that appealing to it hardly proves any point.

In fact, there is an alternative and more plausible explanation for Sidgwick's axiom. This explanation comes from a theory in moral psychology that has received much attention: the Moral Foundations Theory of the psychologist Jonathan Haidt.[182] According to Haidt's theory—which is based on "its scientific usefulness for both answering existing

---

[179] According to Rawls, "Since the history of moral philosophy shows that the notion of moral truth is problematical, we can suspend consideration of it until we have a deeper understanding of moral conceptions." See Rawls, "The Independence of Moral Theory", p. 7.

[180] Greene, actually, is sympathetic to the view that all moral judgments have some emotional component and doubts whether there is even a clear distinction between reason (or what he calls "cognition") and emotion. See Greene, "The Secret Joke of Kant's Soul", p. 41.

[181] Lazari-Radek and Singer, *The Point of View of the Universe*, p. 33-65.

[182] I owe this point to Peter Andes, a fellow graduate student at the University of Alberta, who discusses how Sidgwick's axioms might be debunked by using Haidt's theory: "Sidgwick's Dualism of Practical Reason, Evolutionary Debunking, and Moral Psychology," (unpublished manuscript).

questions about morality and allowing researchers to formulate new questions"[183]—we come equipped with five "moral receptors" that form the basis of all moral judgments. These receptors work in an analogous way to sensory perception: in the tongue, for example, we have receptors for salt, sweet, bitter, sour, and, glutamate, and all these work together with other senses to give us the great variety of gustatory experiences. Likewise, moral receptors interact with cultural learning to produce the variety of our moral judgments.[184] In particular, two of the moral receptors are relevant in the context of our discussion. Firstly, the *care/harm foundation*, which triggers the emotions of compassion toward the suffering of a victim or of anger towards a perpetrator, and it evolved in response to the adaptive challenge of caring for vulnerable offspring, detecting signs of suffering, distress, or neediness. Secondly, the *fairness/cheating foundation*, which triggers the emotions of anger or guilt towards cheating or deception (depending of whether it is caused by oneself or by others) and gratitude towards reciprocity, and it evolved in response to the adaptive challenge of living alongside other human beings.[185] By looking at these two receptors, it is easy to see how they might be motivating Sidgwick's axiom of benevolence. The *care/harm foundation* is likely providing a motivation for caring for the good of others in the first place. It is the basis for caring about suffering at all, and for saying that we are "morally bound to regard the good of any other individual." This would only get us to care about those whose suffering we can acknowledge. The fairness/cheating foundation provides the additional basis of impartiality, of saying that we should regard "the good of any other individual as much as his own." This is because fairness typically involves a tit-for-tat, a sense equal consideration for all parties that engage in an exchange.

Certainly, Haidt's theory is open to empirical revision, but it seems a more likely explanation than the one that Lazari-Radek and Singer offer—that some moral judgments are purely explained as spandrels of reason. What is more, Haidt's is not the only theory that emphasizes *care/harm* and *fairness/cheating*. In fact, what distinguishes Haidt's theory—and what might render it controversial—is not that it includes these two

---

[183] "Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism," *Advances in Experimental Social Psychology* 47 (2013): 55–130, p. 57.
[184] Jonathan Haidt and Craig Joseph, "Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues," *Dædalus* 133, no. 4 (Fall 2004): 55–66.
[185] Graham et al., "Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism."

receptors, but that it also includes other three. In other words, most of the alternative theories of the psychological bases of morality are monistic and emphasize either a sense of justice,[186] a sensitivity to harm,[187] related notions of generalized human welfare,[188] or a combination of these.[189] Therefore, an empirical assault against the Moral Foundations Theory would not be enough; one would also need to show why any theory that includes *care/harm* and *fairness/cheating* receptors is incorrect as well.

Lazari-Radek and Singer may object that the moral receptors we discussed before might explain altruism, but not universal altruism, because this later concept makes no evolutionary sense. This might be true for moral receptors considered individually, since the *care/harm* receptor would lead us to be concerned only about our genetic relatives, and the *fairness/cheating* receptor does not, by itself, motivate feelings of altruism, unless to correct an injustice. But I provided an account of how these receptors *combined* would lead to universal altruism, so the objection cannot be that Sidgwick axiom is left unexplained by out evolutionary-influenced faculties. Moreover, even if it were true that these moral receptors are not enough to explain Sidgwick's axiom, I believe it would not be enough for Lazari-Radek and Singer's case. They have a stronger burden of proof: they need to show not only that Sidgwick's axiom of benevolence is a spandrel of reason, but that it is *exclusively* a spandrel of reason. Accordingly, it might be true that we grasped universal altruism partly through reason, but this does not mean that we did not need some moral receptors. Reason needs something to work on, and one way in which we have reached universal altruism is probably by expanding our initial moral emotions of care and fairness towards all sentient beings, rendering universal altruism as a spandrel of *both* reason and moral receptors. We could certainly go back-and-forth in an endless discussion of whether the axiom is really based on reason or not, but it seems to me that theirs is not the obvious explanation and that the one I provided deserves at least a response.

---

[186] L. Kohlberg, "From Is to Ought: How to Commit the Naturalistic Fallacy and Get Away with It in the Study of Moral Development," in *Psychology and Genetic Epistemology*, ed. T. Mischel (New York: Academic Press, 1971), 151–235; Baumard N. André and D. Sperber, "A Mutualistic Approach to Morality," *Behavioral and Bran Sciences* 36 (2013): 59–122.
[187] K. Gray, L. Young, and A. Waytz, "Mind Perception Is the Essence of Morality," *Psychological Inquiry* 23 (2012): 101–24.
[188] Sam Harris, *The Moral Landscape: How Science Can Determine Human Values* (New York: Free Press, 2010).
[189] E. Turiel, *The Development of Social Knowledge: Morality and Convention* (Cambridge, England: Cambridge University Press, 1983).

Lazari-Radek and Singer might also reply that Sidgwick does not ground his axiom by appeal to any moral receptors, only by the use of reason. But while it could be true that the axiom is solely grounded on reason, there is a difference in saying that it is *explained* by reason. Recall the distinction between the context of discovery and justification. Lazari-Radek and Singer are claiming that Sigdwick's axiom is *justified* by reason (the context of justification), but what matters to us in evolutionary debunking is whether it is *explained* by reason (the context of discovery). The basis for evolutionary debunking was the claim that those judgments that are explained by evolution should be distrusted, and this is why Lazari-Radek and Singer try to argue that Sidgwick's axiom is not explained by evolution. So even if they are right in that the Sidgwick's axiom is justified by reason, this does not show that it is also explained by reason, which leaves the axiom open for evolutionary debunking.

One last thing to note about the previous discussion is that it also reveals a strong disanalogy among judgments that are spandrels of reason. Lazari-Radek and Singer's claim was that it is likely that spandrels of reason are truth-tracking *because* we can grasp mathematical and logical truths through reason. Consequently, if some moral judgments are also spandrels of reason, we could assume that they are also probably moral truths. This inference might work, however, only if mathematical and logical truths were of the same nature as moral truths. Huemer, for example, says that "moral intuition differs from mathematical intuition in the way that perceptions of cars differ from perceptions of trees—that is, merely in having different objects."[190] But moral and mathematical judgments do not seem to vary only with regard to their object—moral receptors surely also make them different. A judgment like "p and not-p cannot be both true at the same time" is not of the same nature as a judgment that includes notions of moral obligations, the good, and concern for the good of others. The difference, again, is probably partly due to the fact that the former does not involve any moral receptors whereas the latter does, as it is likely to assume from our discussion of Haidt's theory. If this is correct, then Lazari-Radek and Singer's inference falls apart: we cannot assume that because some spandrels of reason are likely truth-tracking, all of them will be, especially if there are important

[190] Huemer, *Ethical Intuitionism*, p. 215-216.

differences between these spandrels and if they are also partly explained as spandrels of other faculties.

The fourth and last problem with the genetic argument of Lazari-Radek and Singer is that is seems that it cannot avoid the genetic fallacy, as there seems to be a difference between the context of discovery and the context of justification. Consider this judgment mentioned by Nye: *we have most practical reason not to hit ourselves with a hammer when this would serve no further purpose*. Nye says that this judgment might well be explained as an evolutionary adaptation (the context of discovery), but why should we care, he asks, if we can defend it by a defensible principle (the context of justification), namely, the fact that an act will cause one pain is (at least ceteris paribus) a practical reason for one not to perform it.[191] Thus, it seems that the judgment cannot be dismissed merely because it can be explained by evolution; rather, it should be assessed by the reasons we can give for or against it. In other words, we should check whether it is consistent with other plausible beliefs we would hold, and now we are getting quite close to reflective equilibrium.

