2014-08-07

Moral Knowledge Without Justification? A Critical Discussion of Intuitionist Moral Epistemology

Philipp Schwind

University of Miami, philipp.schwind@gmail.com

Follow this and additional works at: https://scholarlyrepository.miami.edu/oa_dissertations

UNIVERSITY OF MIAMI

MORAL KNOWLEDGE WITHOUT JUSTIFICATION? A CRITICAL DISCUSSION
OF INTUITIONIST MORAL EPISTEMOLOGY

By

Philipp Schwind

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

August 2014

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

MORAL KNOWLEDGE WITHOUT JUSTIFICATION? A CRITICAL DISCUSSION
OF INTUITIONIST MORAL EPISTEMOLOGY

Philipp Schwind

Approved:

_____
James Nickel, Ph.D.
Professor of Law and Philosophy

_____
Brad Cokelet, Ph.D.
Assistant Professor of
Philosophy

_____
Otavio Bueno, Ph.D.
Professor of Philosophy

_____
M. Brian Blake, Ph.D.
Dean of the Graduate School

_____
Robert Audi, Ph.D.
John A. O'Brien Professor of Philosophy
University of Notre Dame

SCHWIND, PHILIPP   (Ph.D., Philosophy)

<u>Moral Knowledge Without Justification? A Critical</u>   (August 2014)
<u>Discussion of Intuitionist Moral Epistemology</u>

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Professor James Nickel.
No. of pages in text. (237)

In this dissertation I discuss the epistemology of ethical intuitionism, in particular the claim that mature moral agents possess self-evident moral knowledge. Traditional intuitionists such as W.D. Ross have claimed that by reflection, we can acquire knowledge of our basic moral duties such as the duty of veracity or benevolence. Recent defenders of intuitionism such as Robert Audi have further developed this theory and argued that adequate understanding can be sufficient for moral knowledge. I criticize this view and argue that such accounts fail to make a convincing case for a foundationalist moral epistemology. Instead, I propose to separate the question of how we acquire moral knowledge from an account that justifies moral beliefs. In response to the first issue, I draw an analogy between our moral intuitions and chosmkian linguistics; in both areas, I argue, human beings possess a universal, unconscious and (partly) inaccessible system of rules that explains how we come to learn language and to make moral judgments. In regards to the justificatory issue, I address recent evolutionary debunking arguments designed to undermine the claim that our moral judgments track stance-independent truths. I try to show that this conclusion only follows under the assumption of an instrumentalist interpretation of moral reasoning which the intuitionist is not forced to accept.

Table of Contents

List of Tables

Chapter 1: Introduction

This dissertation is about the epistemology of ethical intuitionism. The guiding

question is whether and how we can acquire moral knowledge of fundamental moral

principles. Before turning to that question, I will introduce the core elements of ethical

intuitionism, explain what makes the theory attractive, and briefly summarize the main

arguments of my dissertation.

1.1 The core beliefs of ethical intuitionism

A central motivation behind intuitionism is the belief that an account of our moral

duties should be informed by the deeply held intuitions of mature agents. In order to

identify those principles, intuitionists like W.D. Ross believe that philosophers do not

need to devise ingenious procedures or invent new rules; rather, they should focus on the

moral principles that experienced people accept as common sense beliefs. As a result of

this method, Ross argues that we come to accept a plurality of irreducible moral duties.

Some of these appeal to the consequences of our actions while others derive from the

special relationship in which we stand to other people: the duties to keep promises, to pay

reparation for our wrongdoings, to show gratitude to those who have benefited us, to

bring about a distribution of benefits in proportion to merit, to benefit others, to improve

our own condition, and not to harm others.[1] According to Rossian intuitionists, it is

impossible to reduce this list to a smaller set of principles. Ross' list allows us to derive

---

[1] See Ross (1930), 21. This list is, as Ross stresses, only provisional (23). Later in *The Right and the Good*, he even suggests subsuming duties of justice, beneficence and self-improvement "under the general principle that we should produce as much good as possible" (27).

our overall moral obligations from one or more items on the list.[2] The duty to obey the laws, for instance, stems from the duty of gratitude, the obligation to uphold an implicit promise made to one's country and from the duty of beneficence.[3]

To be more precise in the characterization of the fundamental morally relevant consideration, Ross assigns them the status of '*prima facie* duties'.[4] The term, he admits, is misleading: neither does a *prima facie* duty necessarily express an actual obligation, nor is it, as the standard usage of the term "*prima facie*" suggests, based on a first impression that is conditional upon closer inspection. Instead, a *prima facie* duty describes a characteristic of an act (such as its being a lie, benefiting somebody else etc.) which would, if it were the only morally relevant feature present, result in an all-things-considered duty. If several *prima facie* duties are present in a particular case, it is nevertheless does not follow that we have several obligations at the same time. In such a case, there is only one "duty proper" which "arises from [the] *whole* nature" of the act.[5] Importantly, even though the *valence* of *prima facie* duties (the fact that they count in favor or against performing an action) remains stable across situations, their *strength* (their tendency to outweigh competing *prima facie* duties) is context-dependent and escapes codification. No algorithm determining the relative strength of *prima facie* duties

---

[2] See Ross (1930), 27.
[3] See Ross (1930), 27f.
[4] See Ross (1930), 19f. As Stratton-Lake (2000), ch. 5 discusses, there is more than one way in which Ross defines the notion of *prima facie* duty.
[5] Ross (1930), 20. It would be incorrect to say that the "duty proper" results from a judgment about which *prima facie* duty is the weightiest consideration. As a moral realist, Ross does not believe that what is right is determined by our judgment. Our moral duty reflects a mind-independent fact. Only in order to *recognize* what we should do, we have to exercise judgment.

can do justice to the complexities of moral life. A *prima facie* duty that trumps certain competing considerations in one situation might be irrelevant in another.[6]

Ross' normative view stands between two competing ethical theories. On the one side, consequentialists believe that our moral obligations are solely determined by the outcome of our actions.[7] While Ross agrees that we have a *prima facie* duty to benefit others, he denies that this is the only criterion by which our actions are to be judged. On his plural account of moral duties, what we should do also depends on factors that are independent of the results of our actions. Those considerations are grounded in the rights that other people hold against us in virtue of the fact that we stand in a special relation with them (e.g. if we make a promise, the promisee gains a right to the fulfillment of our promise).[8] Even though a violation of *prima facie* duties that fall into the deontic category often also causes bad consequences (for example, breaking a promise usually undermines trust and violates expectations), Ross argues that it would contradict common sense to reduce them to consequentialist claims. This locates him in the deontic camp, but the value an action creates can also be a genuine source of obligation. This puts him at odds with pure forms of deontology.