Granted, as Ward and Nye point out, genetic arguments can give us reason to think that the premises one *would* use in supporting a claim are false or that our evolution-based prejudices might be leading us to blindly accept something. This might be the case for some case-intuitions that are sometimes carelessly thrown against Singer's views without further argument, such as that life is "sacred" (whatever that means) or that animals are "inferior" to humans. It is also likely that this happens with regard to the alarm-type emotions that Greene thinks drive deontological judgments. However, we should be careful to recognize that the real problem lies in that we are not able to give additional reasons to support the judgment in question—or should I say, to adopt it in reflective equilibrium—rather in merely pointing its evolutionary origin, which, nonetheless, might be valuable as a heuristic to know which judgments might not withstand scrutiny.

[191] Nye, "Directly Plausible Principles", p. 623-624.

* * *

With this said, I conclude Chapter 3. In it, I argued that none of the proposed genetic arguments render case-intuitions as systematically unreliable. Genetic arguments, at any rate, serve as heuristics, as a signal for judgments that we would not likely be able to justify after careful consideration. The ultimate judge for whether an intuition or judgment is reliable or not, however, remains its ability to withstand scrutiny, by considering of reasons in favor of against it. This process is mainly reflective equilibrium, and it thus remains the best way we have to test intuitions and the judgments triggered by them. In the next chapter, I address one last concern that could be raised against about case-intuitions: their stability. In a certain way, this concern links nicely with genetic arguments, because they hint at the idea that what is problematic about case-intuitions is that they are arbitrary: culture and evolution could have turned out differently, leading us to hold different intuitions. Let us look at this issue in more detail.

Chapter 4 — Instability and Morally Irrelevant Features

In this last chapter, I will consider a final objection that can be made against case-intuitions: that they are unstable. Roughly speaking, the claim is that case-intuitions are unreliable because they change in response to features which are arguably morally irrelevant, such as the wording of a case, the emotions we experience while thinking about a case, or even the passing of time. For example, in Schnall's study, which we discussed in Section 3.1, participants evaluated more severely certain moral judgments about marriage, sex, and public policies if they were experiencing feelings of disgust, which were induced, for instance, by a commercially available "fart spray." [192] Presumably, this feeling of disgust should not make a difference in our response or in the intuitions we have. But it does. And given that there are that case-intuitions seem particularly prone to the influence of these arguably morally irrelevant factors, we might have reason to be suspicious of them.

Peter Singer articulated this argument in discussing the work of Peter Unger. Singer says that Unger presents the reader with various cases which differ only in morally irrelevant features, and yet our responses to the cases change, and it is hard to reconcile them with each other. Unger then explains why we have these contradictory responses, and, according to Singer, this "is devastating for the view that we should take our intuitive responses to particular cases as the test of a sound theory, because the explanations show that our intuitive judgments are based on things that are obviously of no moral significance at all."[193] In other words, the argument seems to be that since case-intuitions change in response to morally irrelevant features, we have reason to distrust them systematically.

I should note that this type of argument might not be precisely phrased in the way I did, as it often comes in disguise. As we saw with the genetic argument based on emotions, one interpretation as to why emotions might be problematic is that they are the primary morally irrelevant factors that leads us to change our response to a case when everything else is held equal. Although various authors often make an argument of this kind against some intuitions, it should not surprise us if we do not encounter the word "instability"

[192] Schnall et al., "Disgust as Embodied Moral Judgment."
[193] Singer, "A Response", p. 316.

when this argument is made, since what is often discussed are particular instances of an instability and it can also be hard to pinpoint what exactly does this instability involve.

In the next section, I will begin by clarifying what exactly could we mean when we say that intuitions are "unstable." The section after the next argues that instability by itself does not show that case-intuitions are unreliable; it only shows that we cannot consistently hold conflicting intuitions. The final section defends that claiming that a feature is morally irrelevant or not presupposes an ethical principle, which is presumably grounded presumably by using reflective equilibrium

4.1 The instability of intuitions

What do we mean when we say that an intuition or a judgment is "unstable"? Consider what Rawls mentions in one of his earlier works:

> It is required that the judgment be stable, that is, that there be evidence that at other times and at other places competent judges have rendered the same judgment on similar cases, understanding similar cases to be those in which the relevant facts and the competing interests are similar. [194]

In this passage, Rawls refers to an agreement between judges, that is, a criterion about intersubjective agreement for a judgment to be stable. We discussed previously, however, why disagreements between people would not be problematic for reflective equilibrium (Section 2.4), so let us put this feature aside. A second aspect that Rawls mentions is that the agreement has to be about similar cases, that is, about cases that share relevant facts. In other words, our judgments are not stable if they change between cases which vary only with regard to morally irrelevant features. One example of a morally irrelevant feature would be the mere passing of time, as Rawls notes:

> The stability must hold, by and large, over the class of competent judges and over their judgments at different times. Thus, if on similar cases of a certain type, competent judges decided one way one day, and another the next, […] then none of

[194] Rawls, "Outline of a Decision Procedure for Ethics", p 183.

> these judgments would be stable judgments, and therefore none would be considered judgments.[195]

To be clear, Rawls does not mention explicitly that the mere passing of time is a morally irrelevant feature, but I think this is a charitable interpretation for his account, for it seems like the reason why temporal stability matters is because the passing of time should not be the deciding factor in changing our judgment. Let us, then, summarize Rawls's account as follows: *stability involves holding the same judgment over cases that share morally relevant features* (and the mere passing of time is not one of them). Contrarily, instability involves either not holding the same judgment over cases that share morally relevant features or holding the same judgment over cases that vary only with regard to morally irrelevant features.

An illustration of this kind of instability comes from Lazari-Radek and Singer's discussion of the experience machine. In the original formulation of this thought experiment, we are asked if we would plug into a machine that can give us any experience that we desire: writing a great novel, making a friend, or reading an interesting book. Although we would be floating on a tank with electrodes attached to our brain all the time, we would not know we are there, we would think everything is actually happening, and there would never be practical malfunctions with the machine.[196] Many people have the intuition that we should not plug into the machine. However, their intuitions seem to vary once we vary some arguably morally irrelevant elements of the case. In Greene's formulation of the thought experiment, we suddenly wake up in a white room, and a woman tells us that the life we have been experiencing is a machine program selected by us forty years ago. We are awakened at ten-year intervals to ensure that we are satisfied, and the records show that at three previous interruptions we have deemed the program satisfactory. Would we plug in again? In response to this scenario, many more people are willing to. But Lazari-Radek and Singer argue that there are not any morally relevant differences between the two cases, so we should not give much trust to our case-intuition about these cases.[197] According to Rawls's account, this could be considered a case of

---

[195] Rawls, p 183.
[196] Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974), p. 43.
[197] Lazari-Radek and Singer, *The Point of View of the Universe*, pp. 254-261.

instability in our intuitions, because we did not hold the same judgment in cases that shares morally relevant features.

Rawls's considerations also apply to other empirical studies that presumably show the instability of our intuitions, although some of these studies a different in one respect: that the changing judgement is about the same case, not between similar cases. Schnall's study is one such example. In the study, participants were presented with the *same* set of moral judgments, but their responses varied depending on the feelings of disgust they experienced.[198] Studies on framing and ordering effects constitute other such examples. Tversky and Kahneman, for example, found that *the same* hypothetical policy was more likely to be accepted if described as saving 200 people (out of a total of 600) than if it was described as letting 400 people die. Also, Liao et al. found that subjects' responses about a set of trolley dilemmas changed depending on what dilemmas they had considered before.[199] Lastly, participant's response to a same set of moral decisions changed depending on how specific the identity of a person was. In a study by Small and Loewenstein, participants were given 10 dollars, and then drew random cards to determine whether they could keep the money (a nonvictim) or not (a victim). Nonvictims were then paired with victims and were allowed to give a portion of their money. Yet, nonvictims that decided how much money to give *before* drawing a specific victim's number (person #?) gave 60% less than those that decided how much money to give *after* drawing a victim's number (e.g., person #4).[200] But this, Greene says, is absurd, and it is explained by the levels of sympathy and pity, which are morally irrelevant.[201] Moreover, Rawls consideration about the passing of time could also refer to a single case. For example, we could be presented with Nozick's experience machine but change our response daily, holding all else equal. We would then have an instance of a varying

---

[198] Schnall et al., "Disgust as Embodied Moral Judgment."
[199] A. Tversky and D. Kahneman, "The Framing of Decisions and the Psychology of Choice," in *Environmental Impact Assessment, Technology Assessment, and Risk Analysis* (Berlin: Springer-Verlag, 1985); S. Liao et al., "Putting the Trolley in Order: Experimental Philosophy and the Loop Case," *Philosophical Psychology* 25 (2012): 661–71; cited in Nye, "Directly Plausible Principles," p. 622.
[200] D. A. Small and G. Lowenstein, "Helping a Victim or Helping the Victim," *Journal of Risk and Uncertainty* 26 (2003): 5–16.
[201] Greene, "The Secret Joke of Kant's Soul", pp. 49-50.

response about one same case caused by a morally irrelevant factor, which seems to render our intuitions suspicious.