This normative account is only the starting point for a comprehensive theory of ethical intuitionism and invites further questions: Why is it that common sense serves as a guide to the truth? If we possess moral knowledge, by what process do we gain it? How shall we interpret moral discourse? Do moral facts exist? What is their nature, and how

---

[6] See Ross (1930), 23f.
[7] For present purposes, I ignore the various forms of act- and rule-consequentialism.
[8] See Ross (1930), 34-36. Some obligations we hold towards other people – e.g. the duty to refrain from harming them – are not grounded in a special relationship, but are grounded in general rights everybody possesses.

do they relate to non-moral facts? Even though there is no formally accepted set of

beliefs that all intuitionists would subscribe to, there are significant commonalities

between most of them. I will again stick with Ross' view, noting that even though he

reflects the position of many of his fellow intuitionists, his account is not the only one

possible.[9]

Since Rossian intuitionists accept common sense as a guide to truth in regards to the

content of moral principles, it is only natural for them to take our understanding of moral

discourse also at face value and to interpret moral utterances as factive statements with

truth-values.[10] According to this cognitivist view, moral utterances refer to external

facts.[11] This claim does however not imply that moral facts exist. For this to be the case,

moral realism, the view that there are stance-independent moral facts, would also have to

be correct. Another option compatible with cognitivism is moral anti-realism, the denial

of moral realism. This latter view paired with cognitivism would imply an error theory in

regards to moral discourse: our moral utterances are systematically false because they

presuppose a realm of facts that does not exist.[12] Rossian intuitionists reject the error

theory and adopt both cognitivism and moral realism.[13]

The kind of moral realism that intuitionists adopt is non-naturalism, the view that

moral facts are normative facts that cannot be discovered by or reduced to empirical

---

[9] Since the history of ethical intuitionism plays no central role in this dissertation (apart from an account of the reasons that led most philosophers to reject Ross view and their relation to the intuitionist revival since the 1980s), I will mostly ignore the rich and interesting views of Sidgwick, Prichard, Ewing and others.
[10] See Ross (1930), 85.
[11] Ross (1939), 278. Quoted from Stratton-Lake (2002), xiv.
[12] For such a view, see Mackie (1977).
[13] See Ross (1930), 81.

science.[14] But even if the cognitivist interpretation of moral discourse and non-naturalistic moral realism are both plausible theories, it is still possible that all of our moral beliefs are false. What we need is an indicator that mature agents have epistemic access to moral facts. The intuitionist' trust in common sense is only warranted if such an account can be given. In response to this challenge, Ross offers a foundationalist moral epistemology.[15] Our reflected beliefs concerning the contents of *prima facie* duties are self-evident, meaning that we are justified in holding them without the need of further justification.[16] We come to believe them by analogy to mathematical knowledge: after considering a number of instances, we recognize the underlying rule.[17] The resulting self-evident moral knowledge, Ross claims, is as trustworthy as scientific data.[18]

As we have seen, the main elements of ethical intuitionism are a pluralist theory of moral duty (including a characterization of those duties as *prima facie*), a cognitivist interpretation of moral discourse, and a defense of non-natural moral realism paired with foundationalism in epistemology. In order for moral intuitionism to be plausible overall, all of these claims require separate defense. The normative view is the least controversial. The idea that there is an irreducible plurality of moral claims enjoys wide support and underlies one of the main approaches in biomedical ethics.[19] Ethical cognitivism along with non-naturalist moral realism, even if far from uncontroversial, are defended by many

---

[14] Ross offers no direct support for this claim. Instead, he derives it from an argument that moral terms are simple and unanalyzable. Since there is no simple *natural* property that can be identified with moral properties, moral properties must be non-natural. (See Stratton-Lake (2002), xix) The question of whether that is a plausible argument would go beyond the scope of the present exposition of Ross' main claims. Stratton-Lake (2002), xvi-xxvi offers an illuminating discussion.

[15] Many, but not all intuitionists are also foundationalists. For an exception, see Dancy, (1993).

[16] See Ross (1930), 20f Fn1.

[17] See Ross (1930), 32f. For a more detailed account of Ross' moral epistemology, see chapter 2.

[18] See Ross (1930), 40f.

[19] See Beauchamp and Childress (2013).

contemporary philosophers.[20] The claim made by intuitionism that is most difficult to defend concerns moral epistemology. Even though Ross' foundationalism has skilled defenders who have developed and refined his arguments,[21] it is fair to say that philosophical common sense remains skeptical about the prospects of the doctrine. If intuitionists want to increase the attractiveness of their view, a convincing account of epistemological foundationalism might be their toughest challenge. What makes this task especially pressing is the central role that epistemology plays for intuitionism. Ross' moral epistemology fulfills three tasks:

1. His account of self-evidence offers a method for identifying the content of our well-considered judgments.

2. It provides a justification of those beliefs.

3. By giving credibility to common sense moral judgments, Ross' epistemology rules out an error theory. Below, I will lay out how I address this challenge. But first, I want to explain what makes intuitionism worth defending.

1.2 What makes ethical intuitionism an attractive theory

The case can be made that respect for common sense is a virtue intuitionism exhibits to a higher degree than competing ethical theories. This, however, is not the only advantage in its favor. On a deeper level, intuitionism reflects the different perspectives that human beings can adopt towards the world. As Thomas Nagel has argued, the

---

[20] See for example Foot (1958),
[21] Audi, Shafer-Landau and Stratton-Lake are the most prominent ones.

different kinds of obligations we have originate from different sources.[22] Human beings

view the world from individual, relational, and impersonal standpoints, and all of those

perspectives contribute reasons for what we should do. In our moral deliberations, we

have to integrate those different and potentially competing considerations. From the

individual perspective stem *prima facie* duties to improve ourselves. The relational

standpoint results in such duties as the obligation to fulfill promises, not to harm others,

and to show gratitude to our benefactors. The detached view gives rise to the duty of

beneficence.[23] To the same extent that those points of view are fundamentally tied to our

human nature and cannot be reduced to one another, we have reason to believe that such

a reduction cannot be achieved for the different kinds of obligations which result from

them as well.[24] This also explains why it is impossible to codify how the different

obligations interact with each other. For those rules to exist, we would have to put all

prima facie duties on one single scale. This can be done for obligations that originate

from the same source (for example when various rules are derived from a utilitarian

theory), but not for theories where duties from various standpoints converge.[25] In

concrete situations, it will therefore remain a matter of judgment what weight should be

assigned to different kinds of morally relevant considerations. This also gives additional

---

[22] See Nagel (1979), 128-141. In his article, Nagel uses obligations and values interchangeably. Since the present focus is on the right and not on the good, I will present his argument in terms of obligations or duties.

[23] Nagel's list of morally relevant considerations differs somewhat from Ross's. In one respect, it is more comprehensive and reveals a blind spot in Ross' account. Nagel lists obligations that arise from our commitments to other people ("specific obligations"). But next, Nagel adds a category of "rights that everyone has" which include "[r]ights to liberty of certain kinds, or to freedom from assault or coercion" (Nagel (1979), 129). In Ross' theory, there is no separate *prima facie* duty to respect those rights (which could be labeled as "rights to autonomy"). Often, they will be encompassed by the duty to beneficence, but Ross' arguments that the *prima facie* duty of fidelity cannot be reduced to beneficence would apply here as well.

[24] See Nagel (1979), 133.

[25] See Nagel (1979), 131.

support to Ross's claim that we should have more confidence in our abstract beliefs about the content of *prima facie* duties than in our judgments concerning concrete cases. It is more difficult to figure out which perspective provides the strongest reason in a concrete and complex situation than it is to recognize the different kinds of ethically relevant considerations attached to the various perspectives that we can adopt. While not threatening to undermine the irreducibility of Ross' moral principles, considering ethical intuitionism through the lens of Nagel's argument provides a deeper and more satisfactory explanation for his theory than the mere appeal to the benefits of sticking close to common sense morality. It also shows how various features of Rossian intuitionism are connected (the plurality of *prima facie* duties, their content, the lack of rank-order among them and the different degrees of certainty that attach to different levels of ethical knowledge).

1.3 Overview

The dissertation starts with a discussion and criticism of Ross' moral epistemology (chapter 2). This leads to a consideration of Robert Audi's view which in many respects clarifies and improves upon its predecessor. I also discuss some objections against Audi's understanding of self-evident moral knowledge (chapter 3). This motivates the search for an alternative justification of moral knowledge. Such a theory comes in two parts: first, the analogy between linguistic and moral knowledge helps to explain how moral agents form their moral beliefs (chapter 4). But are those beliefs also true? In response, I turn to one of the main criticisms of the possibility of such a view, evolutionary debunking arguments. I consider how intuitionists can reply to this challenge (chapter 5). Finally, I consider two separate issues that are also relevant for intuitionism:

1. How can the notion of *prima facie* duty be defended against moral particularists who claim that all reasons are context-dependent? (Appendix 1)

2. Lastly, I critically discuss Beauchamp and Childress' four-principle-account in bioethics which claims to be an application of intuitionist ethics. (Appendix 2)

Chapter 2: Ross distinguishes between two levels of moral deliberation, the abstract level of *prima facie* duties and the concrete level of duties all things considered. Establishing our duty all things considered involves determining the empirical details of the case and weighing the strength of the various morally relevant considerations that are present. Such a process is messy and we should not put much trust in the resulting ethical judgments. When thinking however in abstract terms about the kinds of duties that we have, our well-reflected beliefs can under certain conditions amount to knowledge. There are two methods that we can employ: conceptual reflection and intuitive induction. In conceptual reflection, careful consideration is given to candidate *prima facie* duties. After mature agents have repeatedly and over an extended period reflected on the meaning of ethically relevant claims, such as that lying counts against an action or the proposal that our duties ultimately reduce to a utilitarian rule, they form a considered judgment. According to Ross, this judgment is as trustworthy as our insight into a rule of mathematics. No additional support is needed in order to possess moral knowledge (Ross is ambiguous as to whether such support is possible). Second, intuitive induction derives knowledge of *prima facie* duties by analyzing similar cases. For example, if an agent has repeatedly judged actions to be wrong because they have involved a lie, this might lead to the recognition that the wrongness of lying always gives a *prima facie* reason not to perform an action.

Critics of Rossian intuitionism have raised two worries. First, they have argued that Ross' moral epistemology fails to account for moral disagreement. Given Ross' claim that self-evidence is a mark of true moral beliefs, it should not be possible for two agents to both engage in conceptual reflection or intuitive induction and arrive at contradictory conclusions. In fact, however, moral disagreement is common, raising worries whether Ross' epistemology is indeed a reliable guide to moral knowledge. Second, critics such as Mackie have argued that moral facts are unlike any other facts in virtue of their action-guidingness. This not only makes their existence difficult to justify, we would also have to possess a special epistemic faculty in order to detect them. They suggest that it is more plausible to assume that moral facts don't exist and that we consequently don't have epistemic access to them. There are good replies on behalf of intuitionism against both of those criticisms, but the fact that they appealed to many philosophers shows that Ross theory was not clear or explicit enough in his explanation of how we gain moral knowledge and of how these our moral competences can fail us. I then turn to some worries of my own against Ross' epistemology: his account of conceptual reflection claims that moral beliefs are vindicated if they pass a test of reflection and are not subject to defeaters (such as lack of reflection, maturity etc.). In order to have positive support for moral knowledge, we would however need a more detailed description of the test of reflection and an explanation of why our moral and conceptual competency is able to identify correct moral propositions. Ross theory of intuitive induction relies on an analogy between moral and mathematical knowledge, but I argue that the disanalogies between those two realms are too significant for the comparison to be helpful. Even if neither of the arguments against Ross amounts to a refutation of intuitionism, the

problems they raise can only be addressed with a more fully worked-out moral epistemology.

Chapter 3: Robert Audi, whose theory is the focus of the third chapter, presents such a theory. His view, moderate intuitionism, aims to achieve two things: (1) showing that many implausible claims that philosophers thought were implied by intuitionism are not essential to the theory and (2) presenting a worked out conception of self-evident moral knowledge. I discuss his theory and consider whether Audi can respond to the challenge of moral disagreement. In his account, (roughly) if an agent adequately understands a self-evident proposition, he has justification for believing it. A number of things, Audi argues, follow from this, among them including: (1) it is possible to understand a self-evident proposition without believing it; (2) adequately understanding a self-evident proposition does not imply the second-order knowledge that the proposition is self-evident. I clarify and discuss those claims. For example, it is a problem if the gap between adequate understanding and belief is too wide, allowing rational people to possess and understand all the relevant information about moral principles but fail to have correct moral beliefs. Hence, it needs to be specified under what circumstances the gap can arise. Regarding the second claim, it can be objected that lack of awareness about the self-evident nature of a proposition can have practical implications; complete understanding of the proposition is therefore not possible without it. Audi's account specifies what follows from understanding a self-evident proposition (it justifies our belief in it); however, it does not define the concept.

In response to this task, Audi's writings contain two suggestions. First, self-evident propositions can be understood as expressing paradigmatic reasons. In analogy to object

perception, we recognize paradigmatic reasons only if we reflect on the notion of a moral reason from a certain perspective (e.g. given a concrete experience in which that paradigmatic reason is instantiated); this explains why not all paradigmatic reasons are entailed in the concept of a moral reason. According to Audi's second attempt to define self-evident propositions, understanding such propositions puts us in contact with abstract entities. While this theory offers a promising account, it is unclear how to flesh out the details of the view.