In summary, we can say that the stability of an intuition involves one of these two situations: 1) holding the intuition over one case even when other morally irrelevant features vary, or 2) holding the intuition over cases which share morally relevant features. If one of these two conditions is not met, the foundationalist might say, we would have reason to suspect the intuition in question.

4.2 Does instability undermine case-intuitions?

The first problem for arguments based on the instability of case-intuitions is that even if we concede that there can be situations in which we have conflicting case-intuitions, it is not clear why we should conclude that both of them are unreliable and, moreover, that all case-intuitions are.

To examine this, consider Tversky and Kahneman's study about framing effects. The authors presented various participants with a policy item and asked them to evaluate which one would they prefer. The experimental manipulation was that, to some participants, the policy was described in terms of "200 people will be saved" (out of 600) and, to other participants, in terms of "400 will die." Participants' responses varied, even though the outcomes of the policies were identical.[202] The foundationalist might conclude that studies like this show that we should withhold judgment when case-intuitions are involved and that we should render case-intuitions as unreliable. But this conclusion seems unwarranted. The three intuitions that are in conflict when we think about Tversky and Kahneman's study are presumably the following:

(a) We should implement the policy that saves 200 lives.
(b) We should not implement the policy that lets 400 people die.
(c) The wording of a case is morally irrelevant, so it is absurd to change one's evaluation of a case based on it.

[202] Tversky and Kahneman, "The Framing of Decisions and the Psychology of Choice."

Participants held intuitions (a) and (b), and this seems wrong given that intuition (c) seems highly plausible. But this does not show that (a) and (b) should be deemed unreliable; it just shows that people should not consistently hold both. A revision is called for, and the person has to choose between holding (b) and (c), or (a) and (c). We see, then, that we are not forced to abandon both case-intuition, as one of these intuitions might be more justified than the other, and, to determine that, we should consider further cases and principles.

What grounds would we have for making the stronger claim that both case-intuitions are unreliable? The argument seems to go something like this: case-intuitions often lead us into situations where there are apparent inconsistencies between our beliefs, situations in which there is a highly plausible principle which cannot be held in conjunction with case-intuitions about identical or seemingly similar scenarios. Phrased like this, however, the argument seems too weak. The presence of conflict between beliefs or the tendency to lead to a conflict cannot be a sufficient reason to dismiss case-intuitions. At any rate, this argument would only warrant caution when making judgments about particular cases and further examination of our beliefs.

To see the implausibility of this kind of reasoning, imagine that we made a similar charge against principle-intuitions. Assume we had a set of participants to whom we asked whether this transitivity principle is correct or not: *If A is better than B, and B is better than C, then A is better than C*. Presumably, most participants will say that it is correct. But assume that we made an experimental manipulation and gave a hallucinatory drug to some participants which drastically altered their cognitive abilities. Presumably, the responses of this second group will be different from those of the first group, even if the majority still thinks that the transitivity principle is correct. Given the slight change in the responses, however, an argument about the instability of the principle-intuition could be made on the basis of these three responses:

(a) The transitivity principle is correct.
(b) The transitivity principle is incorrect.
(c) The influence of a hallucinatory drug is morally irrelevant, so it is absurd to change one's evaluation of the transitivity principle based on it.

If we followed the same reasoning as the foundationalist regarding case-intuitions, we would have reason to dismiss (a) and (b). But this is not an obvious conclusion, as there are other reasonable options. One of them would be to assess each intuition independently and figure out which one we want to keep. Another option would be to just dismiss the intuitions which was made under the influence of the hallucinatory drug. Thus, we cannot conclude that any type of intuitions are unreliable merely based on the fact that they are often in conflict.

4.3 What is a morally irrelevant factor?

Let us now address a second problem for arguments based on the instability of case-intuitions. As we discussed before, arguments based on the instability of intuitions rely on a fundamental idea: morally relevant or irrelevant factors. Although some authors think that it is easy to tell what matters and what does not, the story is not as straight-forward. Arguably, what we believe is morally relevant or not depends on an underlying ethical principle. R. M. Hare—who had a considerable influence on Singer—emphasized this relation between relevance and principles in one of his works:

> The consequences that we think morally relevant are going to be those which are mentioned in whatever moral principles we apply to the situation. For example, if we think that one ought not to kill any human being after conception, then we shall think it morally relevant that to administer a certain drug would kill an embryo. But if we do not accept such a principle, we shall not think it relevant. So the question of relevance boils down to the question of what moral principles we should accept.[203]

Let us illustrate Hare's remark by considering Singer's famous discussion of our duties to help people in other countries. Singer asks us to imagine a situation in which we could save a drowning child in a pond from death but that would involve ruining our expensive clothes. Surely, ruining our clothes is insignificant compared to the life of the child, so it seems that we ought to save the child. Yet, Singer argues that there are no

[203] R. M. Hare, *Essays on Religion and Education* (Oxford: Claredon Press, 1992), p. 68; quoted in Rhys Southan, "Peter Singer, R.M. Hare, and the Trouble With Logical Consistency," *Essays in Philosophy* 18, no. 1 (2017).

morally relevant differences between this case of a drowning child and the current situation of affluent people with regard to poor people in other countries: right now, they could donate some money instead of buying a luxury item (analogous to ruining our expensive clothes in the pond) and save the life of someone in a distant country (analogous to saving the child in a pond). Therefore, it seems that affluent people ought to donate money to save the lives of people in distant countries.[204] Singer's argument, however, only works if there are really no morally relevant differences between both situations. For instance, one obvious difference is the physical proximity in both situations. Singer argues that this is morally irrelevant, [205] but this is an intuition about a principle, namely, that *physical proximity is not morally relevant*. Southan also notes that there are other assumptions about what is relevant in play:

> What we categorize as morally significant will depend on other assumptions that Singer does not clearly address here. If moral significance reduces to pleasure and pain, the pleasure we derive from a favorite outfit could be morally significant, as could the pain of losing that outfit.[206]

In sum, we can see that when Singer says that something is morally relevant or not, he does so based on an intuition or belief in a further ethical principle: that a certain feature is morally relevant or not. Singer sometimes offers a partial list of what he considers to be morally relevant or irrelevant characteristics, such as species membership (irrelevant), pleasure and pain (relevant), rationality (relevant), the use of language (relevant), and autonomy (relevant).[207] Each of the elements of this list, in turn, can be thought of as expressing an ethical principle.

However, we now encounter one problem for arguments based on the instability of case-intuitions: how are the principles about what is morally relevant or not supposed to be grounded? They cannot be justified based on further consideration of their consequences or on careful examination of other related judgments—one way in which philosophers usually evaluate principles—for this will presumably involve an appeal to

---

[204] Singer, "Famine, Affluence, and Morality."
[205] Singer, *Practical Ethics*, p. 232.
[206] Southan, "Peter Singer, R.M. Hare, and the Trouble With Logical Consistency", p. 15.
[207] See Southan, p. 9.

case-intuitions. For example, to clarify whether autonomy is morally relevant or not, a foundationalist could not appeal to thought experiments or scenarios—as one normally does—for then he or she would be using case-intuitions as evidence to tell whether a principle seems plausible or not, and this would render his argument against case-intuitions as contradictory. In contrast, the reflective equilibriumist would have no problem in doing this. Although he or she might ultimately agree with Singer in that certain things are morally irrelevant, he or she would be using reflective equilibrium to reach an answer, and so figuring out what is relevant or not would presuppose reflective equilibrium.

The alternative solution of the foundationalist is to appeal to the plausibility of the ethical principle itself. As Huemer mentions, "How does Singer know that physical proximity is not morally relevant? The answer seems to be that that is 'obvious' or 'self-evident'—in short, it is an intuition."[208] If this is the only viable strategy, however, it is not evident that all principles about what is morally relevant or not are ubiquitously self-evident. Admittedly, some of them are highly plausible. For example, the mere passing of time, the wording or ordering of a case, and the sheer force of habit seem to be, indeed, morally irrelevant factors. But there is some room for doubt for whether other factors also are—factors that authors like Singer and Greene sometimes take for granted.

Consider the case of emotions. In his discussion of Small and Loewenstein's study, for instance, Greene suggests that some of the participants' responses are absurd if they are explained by the levels of sympathy and pity of the participants, as they seem to be.[209] These responses, however, would only be absurd if sympathy and pity are indeed morally irrelevant factor; otherwise, if they actually had moral weight, then the participants' responses would actually make sense and be justified. And we could say the same thing of Schnall's study. If the feeling of disgust participants experienced is morally relevant, then their responses would be justified. Of course, Greene could respond that it is simply absurd to think that these factors are morally relevant, but this seems overly dogmatic. And even if

[208] Huemer, *Ethical Intuitionism*, p. 152.
[209] Greene, "The Secret Joke of Kant's Soul", pp. 49-50.

it is likely that they are indeed morally irrelevant, it seems reasonable to leave some room for doubt, as it could be the case that we are being overconfident.

More problematically, Huemer mentions that if we understand a morally irrelevant property as one which affects whether a thing is good, bad, right, or wrong, subjectivists would say that this is simply a property that affects our attitudes. Therefore, if sympathy, pity, or disgust really affect our attitudes, then that just proves that they are morally relevant.[210] The natural response might be to argue that we are not talking about actual attitudes that people hold, but those attitudes they would have if fully informed and rational. However, reaching that ideal point arguably requires further examination of other moral and non-moral judgments, not the self-evidence of the principle in question, and this would get us into the territory of reflective equilibrium.

Consider now Singer's famous claim that species membership is a morally irrelevant feature. In his important book, *Animal Liberation*, Singer mentions that, "Speciesism […] is a prejudice or attitude of bias in favor of the interests of members of one's own species and against those of members of other species."[211] Granted, this is a highly plausible claim, as evidenced by the fact that many people find it compelling. But still, it seems that it should be at least open to revision, as someone might not find it plausible after reflection. Shelly Kagan, for instance, argues that Singer's attempt to show that speciesism is a prejudice is unsuccessful (although he still thinks that our treatment of animals is unjustified). Most importantly, he mentions that "if we are going to objectively evaluate the charge that speciesism is a mere prejudice and nothing more, we had better not build the claim that it is a prejudice into the very definition of the term."[212] This seems like a reasonable constraint, independently of whether we think that speciesism is ultimately shown to be a prejudice or not. Even Singer would certainly concede this point, as he offers a further principle in support of his charge against speciesism. He grounds his claim on a further principle, the *principle of equal consideration of interests*: "the interests of every being affected by an action are to be taken into account and given the same weight as

---

[210] See Huemer, *Ethical Intuitionism*, p. 152.
[211] Peter Singer, *Animal Liberation* (New York: Harper Colling Publishers, 2002), p. 6.
[212] See Shelly Kagan, "What's Wrong with Speciesism? (Society of Applied Philosophy Annual Lecture 2015)," *Journal of Applied Philosophy* 33, no. 1 (2016), p. 2.

the like interests of any other being."[213] Surely, this principle might not seem self-evident to many, although it might be after further examination. But Singer's response suggests that even highly plausible principles about morally relevant or irrelevant features should not be taken at face value without further consideration, which might involve looking at other plausible principles and case-intuitions. If we do this, however, we are likely in the territory of reflective equilibrium.

Why is all this relevant for the foundationalist argument against case-intuitions? Well, the argument about instability is based on the assumption that there is a set of morally irrelevant features that arguably influence case-intuitions. But now we see that this list might not as sufficiently extensive as to warrant systematic dismissal of case-intuition. Some of the factors in this list are surely uncontroversial: ordering and framing effects, the mere passing of time, etc. The rest, however, seem to require further examination of a type that presupposes reflective equilibrium. And if it seems reasonable, for instance, that the highly plausible principle-intuition in favor of speciesism should be open to revision, why would we not think the same of other principle-intuitions—some of which are less plausible—about morally relevant or irrelevant features? This presumably applies to the sort of morally irrelevant factors that foundationalists often present against case-intuitions, such as emotions (e.g., feeling of disgust, jealousy, outrage), alleged biases, physical proximity, and so on. Let me emphasize, again, that my point is not that these features are morally irrelevant—they might well be. My point is that we need to engage in reflective equilibrium to find that out.

The most forceful response to my argument is, I believe, to grant we need to examine other cases and principles to test a principle, but argue that there is no justificatory relation between the principle and the cases. The foundationalist could argue that cases are merely used to *illustrate* or *clarify* the principle in question, not as evidence in favor of the principle. Thus, there is no circularity, and reflective equilibrium—which states that the relation is evidential—need not be included as premise in the argument. Singer hints at this interpretation in a reply to F. M. Kamm. He states that general principles are not supported or undermined by intuitions about particular cases:

---

[213] Singer, *Animal Liberation*, p. 5.

> It is true that I have appealed to intuitive judgments about hypothetical cases, and sometimes also about imaginary ones, at various places in my writing. But it is one thing to generate intuitions about particular cases in order to show that a distinction widely believed to be morally significant is not really doing the work we ordinarily think it does in grounding our judgments of right and wrong, and another thing altogether to give these intuitions probative status.[214]

In this passage, Singer uses the word "show" to describe the function that case-intuitions play, yet "show" could mean various things, and Singer does not clarify exactly what he means. To be charitable to Singer, nonetheless, we should try and think what this non-probative relation between principles and cases might involve, and it is perhaps best to look at the account of Howard Nye, who has written on this issue and suggests that his approach is continuous with the one that Singer and other foundationalists like Sidgwick and Ross use in some of their works.[215]

Nye argues that case-intuitions serve two purposes: 1) they suggest principles to us, and 2) they illustrate or otherwise help us clarify what a principle is really saying.[216] The first purpose is straight-forward: a case might help us to think of a principle which did not occur to us before. Although this seems a likely role that case-intuitions play, it is probably not the non-probative relation that we are looking for in grounding principles—or the one that Singer refers to in the previous passage—as in the cases we discussed before, it seems that we already had a principle and we were just trying to see whether it was defensible or not. What interests us then is the second purpose that Nye mentions.

In his formulation, Nye refers to two main concepts: "illustrate" and "clarify." The role of clarifying is "to help us understand what the principles are really saying," and the role of illustration seems to be to bring out "the genuinely direct plausibility or implausibility of the principles."[217] This is exemplified by Nye when he makes reference to Singer's drowning child case and says, "The point of the Pond case was […] to give us a more concrete understanding of how plausible the idea contained in [Singer's proposed

---

[214] Singer, "A Response", p. 315.
[215] Nye, "Directly Plausible Principles," p. 613.
[216] Nye, p. 610.
[217] Nye, p. 630.

principle] really is saying – in itself and independent of what else it supports or entails."[218] It is unclear, however, whether it is even useful to distinguish between clarifying and illustrating, since illustrating can sometimes seem to have the function of clarifying, as when Nye says, "By giving a concrete illustration […], the Pond case helps us better appreciate what [Singer's principle] is really saying, in a way that amplifies its direct plausibility." [219] Thus, I take it that Nye's overarching idea—which is the one that interests me—is that cases allow us to fully understand what a principle is really saying and, in turn, bring out its direct plausibility or implausibility.

In general, I am sympathetic to Nye's approach, as he also opposes genetic arguments and favors critical scrutiny of our judgments, but I have to say that I have a hard time making sense of the idea that cases allow us to better *understand* a principle. I find this suggestion confusing because it seems to me that I fully understand some of the principles we have examined so far—it is just that I am unsure whether I would endorse them after reflection. Consider the case of speciesism again. Kagan articulates in the following way what might be the principle behind Singer's idea that speciesism is a prejudice:

> *Other things being equal, human interests count more than corresponding animal interests*. (That is, even when given interests that are otherwise similar, human interests get special consideration, more weight than the corresponding animal interests).

Stated like this, it seems to me that I have perfectly good understanding of this principle, even if I have not examined all possible instances of it. Now, I might imagine an illustration of the principle in which a human interest outweighs the equivalent interest of some animal, and this case-intuition might make me think that the principle cannot be plausible. But I would not say that the case allowed me to better *understand* the principle. I would say that it just presented me with a conflicting intuition that reduced my credence in the principle. Think about an analogous case with science. Newton's first law of motion states that *every object will remain at rest or in uniform motion in a straight line unless compelled*

[218] Nye, p. 630.
[219] Nye, p. 626.

to change its state by the action of an external force. According to the idea that an illustration of a principle allows us to better grasp its content, if I see an apple falling, presumably, I would get a better understanding of the law. But surely, I understood altogether what Newton's law meant—the falling apple was just an instance of it.