Overall, despite Audi's helpful distinctions and definitions, many philosophers will not be convinced by the epistemology of ethical intuitionism. Therefore, the theory would benefit from a further independent explanation regarding our knowledge of *prima facie* duties.

Chapter 4: For traditional intuitionist accounts, self-evidence provides both the content of and the justification for our moral knowledge. In order to discover fundamental moral principles, we have to attend to what we are most certain of. Further, adequate understanding of those self-evident propositions also justifies those same beliefs. Instead, given the difficulties with self-evidence based accounts, I suggest treating those two questions separately. We should investigate which principles underlie the well-considered moral judgments of mature agents and afterwards consider whether those judgments are justified. One advantage of this approach is that it turns an introspective first-person method into an empirically verifiable theory in moral psychology. I argue that such a descriptively account is something that Ross himself was attempting to achieve, even if at his time, the required tools weren't available. In order to assess whether Ross' list of *prima facie* duties matches with such an empirical account, I

first assess what predictions can be derived from his theory. This includes among others the claim that there is a plurality of first principles, the assumptions that some of those duties are consequentialist while others are deontological, the content of the duties, and the epistemic stand that people take towards them. As a theoretical framework for testing these hypotheses, I use John Rawls' idea that modern linguistics can serve as a model for understanding moral psychology. Noam Chomsky argued that a generative grammar underlies our linguistic competence. The same, Rawls claims, is the case for our knowledge of the principles that underlie our moral judgments. On this view there is a universal system of rules to which we have at best limited conscious access. Building on John Mikhail's work that develops Rawls' idea by presenting a large number of individuals with variations of trolley cases and by analyzing their responses, many of Ross' claims turn out to be accurate. Even though there are, as I try to show, (a) important disanalogies between morality and linguistics and (b) significant differences between Mikhail's reconstruction of the moral grammar that guides everyday decision making and Ross' theory (for example, Mikhail's theory rejects the idea that duties are *prima facie*), intuitionism is strengthened by those empirical findings that serve as arguments in favor of the claim that common sense morality can be described by a system of *prima facie* duties.

Chapter 5: Even if it can be shown that the assumptions Ross makes about common sense morality are descriptively adequate, this leaves open whether those beliefs correspond to stance-independent facts, as many intuitionists would like to argue. Instead of offering a direct response to this worry, I discuss a major obstacle which, if successful, would show that we have no reason to believe that our moral beliefs track the truth.

Debunking arguments start with a causal premise. According to this premise, a set of beliefs can be explained by a certain causal process. The second epistemic premise describes this process as not truth tracking. It is concluded that the set of beliefs in question are therefore unjustified. This does not make it impossible for our beliefs to match with stance-independent facts, but it would be a cosmic coincident if out of all the possible sets of beliefs, the non truth-tracking process would have steered us towards the correct one. I consider two causal factors that contribute to our moral beliefs and that satisfy the epistemic premise: psychological defeaters and the impact of natural selection. Walter Sinnott-Armstrong has argued that fundamental moral beliefs are systematically distorted by factors such as (among others) partiality, emotion, order and priming effects. Taken together, they cast doubt on the reliability of all moral judgments. In order to make the case that the intuitions that the ethical intuitionist is interested in fall into a subgroup of moral judgments that has a higher chance of being correct (because, the intuitionist might argue, they have been formed under conditions that are free of the distorting factors Sinnott-Armstrong lists), inferential justification is needed. But since it is a core belief of intuitionism that fundamental moral beliefs do not stand in need of inferential justification, this move is not available to intuitionists. Hence, we have to assume that our moral beliefs are distorted by psychological factors and are therefore unreliable. Against Sinnott-Armstrong's criticism, I argue that using descriptive claims which show that the moral judgments the intuitionist is interested in are more likely to be true than others does not undermine the foundationalist status of those beliefs.

I next consider evolutionary debunking arguments claiming that normative beliefs are (partly) shaped by evolutionary pressures. Those pressures have selected for survival and

reproductive fitness, but not for truth. Therefore, we have no reason to assume that moral beliefs are correct. Michael Huemer offers a reply on behalf of intuitionism: he attempts to single out intuitions that are immune from selective pressures; those beliefs can then be used to verify other moral intuitions. The problem with this argument is that the abstract principles that survive this process (e.g. "If x is better than y and y is better than z, then x is better than z") are not of much use since no moral beliefs can be derived or tested by them. I advance two worries of my own which might at least weaken the strength of debunking arguments. First, it might be less of an accident than debunkers such as Sharon Street assume if our moral beliefs turned out to be true. After all, even Street accepts that what is genetically inheritable are only basic evaluative tendencies, and there are far less of those than there are sets of full-fledged normative beliefs. As a consequence, it would be much less surprising if the basic evaluative tendencies we have acquired overlap with a mind-independent moral truth.

Further, if it can be established that evolutionary pressures have impacted only basic evaluative tendencies, there is room for other factors to play a role as well. One such factor that has not been given enough attention in the discussion about evolutionary debunking arguments is the role of moral reason. According to rationalist theories of moral reason, reflection is able to provide us with insight into the correct moral principles. This, I argue, would make debunking arguments less devastating than their defenders make them appear to be.

Chapter 2: Ross' Moral Epistemology and its Critics

2.1 Introductory Remarks

In this chapter, I will first present Ross' arguments for his epistemological views. In a second step, I consider the most influential and representative arguments that have been raised against his theory. Even if a number of those critics haven't paid close attention to Ross' own arguments, their worries reflect a general dissatisfaction with ethical intuitionism and help to understand why the theory fell into disrepute for much of the 20th century. An analysis of the reaction to Ross' views also provides an insight into the challenges that contemporary moral philosophers inclined towards ethical intuitionism have faced when they attempt to present a version of the theory that would be taken seriously even by critics. In the final section of this chapter, I provide arguments of my own against Ross' view.

2.2 Ross' moral epistemology

2.2.1 Levels of moral thinking and their epistemic status

In *The Right and the Good*, Ross distinguishes between two levels of moral thinking, one concerned with general moral principles and the other with our duty all things considered which applies to concrete circumstances. The abstract level contains *prima facie* duties, whereas the concrete level specifies our 'duty proper' or 'duty sans phrase'. The two levels differ in their epistemic status. We are able to

achieve knowledge at the level of *prima facie* duties, but Ross argues that our

insight into 'duty proper' is never more than "probable opinion."[26]

The knowledge achieved at the level of *prima facie* duty is based on

common-sense convictions: "[T]he moral convictions of thoughtful and well-

educated people are the data of ethics just as sense-perceptions are the data of a

natural science."[27] In taking common-sense as the foundation of moral knowledge,

Ross puts himself in opposition to moral philosophers such as utilitarians who

claim that we can derive our moral duties from one underlying principle:

> I would maintain, in fact, that we are apt to describe as 'what we
> think' about moral questions contains a considerable amount that we
> do not think but know, and that this forms the standard by reference
> to which the truth of any moral theory has to be tested, instead of
> having itself to be tested by inference to any theory.[28]

It is important to stress that the certainty that characterizes our moral knowledge

only extends to *prima facie* duties which specify that certain features such as

honoring fidelity or benefiting somebody count in favor, other features such as

causing harm count against an act; however, we cannot achieve the same

confidence in judgments that weigh the *strength* of our duties against each other.[29]

Ross offers two reasons for the claim that the epistemic status of our duties

all things considered is weaker than our knowledge about *prima facie* duties.[30]

First, when confronted with a situation in which several *prima facie* duties with an

opposite valence conflict, there are no general principles giving rules which cover

---

[26] Ross (1930), 31.
[27] Ross (1930), 40. See as well 23, 29 and 39f. See as well his (1927), 123.
[28] Ross (1930), 40.
[29] See Ross (1930), 30f. and 41f. In his (1927), see 123.
[30] See Ross (1930), 19 & 30 and his (1939), 190f.

the complex and context-dependent interplay between various *prima facie* duties; all we have are defeasible heuristics such as the claim that "a great deal of stringency belongs to the duties of 'perfect obligation' – the duties of keeping our promises, of repairing wrongs we have done, and of returning the equivalent of services we have received."[31] This lack of certainty makes the conclusion about our overall duty more prone to error. The common man is, Ross claims, well aware of the contrast between the self-evidence which we ascribe to our knowledge of *prima facie* duties and the "moral risk" we take in making overall moral judgments.[32] Second, even if it seems to us that only one *prima facie* duty bears on a situation, it is always possible that we might have overlooked the existence of countervailing considerations. For example, it is to be expected that every act has in the long run positive as well as negative consequences, making it unlikely that we will ever encounter a case that is free of conflicts between duties.[33]

Ross does not, however, believe that the lack of rules guiding our judgments about overall duty is a problem that his theory is facing alone; his strongest competitor, utilitarianism, does not offer a comprehensive rule system that helps to weigh different forms of good either.[34]

The claim that we can never have knowledge regarding our duties all things considered is a strong claim, and it makes only sense if we keep in mind Ross'

---

[31] Ross (1939), 41f.
[32] See Ross (1930), 30.
[33] See Ross (1930), 31 as well as his (1939), 172f. What gives plausibility to this claim is Ross' argument that only the consequences of an act, but not the motive determine its rightness. Furthermore, in order to fully assess an act, there is – according to Ross – no cut-off point in time beyond which distant consequences won't affect its ethical status.
[34] See Ross (1930), 23f.

understanding of what constitutes moral knowledge. Most contemporary

philosophers would adopt a weaker notion and consequently allow for moral

knowledge regarding all things considered even in the face of the uncertainties that

Ross mentions. Even if we accept that there are no principles determining how to

weigh *prima facie* duties against each other, in many cases there is little doubt

which moral consideration is the weightiest. For example, if I consider murdering

somebody else to steal their money in order to buy myself a new car, it would be

quite an exaggeration to claim that due to the lack of principles assigning a fixed

weight to the *prima facie* duties that are at stake (non-maleficence versus

beneficence), I do not know that my intention is immoral. Further, the requirement

that we need to take into consideration the consequences our acts will have in the

indefinite future in order to know its moral status is not widely shared and would

for most philosophers not constitute an objection again the knowledge of all things

considered duties.

The fact that Ross never further develops his account of how we arrive at

judgments regarding our overall duty suggests that he does not believe that

additional ethical inquiry can illuminate the process of solving conflicts between

various duties. Quoting Aristotle's dictum that "[t]he decision rests with

perception," he argues that our practical judgment is "the only guide we have to

our [overall] duty."[35] Instead, Ross focuses his attention on explaining how we

achieve and justify our knowledge of *prima facie* duties.

---

[35] Ross (1930), 42. Relying on Aristotle's conception of moral judgment offers Ross a rich account of
moral deliberation, but he does not further develop his reference to Aristotle.

But before turning to the epistemology of *prima facie* duties, it is important to note Ross' view of moral deliberation that is implicit in his remarks above. Although there are no exact rules determining how we should reason from our perception of a concrete situation to a judgment about our overall duty, it is clear that (a) we can have firm knowledge of what *prima facie* duties consist in (i.e., we can know the content of *prima facie* duties in general, but not how the various *prima facie* duties that bear on a situation are weighed against each other), (b) that one, but most likely more than one *prima facie* duty bears on what our duty proper in the situation is,[36] (c) that in concrete situations, we have to establish which *prima facie* duties play a role,[37] (d) that we have to weigh various *prima facie* duties against one another in order to reach a judgment about our overall duty, (e) that context, not rules, at least partly determines the strength of each *prima facie* duty.

2.2.2 How we gain general moral knowledge

How do we gain knowledge of *prima facie* duties? According to Ross, there are two ways: through reflection and induction. Both processes result in a state of mind where the *prima facie* duties are self-evident to us.

---

[36] See Ross (1930), 41.

[37] This is implicit in Ross (1930), 41f: "Every act therefore, viewed in some aspects, will be *prima facie* right, and viewed in others, *prima facie* wrong […]. For the estimation of the comparative stringency of these *prima facie* obligations no general rules can, so far as I can see, be laid down. We can only say that a great deal of stringency belongs to the duties of 'perfect obligation' – the duties of keeping our promises, of repairing wrongs we have done, and of returning the equivalent of services we have received. For the rest, the decision rests with perception" In order to exercise practical judgment, we have to be aware of the *prima facie* duties that are at stake.

2.2.2.1 Conceptual Reflection

By reflection, Ross means the attentive consideration of a proposition:

> That an act, *qua* fulfilling a promise […] is *prima facie* right, is self-evident; not in the sense that it is evident from the beginning of our lives, or as soon as we attend to the proposition for the first time, but in the sense that when we have reached sufficient mental maturity and have given sufficient attention to the proposition it is evident without any need of proof, or evidence beyond itself. It is self-evident just as a mathematical axiom, or the validity of a form of inference, is evident. The moral order expressed in these propositions is just as much part of the fundamental nature of the universe (and, we may add, of any possible universe in which there were moral agents at all) as is the spatial or numerical structure expressed in the axioms of geometry or arithmetic. In our confidence that these propositions are true there is involved the same trust in our reason that is involved in our confidence in mathematics; and we should have no justification for trusting it in the latter sphere and distrusting it in the former. In both cases we are dealing with propositions that cannot be proved, but that just as certainly need no proof.[38]

Ross makes a number of claims in this rich passage. His aim is to explain how a process that can be called 'conceptual reflection' leads to self-evidence knowledge of *prima facie* duties. Two conditions qualify the appropriate kind of reflection: first, the agent needs to be in a state of mental maturity. This excludes children, those who are inexperienced in moral matters and psychopaths from obtaining moral knowledge in this way. Second, the proposition to be examined has to be familiar to the agent in the sense that he or she has reflected on it (a) over an extended period of time ("not […] attended to the proposition for the first time") and (b) has thought about it with due care ("given sufficient attention to the proposition"). This implies that moral knowledge is for Ross not the result of an

---

[38] Ross (1930), 29f.; see as well 12f. and 24.

immediate judgment about a moral problem; it also implies that no defeaters such as emotional arousal or a bias distort the reflection.