What might Nye mean, then, by *understanding* what a principle is really saying? Understanding cannot refer to what a judgment entails. Nye clearly states that understanding a principle is "independent of what else it supports or entails," and it is also implausible to think so. If we had the principle *Suffering is wrong*, to fully understand it we would have to consider the infinite number of real and imaginary cases to which it would apply, but this would not only be impossible, but also arguably unnecessary to have an understanding of the principle.

Understanding might involve some clarification of the terms in a principle. For example, if I had the principle *Contraception is unnatural*, cases might allow us to clarify what do we mean by "unnatural": does it refer to statistical deviation, to alignment with the goals of evolution, or to something else? Yet, although this is a useful task—and one that reflective equilibrium also engages in—it cannot be the whole story, for there are times when cases do not seem to further clarify concepts. Assume, for instance, that a person disagreed with the principle *physical proximity is morally relevant* and that he had a rough definition of each of the terms in this principle. He might define *physical proximity* as the distance between two objects in space and *morally relevant* as a something which affects whether a thing is good, bad, right, or wrong. Assume now that he was presented with Singer's drowning child case and Unger's envelope case. Recall, in the drowning child case, we are asked to imagine a situation in which we could save a drowning child in a pond from death but that would involve ruining our expensive clothes.[220] In the envelope case, we are asked to imagine a situation in which we receive an envelope from a trustworthy aid organization asking us to donate money that will help save various lives (and assume that this will be the case).[221] The person, then, might realize that he is willing to save the child but not donate to the aid organization, and think this is inconsistent because there are no

[220] Singer, "Famine, Affluence, and Morality", p. 231.
[221] Peter Unger, *Living High & Letting Die: Our Illusion of Innocence* (Oxford: Oxford University Press, 1996), p. 25.

morally relevant differences between both cases. The person would now have changed his credence in the principle *physical proximity is morally relevant*. But did the person modify the definitions of the terms of the principle throughout the process? No, he just seems to have changed his mind about whether physical proximity is morally relevant.

Understanding could also involve the important task of avoiding conflation of two concepts. In our previous example, the person might have initially said that physical proximity is morally relevant because he was confounding it with *one's ability to help* or *chances of succeeding*. To show this, we could ask the person to imagine that he has the superpower of stretching his arms into incredible lengths (like Mister Fantastic of the Fantastic Four), even miles, countries, and planets apart. We could then present him with a new drowning child case: there is one baby drowning right next to him, as he is walking by a pond, and two other drowning babies at the other end of the pond. Given his superpower, he could save either set of babies without much effort, although not both. But if physical proximity really had moral weight, then the person would have to discount the distance from other competing moral considerations to determine which set of babies to save. If the distance was not enough to prefer saving the single baby over the other two, we could always increase the distance and say the other two babies are in another pond, in a distant ocean, or in another planet, as his superpower would allow him to always save the two babies no matter how far. The person might then come to realize that physical proximity would be irrelevant, as he ought to save the two babies, and that he might have been confounding it with other factors.

Finally, understanding a principle might involve an appreciation of its vividness. By *vividness*, I mean appreciating the degree of intensity of a factor or some degree of what is colloquially refered to as "visualizing," "seeing in the mind's eye," "hearing in the head," or "imagining the feel of."[222] Granted, this is needed to appreciate some principles. For example, to appreciate the principle *Suffering is wrong*, we require some direct knowledge of what suffering feels like or an imaginative appreciation of an amount of suffering which we have not experienced directly or of a different kind of suffering that other persons or

[222] For a more detailed discussion, see Nigel J.T. Thomas, "Mental Imagery," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Spring 2018 (Metaphysics Research Lab, Stanford University, 2018), https://plato.stanford.edu/archives/spr2018/entries/mental-imagery/.

sentient beings might experience.[223] Also, lack of vividness might be the reason why we reject some principles. Consider, for example, the discussion around the Impersonal Total Principle, which states that, *If other things are equal, the best outcome is the one in which there would be the greatest net sum of happiness minus misery*. Against it, Derek Parfit pointed out that this principle implies what he called the Repugnant Conclusion:

> For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better, even though its members have lives that are barely worth living.[224]

Many people, including Parfit himself, take this implication to be a knock-down objection to the Impartial Total Principle. But some authors have suggested that this conclusion is premature,[225] and Ord argues, for example, that we might reject the implication because it is hard to imagine a world with trillions or more people. Our intuitions, Ord says, become shaky, and imagining "a trillion people" rather than "a billion people" becomes almost identical, so we might be undervaluing outcomes.[226] In this way, the lack of vividness of both the Impartial Total Principle and its Repugnant Conclusion makes us reject the former; thus, we might be inclined to say that, without vividness, we do not understand what the principle really says.

At this point, however, I am simply unsure whether there is a substantial or a verbal dispute between this version of foundationalism and reflective equilibrium. Nye will say that cases allow us to clarify definitions, to avoid conflation, and to highlight the vividness of a principle; the reflective equilibriumist will agree with Nye and add that cases also serve as instances of principle. But the practical way in which the procedure is conducted is almost identical. Consider someone who endorsed the speciesist principle (and there are various people who do, though it might be before reflection), that *other things being equal,*

---

[223] Rawls makes a similar point and says that a good judge should be able to imagine the interests of others. See Rawls, "Outline of a Decision Procedure for Ethics," p. 179.
[224] Parfit, *Reasons and Persons*, p. 387-388.
[225] T. Tännsjö, "Why We Ought to Accept the Repugnant Conclusion," *Utilitas* 149, no. 3 (2002): 339–59; Huemer, "In Defence of Repugnance."
[226] Toby Ord, "Overpopulation or Underpopulation," in *Is the Planet Full?* (Oxford: Oxford University Press, 2014), 46–60.

human interests count more than corresponding animal interests. One could try to convince this person that the principle is implausible by providing presumably conflicting case-intuitions. For example, one could ask the person to think of a very smart chimpanzee or of an intelligent being from another planet such as ET (from the movie) or Superman—the superhero—whose interests seem to count although he is not a *homo sapiens*.[227] What would we say is going on here? The foundationalist will say that the cases are making more vivid the implausibility of the speciesist principle and that they might point out a conflation between species and intellectual ability. The reflective equilibriumists will say that the cases might highlight a conflation and some vividness but also that they are counterexamples to the principle.

Nye's interpretation is a plausible one, but it just seems to me to be less parsimonious. I see no reason to prefer it over the much simpler and more intuitive account than the evidential one. And I should emphasize, again, that based on what I argued in Part I, the burden of proof is on the foundationalist to show that the relationship is not evidential, and I just find it hard to see why this would be the case.

I believe the foundationalist might be so reluctant to concede that there is an evidential relation because of a further reason: a concern about the revisionary power of reflective equilibrium. From Nye's discussion of what a slaveholding English aristocrat would have reason to believe, one can infer that his concern is that we might endorse implausible principles if we concede evidential status to intuitions.[228] Singer has also expressed similar concerns about the revisionary potential of reflective equilibrium.[229] These worries, however, seems misplaced to me. Firstly, it is not even clear that the foundationalist can avoid a similar charge so easily. Principle-intuitions, as Huemer mentions, are prone to the problem of overgeneralization, which is "the tendency to judge the truth of a generalization in terms of typical cases, or the sorts of cases that are easy to think of."[230] If this is correct, then it is not so clear that principles will allow us to be as revisionary as foundationalists want, as we will probably think of cases that habitually

---

[227] The ET and Superman examples are taken from Kagan, "What's Wrong with Speciesism? (Society of Applied Philosophy Annual Lecture 2015)," p. 9.
[228] Nye, "Directly Plausible Principles", pp. 628-631.
[229] See Singer, "Sidgwick and Reflective Equilibrium", p. 515.
[230] Huemer, "Revisionary Intuitionism", p. 384.

come to us, either to refute or support a principle. Maybe this explains why some foundationalists were not as thorough as they thought they were, as Hurka maintains: "Moore, Prichard, and Ross often made more confident moral claims than they should have, backed by overly hasty assertions of consensus. […] Sidgwick did the same with his moral axioms, which he did not subject to nearly as careful scrutiny as he did deontological principles and for which he too claimed more agreement than is plausible."[231] Of course, the foundationalist will reply that the fact that some people carry out the method incorrectly does not show that the method is incorrect, but the same thing could be said of reflective equilibrium.

It seems to me that much of the alleged attacks against reflective equilibrium consist, instead, of attacks on the improper use of reflective equilibrium, in which we only consider a narrow set of judgments. If we do this, certainly, it will be hard for a racist, a sexist, or a slaveholder to change his views, as the person will only consider those judgments that confirm what he or she thinks. But, as I mentioned, in reflective equilibrium we ought to seek the systematization of our intuitions about as many judgments as we can—both non-moral judgments and moral judgments—not only the ones that we currently have or that uncritically occurred to us. By doing this, the racist, the sexist, and the slaveholder are then forced to consider other ethical theories or plausible alternatives to their views which they have not considered yet; thus, their best systematization need not be the theory they currently hold—it might as well undergo a radical change. Obviously, considering all possible judgments is impossible to do from a practical standpoint, and one could even be doubtful that such a set of judgments could even be well-defined. But this is intended as an ideal aim, and perhaps the most we can do as a way of heuristic is to consider the most well-established positions. We might not reach the ideal, perfect state of equilibrium, but we need not conform ourselves to an uncritical equilibrium either.

Perhaps a more substantial concern about reflective equilibrium is that it could lead us to reject a highly plausible principle only because it conflicts with many case-intuitions. Frances Kamm's discussions might be the best illustration of why foundationalists are

---

[231] Hurka, *British Ethical Theorists from Sidgwick to Ewing*, p. 118.

reluctant to give evidential weight to case-intuitions,[232] as Kamm willingly adopts principles which might not seem so plausible by giving great weight to case-intuitions.[233] This is certainly an important problem, but there are at least three ways to avoid it. The first one is simply to avoid giving strong coherence probative value, as I discussed in section 1.3, because this will definitely give us stronger reasons to prefer a set of case-intuitions over plausible principles. The second one is to give proper weight to principle-intuitions. As Nye mentions, Kamm seems to give only lip service to this requirement, but this is not how reflective equilibrium ought to be conducted.[234] Lastly, and perhaps most importantly, the weight that case-intuitions have need not be static. Based on what we have discussed, one might think that in reflective equilibrium we merely add up the individual weight of case-intuitions and see whether it outweighs that of the principle. But this does not mean that the weight of either cases or principles cannot be modified throughout the process. For example, one might have initially thought that we have no duties of aid towards people in other countries based on the Envelope case; however, after being presented with Singer's principle[235] and the Pond case, one might change his or her attitude towards the Envelope case. Thus, the decision to whether or not accept Singer's position was not based on a static weight of the Envelope case, but how plausible it seems once other principles and cases were considered.

A good example of how reflective equilibrium can be revisionary is Peter Unger's discussion of our duties to help people in other countries. After considering various cases, Unger repeatedly concludes that many of our intuitions respond to morally irrelevant

---

[232] In fact, Singer and Nye have criticized Kamm's method. See Singer, "A Response"; Nye, "Directly Plausible Principles", pp. 627-628.

[233] As an illustration of what I mean, consider Kamm's Doctrine of Productive Purity (DPP), which has these two absurdly large clauses: (1) If an evil* cannot be at least initially sufficiently justified, it cannot be justified by the greater good that it is necessary (given our act) to causally produce. However, such an evil* can be justified by the greater good whose component(s) cause it, even if the evil* is causally necessary to help sustain the greater good or its components. (2) In order for an act to be permissible, it should be possible for any evil* side effect (except possibly indirect side effects) of what we do, or evil* causal means that we must use (given our act) to bring about the greater good, to be at least the effect of a good greater than it is working itself out (or the effect of means that are noncausally related to that greater good that is working itself out). F. M. Kamm, *Intricate Ethics: Rights, Responsibilities, and Permissible Harm* (Oxford: Oxford University Press, 2007), p. 164.

[234] Nye, "Directly Plausible Principles", p. 627.

[235] "If it is in our power to prevent something bad from happening, without thereby sacrificing anything of comparable moral importance, we ought, morally, to do it." Singer, "Famine, Affluence, and Morality", p. 231.

features.[236] Singer thinks Unger's conclusions pose a challenge to reflective equilibrium: "[…] the view that we must test our normative theories against our intuitions […] now it faces its most serious challenge yet, in the form of Peter Unger's *Living High and Letting Die*."[237] He then mentions, "Clearly, if Unger is right, the method of doing moral philosophy that relies on our intuitive judgments of particular cases is in tatters."[238] However, although Unger thinks that case-intuitions are not often reliable guides, he does not venture to say that all of them are. [239] In fact, Unger reaches his conclusions not by dismissing case-intuitions from the beginning, but by engaging in a process similar to reflective equilibrium. As he mentions,

> […] we'll note some factors that do differentiate between our puzzle cases. […] In trying to answer, each time we'll consult two main guides. On the one hand, we'll note our moral intuitions on particular cases. On the other, we'll note the deliverance of what I'll call our general moral common sense, since this second sensibility is directed at matters at least somewhat more general than the first's proper objects.[240]

This goes to show that reflective equilibrium and revision are not incompatible and that perhaps the best way to show that a case-intuitions is unreliable is to subject it to further scrutiny.

---

[236] Unger, *Living High & Letting Die: Our Illusion of Innocence*.
[237] Singer, "A Response", p. 316.
[238] Singer, p. 317.
[239] Unger, *Living High & Letting Die: Our Illusion of Innocence*, pp. 10-12.
[240] Unger, p. 28.

Conclusions

In this thesis, I argued why reflective equilibrium should be the default method for ethical inquiry. In Part I, I presented a four-premised argument in favor of it:

> *Premise 1:* If a judgment seems plausible to an agent—in other words, if an agent has an intuition regarding that judgment—then, other things being equal, an agent should assign credence to that judgment.
>
> *Premise 2:* Other things being equal, an agent should assign credence to a judgment proportional to the degree of plausibility that the judgment has for the agent.
>
> *Premise 3:* Judgments of all levels of generality can seem plausible to an equal degree.
>
> *Premise 4:* At the outset of inquiry, there is no reason to think that all case-intuitions or principle-intuitions are unreliable in a systematic way.
>
> *Conclusion:* Other things being equal, an agent should assign credence to judgments of all levels of generality—case-intuitions and principle-intuitions—depending on the degree of plausibility that those judgments have for the agent.

I defended the first premise by arguing that rejecting it leads to implausible implications and to an indirectly self-defeating position. The second premise was defended by appeal to our everyday and philosophical experience and, also, by showing that rejecting it leads to implausible implications. Support for the third premise was found by providing plausible examples of both case-intuitions and principle-intuitions and by arguing that even if one type of intuitions is ultimately more plausible, this presupposes reflective equilibrium. Finally, the fourth premise gains support from the way in which we use case-intuitions and principle-intuitions in ethical debates. I defended this premise against the charge that case-intuitions are not universally accessible by arguing that this is not obvious, that principle-intuitions share the same problem, and that we have to engage in reflective equilibrium to find that out.

Part II was an extended discussion of premise 4. In Chapter 3, I argued that genetic arguments based on emotional, cultural, and evolutionary influences do not warrant systematic dismissal of case-intuitions. In section 3.1, I argued that the problematic aspect of emotional influences could not be grounded on (a) its direct plausibility, because of the genetic fallacy; on (b) an analogy with factual judgment, because it is empirically questionable whether emotions bias factual judgment and, even if they do, because there is a disanalogy with moral judgments; and on (c) Joshua Greene's account, because it is based on further considerations about the instability of intuitions and evolutionary debunking. In section 3.2, I argued that a genetic argument based on cultural influences falls prey to the genetic fallacy and to it being self-defeating. In section 3.3., I argued, against Lazari-Radek and Singer, that their evolutionary debunking argument presupposes moral realism, that it is based on the vague notion of "reason," that principles which presumably avoid evolutionary debunking can be explained by Haidt's theory of moral receptors, and that there is a genetic fallacy. Finally, in Chapter 4, I mentioned that the argument based on the instability of intuitions does not warrant a dismissal of case-intuitions and that saying that something is morally relevant—the basis of this type of arguments—presupposes moral principles which arguably need to be examined in reflective equilibrium. I ended by discussing Nye's account—which I consider the best alternative to reflective equilibrium—but I argued that it is not parsimonious, that in practice it might be indistinguishable from reflective equilibrium, and that the burden of proof is on his side.