Next, Ross introduces a comparison between moral and mathematical facts that serves two purposes. Concerning moral metaphysics, he claims that true moral statements correspond to objective non-naturalist facts as much as mathematical claims do. Second, moral knowledge of *prima facie* duties has the same degree of certainty that we assign to mathematical truths. Hence, moral knowledge of principles enjoys a maximally trustworthy epistemic status. As he later adds, self-evident knowledge of *prima facie* duties implies that we couldn't imagine it being false.[39]

What motivated Ross to stress so strongly the reliability of our knowledge gained through conceptual reflection? There are two reasons: First, in order to vindicate the claim that we have epistemic access to basic duties which cannot be defined in terms of anything else, Ross has to show that we have sufficiently strong justification for believing them; the justification cannot rest on any other more fundamental claims. Second, the context of his overall project needs to be considered. In *The Right and the Good*, Ross' main objective is to establish that utilitarianism violates common sense and to develop a competing theory that respects the opinions of "the common men." It is therefore of importance for him to establish that moral knowledge based on common-sense has a strong epistemic

---

[39] Ross (1930), 39f.: "[I]f we are told, for instance, that we should give up our view that there is a special obligatoriness attaching to the keeping of promises because it is self-evident that the only duty is to produce as much good as possible, we have to ask ourselves whether we really, when we reflect, are convinced that this is self-evident, and whether we really can get rid of our view that promise-keeping has a bindingness independent of productiveness of maximum good."

weight that cannot be easily overruled by theoretical arguments like the utilitarian

principle of maximizing pleasure and minimizing pain. Otherwise, Ross could not

explain why the fact that utilitarianism has consequences that – as he believes –

contradict our considered opinions counts against it.

However, two limitations on the self-evidence of common-sense moral

convictions have to be noted. First, what appears to us as self-evident moral

knowledge can be defeated if it contradicts another moral belief which "stands

better the test of reflection."[40] Second, not all moral convictions are candidates for

knowledge:

> The main moral convictions of the plain man seem to me to be, not
> opinions which it is for philosophy to prove or disprove, but
> knowledge from the start; and in my own case I seem to find little
> difficulty in distinguishing these essential convictions from other
> moral convictions which I also have, which are merely fallible
> opinions based on an imperfect study of the working for good or
> evil of certain institutions or types of actions.[41]

While some moral convictions turn out to be knowledge, others are merely

"fallible opinions." Nothing in the quote indicates that the distinction between

those moral convictions that can be considered as "essential convictions" from

those that are "merely fallible opinions" falls together with the distinction between

prima facie duties and duty proper.

It is ambiguous whether Ross allows self-evident moral truths to be

supported by further arguments that go beyond reflection. As quoted above, he

claims that moral propositions 'cannot be proved, but […] just as certainly needs

---

[40] Ross (1930), 41.
[41] Ross (1930), 21f. Fn. 1

no proof.'[42] On other occasions, he makes the more cautious claim that moral

propositions don't need further proof,[43] while in "The Basic of Objective

Judgments in Ethics," he explicitly allows for additional proof.[44]

For Ross, moral knowledge is not the result of an isolated reflection. We

should take into account what fellow moral inquirers have come to conclude over

time:

> The existing body of moral convictions of the best people is the
> cumulative product of the moral reflection of many generations,
> which has developed an extremely delicate power of appreciation
> of moral distinctions; and this the theorist cannot afford to treat
> with anything other than the greatest respect.[45]

2.2.2.2 Intuitive Induction

Apart from conceptual reflection, Ross offers a second account of how to

gain moral knowledge: through intuitive induction, we apprehend abstract truths

by considering a range of individual cases.[46] Human beings are, Ross argues, not

born with knowledge of moral facts; still less are we aware from birth that these

facts are self-evident. Instead, moral knowledge needs to be acquired. In order to

explain how this process works, he draws an analogy to our learning of

mathematical facts: We see that two and two items added to another result in four

items. After encountering various instances of this kind, we realize that we are

dealing with a mathematical law that holds independent of the items that are being

---

[42] Ross (1930), 30.
[43] See Ross (1930), 29.
[44] See Ross (1927), 121. For more discussion, see Stratton-Lake's (2002), xlix.
[45] Ross (1930), 41.
[46] Ross introduces the term in his (1939), 170. For his more extended discussion on which my
reconstruction is based, see his (1930), 32-34.

counted. Through concrete examples, we come to apprehend abstract facts. According to Ross, the same process is at work in regards to moral knowledge of *prima facie* duties. If we encounter various instances of promise-keeping and possess sufficient intellectual maturity, we conclude that we have a *prima facie* duty to always keep our promises; Ross stresses that this account is not contradicted by cases where the overall duty differs from a *prima facie* duty that is also present since we are aware that countervailing considerations can outweigh any other obligations that we might also have.

Ross notes that the above-mentioned analogy between mathematical and moral facts breaks down in regard to how properties of both kinds interact with each other.[47] While mathematical properties such as the fact that a certain triangle is isosceles depend on characteristics that can be isolated – in the present example, a triangle is isosceles in virtue of the fact that two of its angles are equal and it does not matter what other properties the triangle possesses – in moral cases our overall duties are determined by the interplay of many factors: "[an act's] rightness depends on its whole nature and not on any element of it."[48]

The notion of induction that Ross operates with contrasts with two modern understanding of inference, enumerative induction and abduction. According to the first, the likelihood that a statement is universally true increases with the number of observed instances and remains in principle falsifiable. Abduction, the second alternative form of inference, compares various possible explanations of a fact and adopts the most plausible one. Both enumerative induction and abduction result in

---

[47] Ross (1930), 33f.
[48] Ross (1930), 33.

conclusions that are likely or plausible, but not guaranteed to be true. In his moral epistemology, Ross takes on neither of those methods. Instead, in his book on Aristotle Ross describes the form of induction he uses as "the process whereby after experience of a certain number of particular instances the mind grasps a universal truth which then and afterwards is seen to be self-evident."[49] This explains the analogy to mathematical discovery where our confidence in the truth of "2+2=4" can – once the underlying regularity has been firmly grasped – not be increased by encountering further cases which instantiate this fact.