There remain people who are skeptical of reflective equilibrium, however. In discussion of this method, Huemer mentions that we can attempt to "apply the traditional reflective methods of ethics very carefully, in the hope that they will weed out the worst biases and distortions in our ethical judgments. The latter reaction strikes me as overly complacent."[241] As I said, much of this skepticism about reflective equilibrium is misplaced, but I do believe the skeptics force us to be on the alert for situations in which reflective equilibrium is not conducted properly. One such instance is when a single counterexample is presented as an attempt to dismiss a theory. This sometimes happens, for example, with case-intuitions that conflict with consequentialism or utilitarianism, such as the Transplant

[241] Huemer, "Revisionary Intuitionism", p. 378.

case.[242] Perhaps this is why some utilitarians have enormous distrust for case-intuitions, especially Singer, who often engages in public debates where appeals to isolated case-intuitions are ubiquitous, which can be frustrating. Nevertheless, reflective equilibrium, conducted properly, assesses judgments by their overall standing in relation to other judgments, including those principles under which case-intuitions can be subsumed. So it is not enough to present an isolated case-intuition; it also has to be articulated in relation to a principle and vice versa. As Rawls mentions,

> Objections by way of counterexamples are to be made with care, since these may tell us only what we know already, namely that our theory is wrong somewhere. The important thing is to find out how often and how far it is wrong. All theories are presumably mistaken in places. The real question at any given time is which of the views already proposed is the best approximation overall.[243]

My hope throughout these pages is to have shown that reflective equilibrium ought to be the default method for ethical inquiry, and that alternatives to it have the burden of proof. My original motivation for defending this position came from reading Lazari-Radek and Singer's book, *The Point of View of the Universe*, and, later on, from looking at other works of Singer. I am sympathetic to most of Singer's practical conclusions on what we ought to do, but I found his views on moral epistemology to be questionable. It seemed to me that it was intellectually erroneous simply to dismiss the case-intuitions of opponents, merely because they conflicted with our purported principles. I actually think Singer at his best relies on reflective equilibrium, even if he says otherwise. As we said before, reflective equilibrium is not incompatible, in principle, with any account in normative ethics, including classical utilitarianism. Careful consideration of possible judgments can be hard and it might involve substantial effort to convince opponents, but it seems to me that it is the best we can do.

---

[242] The case of a doctor who kidnaps and takes the organs of an innocent pedestrian as the only means of saving five patients in need of a transplant.
[243] Rawls, *A Theory of Justice*, p. 45.

References

Anderson, Elizabeth. "The Social Epistemology of Morality." In *The Epistemic Life of Groups: Essays in the Epistemology of Collectives*. Oxford: Oxford University Press, 2016.

Andes, Peter. "Sidgwick's Dualism of Practical Reason, Evolutionary Debunking, and Moral Psychology," (unpublished manuscript).

André, Baumard N., and D. Sperber. "A Mutualistic Approach to Morality." *Behavioral and Bran Sciences* 36 (2013): 59–122.

Audi, Robert. *The Good in the Right: A Theory of Intuition and Intrinsic Value*. Princeton, NJ: Princeton University Press, 2004.

Bealer, George. "A Priori Knowledge and the Scope of Philosophy." *Philosophical Studies* 81 (1996): 121–42.

Berkey, Brian. "Climate Change, Moral Intuitions, and Moral Demandingness." *Philosophy and Public Issues* 4, no. 2 (2014): 157–89.

———. "The Demandingness of Morality: Toward a Reflective Equilibrium." *Philosophical Studies*, no. 173 (2016): 3015–35.

Bonjour, Laurence. *The Structure of Empirical Knowledge*. Cambridge, Massachusetts: Harvard University Press, 1985.

Brandt, R. *A Theory of the Good and the Right*. Oxford: Claredon Press, 1979.

Buckwalter, Wesley, and Stephen Stich. "Gender and Philosophical Intuition." In *Experimental Philosophy, Volume 2*, edited by Knobe Joshua and Nichols, 307–46. Oxford: Oxford University Press, 2014.

Cappelen, Herman. *Philosophy Without Intuitions*. Oxford: Oxford University Press, 2012.

Chalmers, David J. "Intuitions in Philosophy: A Minimal Defense." *Philosophical Studies* 171 (2014): 535–44.

Clarke, Stanley G. "Anti-Theory in Ethics." *American Philosophical Quarterly* 24, no. 3 (1987): 237–44.

Dancy, Jonathan. *Ethics Without Principles*. Oxford: Claredon Press, 2004.

———. "Moral Particularism." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2013., 2013. http://plato.stanford.edu/archives/fall2013/entries/moral-particularism/.

Daniels, Norman. *Justice and Justification: Reflective Equilibrium in Theory and Practice*. Cambridge: Cambridge University Press, 1996.

———. "Reflective Equilibrium." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2016. Metaphysics Research Lab, Stanford University, 2016. https://plato.stanford.edu/archives/win2016/entries/reflective-equilibrium/.

———. "Wide Reflective Equilibrium and Theory Acceptance in Ethics." *The Journal of Philosophy* 76, no. 5 (1979): 256–82.

Davidson, Donald. "A Coherence Theory of Knowledge and Truth." In *Truth and Interpretation*, edited by E. LePore, 307–19. Oxford: Blackwell, 1986.

DePaul, Michael R. "The Problem of the Criterion and Coherence Methods in Ethics." *Canadian Journal of Philosophy* 18, no. 1 (1988): 67–86.

Dworkin, Ronald. "In Praise of Theory." *Arizona State Law Journal* 353, no. 29 (1997).

Frykholm Erin. "A Humean Particularist Virtue Ethics." *Philosophical Studies* 172 (2015): 2171–91.

Goldman, Alvin. "What Is Justified Belief?" In *Reliabilism and Contemporary Epistemology*, 29–49. New York: Oxford University Press, 2012.

Goldman, Alvin, and Bob Beddor. "Reliabilist Epistemology." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2016. Metaphysics Research Lab, Stanford University, 2016. https://plato.stanford.edu/archives/win2016/entries/reliabilism/.

Goodman, Nelson. *Fact, Fiction and Forecast*. Cambridge, MA: Harvard University Press, 1955.

Graham, Jesse, Jonathan Haidt, Sena Koleva, Matt Motyl, Ravi Iyer, Sean P. Wojcik, and Peter H. Ditto. "Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism." *Advances in Experimental Social Psychology* 47 (2013): 55–130.

Gray, K., L. Young, and A. Waytz. "Mind Perception Is the Essence of Morality." *Psychological Inquiry* 23 (2012): 101–24.

Greene, Joshua. *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them.* New York: The Penguin Press, 2013.

———. "The Secret Joke of Kant's Soul." In *Moral Psychology, Vol. 3*, edited by Walter Sinnott-Armstrong. MIT Press, 2008.

Greene, Joshua D. “From Neural ‘Is’ to Moral ‘Ought’: What Are the Moral Implications of Neuroscientific Moral Psychology?” *Nature Reviews Neuroscience* 4 (2003): 847–50.

Greene, Joshua, R. Brian Sommerville, Leigh E. Nystrom, John M. Darley, and Jonathan D. Cohen. “An fMRI Investigation of Emotional Engagement in Moral Judgment.” *Science* 293, no. 5537 (2001): 2105–8.

Gupta, Anil. “Definitions.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2015. Metaphysics Research Lab, Stanford University, 2015. https://plato.stanford.edu/archives/sum2015/entries/definitions/.

Haidt, Jonathan. “The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment.” *Psychological Review* 108, no. 4 (2001): 814–34.

———. *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. New York: Pantheon Books, 2012.

Hare, R. M. *Essays on Religion and Education*. Oxford: Claredon Press, 1992.

Harris, Sam. *The Moral Landscape: How Science Can Determine Human Values*. New York: Free Press, 2010.

Hasan, Ali, and Richard Fumerton. “Foundationalist Theories of Epistemic Justification.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2016. Metaphysics Research Lab, Stanford University, 2016. https://plato.stanford.edu/archives/win2016/entries/justep-foundational/.

Haslanger, Sally. “Gender and Race: (What) Are They? (What) Do We Want Them to Be?” *Nous* 34, no. 1 (2000): 31–55.

Honderich, Ted. *The Oxford Companion to Philosophy*. Oxford: Oxford University Press, 2005.

Huemer, Michael. *Ethical Intuitionism*. Basingstoke [England]; New York: Palgrave Macmillan, 2005.

———. “In Defence of Repugnance.” *Mind* 117, no. 468 (October 2008).

———. “Probability and Coherence Justification.” *The Southern Journal of Philosophy* 35 (1997).

———. “Revisionary Intuitionism.” *Social Philosophy and Policy* 25, no. 1 (2008): 368–92.

Hurka, Thomas. *British Ethical Theorists from Sidgwick to Ewing*. Oxford: Oxford University Press, 2014.

Hylton, Peter. "Willard van Orman Quine." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2018. Metaphysics Research Lab, Stanford University, 2018. https://plato.stanford.edu/archives/spr2018/entries/quine/.

Jackson, Frank. "The Role of Conceptual Analysis." In *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Claredon Press, 1998.

Kagan, Shelly. "Thinking about Cases." In *Moral Knowledge*, edited by Ellen Frankel Paul, Fred. D. Miller, and Jeffrey Paul. Cambridge: Cambridge University Press, 2001.

———. "What's Wrong with Speciesism? (Society of Applied Philosophy Annual Lecture 2015)." *Journal of Applied Philosophy* 33, no. 1 (2016).

Kahneman, Daniel, David Schkade, and Cass R. Sunstein. "Shared Outrage and Erratic Awards: The Psychology of Punitive Damages." *Journal of Risk and Uncertainty* 16 (1998): 49–86.

Kamm, F. M. *Intricate Ethics: Rights, Responsibilities, and Permissible Harm*. Oxford: Oxford University Press, 2007.

Knobe, Joshua, and Shaun Nichols, eds. *Experimental Philosophy*. New York: Oxford University Press, 2008.

Kohlberg, L. "From Is to Ought: How to Commit the Naturalistic Fallacy and Get Away with It in the Study of Moral Development." In *Psychology and Genetic Epistemology*, edited by T. Mischel, 151–235. New York: Academic Press, 1971.

Lazari-Radek, Katarzyna de, and Peter Singer. *The Point of View of the Universe: Sidgwick and Contemporary Ethics*. Oxford: Oxford University Press, 2014.

Lazari-Radek, Katarzyna, and Peter Singer. *Utilitarianism: A Very Short Introduction*. Oxford: Oxford University Press, 2017.

Lewis, C. I. *An Analysis of Knowledge and Valuation*. La Salle: Open Court, 1962.

Liao, S., A. Wiegmann, J. Alexander, and G. Vong. "Putting the Trolley in Order: Experimental Philosophy and the Loop Case." *Philosophical Psychology* 25 (2012): 661–71.

Mason, Kelby. "Moral Psychology and Moral Intuition: A Pox On All Your Houses." *Australasian Journal of Philosophy* 89, no. 3 (2011): 441–58.

Moll, Jorge, Paul J. Eslinger, and Ricardo de Oliveira-Souza. "Frontopolar and Anterior Temporal Cortex Activation in a Moral Judgment Task: Preliminary Functional MRI Results in Normal Subjects." *Arq. Neuropsiquiatr* 59, no. 3 (2001): 657–64.

Moll, Jorge, Ricardo de Oliveira-Souza, Ivanei E. Bramati, and Jordan Grafman. "Functional Networks in Emotional Moral and Nonmoral Social Judgments." *NeuroImage* 16, no. 3 (2002): 696–703.

Moore, G. E. *Principia Ethica*. Cambridge: Cambridge University Press, 1903.

Musschenga, Albert W. "Empirical Ethics, Context-Sensitivity, and Contextualism." *Journal of Medicine and Philosophy* 30 (2005): 467–90.

Nado, Jennifer. "The Intuition Deniers." *Philosophical Studies* 173, no. 3 (2016): 781–800.

———. "Why Intuition?" *Philosophy and Phenomenological Research* 86 (2011): 15–41.

Nozick, Robert. *Anarchy, State, and Utopia*. New York: Basic Books, 1974.

Nye, Howard. "Directly Plausible Principles." In *The Palgrave Handbook of Philosophical Methods*, edited by Chris Daly. UK: Palgrave Macmillan, 2015.

Olsson, Erik. "Coherentist Theories of Epistemic Justification." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2014. Metaphysics Research Lab, Stanford University, 2014. https://plato.stanford.edu/archives/spr2014/entries/justep-coherence/.

Ord, Toby. "Overpopulation or Underpopulation." In *Is the Planet Full?*, 46–60. Oxford: Oxford University Press, 2014.

Parfit, Derek. *Reasons and Persons*. Oxford: Claredon Press, 1984.

Pust, Joel. "Intuition." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2017. Metaphysics Research Lab, Stanford University, 2017. https://plato.stanford.edu/archives/sum2017/entries/intuition/.

Rawls, John. *A Theory of Justice*. Cambridge, MA: Harvard University Press, 1971.

———. "Outline of a Decision Procedure for Ethics." *The Philosophical Review* 60, no. 2 (1951): 177–97.

———. "The Independence of Moral Theory." *Proceedings and Addresses of the American Philosophical Association* 48 (1975 1974): 5–22.

Ross, W. D. *The Right and the Good*. New York: Oxford University Press, 2002.

Sandberg, Joakin, and Niklas Juth. "Ethics and Intuitions: A Reply to Singer." *The Journal of Ethics* 15 (2011): 209–26.

Schnall, Simone, Jonathan Haidt, Gerald L. Clore, and Alexander H. Jordan. "Disgust as Embodied Moral Judgment." *Personality and Social Psychology Bulletin* 34, no. 8 (2008): 1096–1109.

Shoemaker, David W. "Utilitarianism and Personal Identity." *The Journal of Value Inquiry* 33 (1999): 183–99.

Sidgwick, Henry. *The Methods of Ethics*. 7th ed. Palgrave Macmillan, 1901.

———. "The Theory of Evolution in Its Application to Practice." *Mind* 1, no. 1 (1876): 52–67.

Singer, Peter. "A Response." In *Singer and His Critics*, 269–335. Oxford, UK; Malden, Mass.: Blackwell Publishers, 1999.

———. *Animal Liberation*. New York: Harper Colling Publishers, 2002.

———. "Ethics and Intuitions." *The Journal of Ethics* 1, no. 9 (2005): 331–52.

———. "Famine, Affluence, and Morality." *Philosophy & Public Affairs* 1, no. 3 (1972): 229–43.

———. *Practical Ethics*. New York: Cambridge University Press, 1993.

———. "Sidgwick and Reflective Equilibrium." *The Monist* 58, no. 3 (1974): 490–517.

Small, D. A., and G. Lowenstein. "Helping a Victim or Helping the Victim." *Journal of Risk and Uncertainty* 26 (2003): 5–16.

Sosa, Ernest. "Minimal Intuition." In *Rethinking Intuition*, edited by M. DePaul and W. Ramsey, 201–40. Lanham, MD: Roman and Littlefield, 1998.

Southan, Rhys. "Peter Singer, R.M. Hare, and the Trouble With Logical Consistency." *Essays in Philosophy* 18, no. 1 (2017).

Stratton-Lake, Philip. "Intuitionism in Ethics." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2016. Metaphysics Research Lab, Stanford University, 2016. https://plato.stanford.edu/archives/win2016/entries/intuitionism-ethics/.

Street, Sharon. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127 (2006): 109–66.

Tännsjö, T. "Why We Ought to Accept the Repugnant Conclusion." *Utilitas* 149, no. 3 (2002): 339–59.

Temkin, L. “Intransitivity and the Mere Addition Paradox.” *Philosophy and Public Affairs* 16 (1987): 138–87.

Temkin, L. S. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. New York: Oxford University Press, 2012.

Thomas, Nigel J.T. “Mental Imagery.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2018. Metaphysics Research Lab, Stanford University, 2018. https://plato.stanford.edu/archives/spr2018/entries/mental-imagery/.

Thomson, Judith Jarvis. “A Defense of Abortion.” *Philosophy and Public Affairs* 1, no. 1 (Autumn 1971): 47–66.

———. “The Trolley Problem.” *The Yale Law Journal* 94, no. 6 (May 1985): 1395–1415.

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

Tversky, A., and D. Kahneman. “The Framing of Decisions and the Psychology of Choice.” In *Environmental Impact Assessment, Technology Assessment, and Risk Analysis*. Berlin: Springer-Verlag, 1985.

Unger, Peter. *Living High & Letting Die: Our Illusion of Innocence*. Oxford: Oxford University Press, 1996.

Walden, Kenneth. “In Defense of Reflective Equilibrium.” *Philosophical Studies* 2, no. 166 (2013): 243–56.

Ward, Andrew C. “The Value of Genetic Fallacies.” *Informal Logic* 30, no. 1 (2010): 1–33.

Weinberg, Jonathan M., Shaun Nichols, and Stephen Stitch. “Normativity and Epistemic Intuitions.” *Philosophical Topics* 29, no. 1 (Spring and Fall 2001).

Williamson, Timothy. *The Philosophy of Philosophy*. Oxford: Oxford University Press, 2007.