In intuitive induction, Ross relies not only on self-evident knowledge regarding *general principles* of *prima facie* duty, but also concerning *prima facie* rightness of *particular acts*.[50] He is committed to this view in order to avoid a circularity in his account of intuitive induction: if our self-evident knowledge that e.g. lying in this particular instance is *prima facie* wrong would have been derived from our general knowledge about the *prima facie* wrongness of lying, our knowledge in the concrete case could not be used to support our knowledge in the abstract case since it would depend on it in a circular way. Instead, Ross claims that we have an independent and direct insight (an action "presents itself *prima facie* as right in virtue of some character it possesses")[51] into the nature of an act which leads us to recognize that it is *prima facie* wrong given the fact that it entails

---

[49] Ross (1923), 223.
[50] Ross (1930), 33. The second case refers to our knowledge that in a particular situation, a certain *prima facie* duty is present, but not how this duty has to be weighed against other principles or whether it is our duty all things considered. As Ross has argued above, these beliefs are necessarily less than certain.
[51] Ross (1939), 173.

telling a lie.[52] Here, he is not claiming that we can know what our duty is all things considered (this would contradict his earlier claim that such knowledge is impossible), but merely that if an act instantiates e.g. lying, we realize that this counts against doing it.

2.2.3 An illustration of Ross' moral epistemology in The Right and the Good

After introducing conceptual reflection and induction as methods for gaining moral knowledge, Ross illustrates his moral epistemology by discussing how the concepts of "right" and "optimific" (i.e. being productive of the best possible consequences) are related to each other.[53] Is whatever is *prima facie* right as well optimific, as G.E. Moore argues?[54] Ross attempts to refute Moore by using three epistemological approaches: immediate apprehension, induction and reflection (or, as he calls it here, "deduction").[55] The merits of the arguments themselves are not the focus here; I will only consider the way the different epistemological methods are presented.

First, Ross discusses whether "immediate apprehension" can cast light on the relationship between the two concepts in question.[56] Although he does not explicitly establish the connection, immediate apprehensions seem to be the same judgments as those concerning particular situations which Ross refers to in his discussion of *intuitive induction*. In both cases, the judgments are concrete, self-

---

[52] "[W]e *see* the *prima facie* rightness of an act which would be the fulfillment of a *particular* promise" Ross (1930), 33 (my emphasis). Ross (1939), 169f. also comments on the difference between the general knowledge of *prima facie* duties and our independent moral knowledge in concrete situations.
[53] See Ross (1930), 34-39.
[54] See Moore (1912), 181 (from Ross (1930), 34).
[55] Ross (1930), 39.
[56] See Ross (1930), 34-36.

evident and used in order to provide evidence for a more general belief. In the

present context, the appeal to immediate apprehension does not serve to establish a

positive thesis; instead, it has the negative function of showing that there are

scenarios where it is evident that the right and the optimific fall apart. If we

consider cases where an agent has two options, one being the fulfillment of a

promise which produces 1000 units of good and the other a different act which

produces 1001 units of good, we hold that the fact that the second act produces a

slight surplus of good does not make the second rather than the first act our duty.

This is obvious to us even if we are not sure what exactly the concept of promising

involves.[57] Discussion of these and similar cases lead Ross to the conclusion that

there is "no self-evident connexion between the attributes 'right' and

'optimific'."[58]

Second, Ross turns to induction.[59] In order to establish through induction

that we know the coextensiveness of the right and the optimific, two conditions

have to be fulfilled. First, an extensive number of acts needs to be examined;

second, our judgments about these cases have to take account of their

consequences "far into the future." Given these demanding requirements, there is

little hope that we can find out whether the assertion in question is self-evident. At

most, we can claim a preponderance of cases where the right is optimific. In sum,

an investigation which involves "enormous complexity" and offers merely

---

[57] See Ross (1930), 35.
[58] Ross (1930), 35. As an aside, this confirms Ross's standpoint that what is self-evident must not be obvious. At the opposite, subtle reflections might be necessary to discover self-evident moral truths.
[59] See Ross (1930), 36f.

"inevitable inconclusiveness […] is worth no one's while."[60] It is not clear whether the method used here is the same as the one that Ross has introduced above under the label of "intuitive induction." In discussing the relation between the right and the optimific, he makes it clear that even the consideration of large number of cases "[would fall] far short of proving the constant connexion of the two attributes."[61] Such a procedure would be employed in enumerative induction, but not with the form of intuitive induction that Ross seemed to rely on in when he introduced the method (see last section). There, in analogy to mathematical knowledge a small number of instances is sufficient to draw our attention to the nature of the proposition in question and to give us self-evident knowledge that does not require further empirical investigation. The present example suggests that Ross uses, other than his introduction of intuitive induction suggests, both the classical and the enumerative form of induction.

Finally, Ross turns to a third method.[62] Although he again leaves it unclear how it relates to the methods introduced earlier, at the end of the passage he summarizes his discussion of Moore's claim by stating that "we do not know either by intuition, by deduction, or by induction that [the right and the optimific] coincide in their application."[63] Since the first method examined is clearly based on intuition and the second on induction, the third method has therefore to be on deduction. In his discussion of epistemological approaches, Ross does not explicitly mention deduction, but the content of the passage in question suggests

---

[60] Ross (1930), 37.
[61] Ross (1939), 36.
[62] See Ross (1930), 37-39.
[63] Ross (1930), 39.

that deduction is what I have called above "conceptual reflection." In it, Ross argues that 'reflection on what a promise is' implies that it is not identical to what is optimific.[64] Whereas in his methodological discussion mentioned above, Ross focused on the conditions that have to be fulfilled for a successful conceptual reflection, here we learn how the method works in practice: disagreements about the content of our general moral principles can be resolved by first figuring out what meaning of moral terms certain theories imply and then by reflecting on whether this meaning accords with common-sense. In the case at hand, utilitarianism claims that the moral value of keeping a promise is a function of its consequences; we have an obligation to keep our word as long as it leads to the best overall outcome (taking into account the damage a broken word causes to the institution of promising). Ross believes that this contradicts our understanding of the concept of promising: "Do we really think that the production of the slightest balance of good, no matter who will enjoy it, by the breach of a promise frees us from the obligation to keep our promise?"[65] Rather, to make a promise means to put oneself in an irreducible relation to another person. This illustrates how Ross believes that disputes about the meaning of moral terms can be decided by reflecting on 'what we really think'. It also shows that he believes that those considerations trump over arguments (in this case, utilitarianism) that do not have (according to Ross) the benefit of being in line with common sense.

---

[64] Ross (1930), 37: "It is plain, I think, that in our normal thought we consider that the fact that we have made a promise is in itself sufficient to create a duty of keeping it."
[65] Ross (1930), 38.

2.2.4 The role of general moral knowledge in The Foundations of Ethics

While his discussion of moral epistemology in *The Right and the Good*
focuses mostly on self-evident knowledge of abstract principles of *prima facie*
duties, his second book on moral philosophy, the *Foundations of Ethics*, puts a
different emphasis in regards to moral knowledge. There, Ross argues that abstract
principles are in most cases not necessary in order to reach a moral verdict in
concrete cases.[66] Although it seems to us that "our perception of particular duties is
always an act of inference, in which the major premise is some general moral
principle",[67] agents who don't possess those concepts (for example, as Ross
argues, primitive human beings from an earlier epoch),[68] are able to make moral
judgments. This suggests a different order or recognition: knowledge of what
duties are relevant in a specific situation precedes knowledge of duties in a general
form (e.g. Ross list of *prima facie* duties). We gain awareness that certain duties
bear on a context "by direct insight" and not by reflection.[69] Ross allows for two
narrow exceptions. First, we might in our early stages of moral development
accept a moral principle on authority – e.g. that lying is wrong –, but once we have
matured, we recognize on our own the moral importance of honesty.[70] Second,
various concrete considerations might result in our believing a general principle,

---

[66] In the *Foundations of Ethics*, Ross also deals at length with the uncertainty of moral judgments in
particular cases [173-190], but his discussion of the topic only develops further his thoughts from *The Right
and the Good* [30-32]. I will therefore not discuss it here.
[67] Ross (1939), 169.
[68] See Ross (1939), 169.
[69] Ross (1939), 171.
[70] Ross (1930), 171: "In most people's lives there is a stage at which they accept some moral principle on
authority before they have really come to recognize its truth for themselves."

e.g. that indiscriminate charity is wrong.[71] In a specific situation, we don't go again

through the complicated process of tracing down all the concrete facts, but rely on

the general principle. Ross takes, however, pain to emphasize that these scenarios

(where we rely on a general principle in order to reach a practical verdict) do not

correspond to normal cases of moral cognition.[72]

Even though in the *Foundations of Ethics,* Ross does not explicitly deny his

earlier claims, it is apparent that his position has developed. In *The Right and the*

Good, no discussion regarding the significance of general moral knowledge can be

found. Implicitly, however, he must assume that the acquisition of knowledge of

prima facie duties is important for us. After all, he presents intuitive induction and

conceptual reflection as methods which common people actually engage in.[73] This

assumption is absent from his discussion in *The Foundations of Ethics*. There,

knowledge of general principles is considered superfluous for the exercise of our

moral judgment, even though many of us are, in principle, capable of

understanding them 'when they are pointed out to us.'[74]

Ross' earlier position certainly provides more resources for rendering

plausible his foundationalist moral epistemology. There, we have a positive

account of how we come to have knowledge of our duties. By characterizing in the

---

[71] See Ross (1939), 172f.

[72] "Our insight into the basic principles of morality is not of this order. When we consider a particular act as a lie, or as the breaking of a promise, or as a gratuitous infliction of pain, we do not need to, and do not, fall back on a remembered general principle; we see the individual act to be by its very nature wrong." Ross (1939), 173. Here, Ross puts more trust into our ability to come to correct conclusions regarding our duty all things considered than he does in his (1930). While there, his intuitionism is directed towards general rules, here it seems to extend towards acts.

[73] See e.g. on (1930), 32. It would also have been an option for him to deny that pursuing knowledge of *prima facie* duties has direct ethical relevance. After all, the pursuit of most philosophical questions is practically useless.

[74] Ross (1939), 172.

Foundations of Ethics our moral faculties as providing "direct insight"[75] into what

prima facie duties demand from us in a given situation and by minimizing the

significance of general moral principles, Ross does not effectively respond to his

critics who want a clearer account of moral knowledge; at the opposite, it offers

even less of an explanation than his earlier account.

2.3 Criticisms of Ross by his contemporaries

In the decades following the publication of Ross's work, ethical

intuitionism, a view that was until World War II one of the most widely held

theories and defended in one form or another by the major moral philosophers of

the time, faced almost unanimous rejection. Often, the criticisms were crude and

did not do justice to the subtleties of the arguments they attacked. One explanation

for this change of climate is logical positivism which brought into disrepute the

idea of irreducible non-natural moral facts, but another notion that attracted much

criticism was the concept of self-evident moral knowledge.[76] In order to

understand the reasons behind the rejection of intuitionism, it is helpful to consider

a number of those criticisms in more detail. Here, I limit myself to arguments

against intuitionist moral epistemology. With the purpose of presenting widely

shared views on the topic, I will consider textbooks on ethics from the time as well

as more specialized literature.[77] It is surprising that the criticisms are – although

---

[75] Ross (1939), 171.
[76] See Brandt (1959), 190: "The most vulnerable point of the nonnaturalist doctrine is its epistemology." See also Frankena (1963), 105.
[77] Although I will sometimes show that the criticisms presented here misread Ross (if they read him at all), I will not try to defend him wherever possible; rather, I want to show which obstacles recent intuitionists faced when formulating their views.

they differ in detail – mostly similar and can be grouped under two headings: (a) given its epistemological commitments, intuitionism cannot account for moral disagreement and (b) intuitionism explains moral cognition by appeal to a mysterious faculty.[78]

2.3.1 Intuitionism and moral disagreement

On both concrete and fundamental levels, there is disagreement among mature moral agents about what our duty is. Many critics have taken this to be a major problem for ethical intuitionism. There are two main arguments that are often raised in this context.[79]

The first argument tries to show that the truth of intuitionism is incompatible with moral disagreement. Richard Brandt, for example, asks how different agents can have conflicting moral intuitions if those intuitions are supposed to be necessarily true.[80] Perhaps, some of them are based on "pseudo intuitions" that should be discarded. This leads to the second argument according to which intuitionists have no means of solving moral disagreements. If the intuitionist' only means in order to establish moral truth is an appeal to self-evidence, Brandt argues that we might have to look for help in areas other than moral philosophy where a distinction between veridical and pseudo intuitions needs also to be drawn. If somebody believes an optical illusion such as the Mueller-Lyer illusion to be true, the issue can be decided by appeal to external standards: we can show that both

---

[78] Stratton-Lake (2002), xlii-xlix proposes a similar division of the reactions to classical moral intuitionism.
[79] For a similar reconstruction of arguments against intuitionism that are based on moral disagreement, see Huemer (2006), 128-154.
[80] Brandt (1959), 192-196 and 200f.

lines are of equal length with the help of a ruler or we can use theories of the

physiology of perception in order to explain why the perception is illusionary. But

in the case of supposed "moral illusions" which are not due to defeaters such as

emotional exhaustion, faulty reasoning etc., intuitionists haven't provided yet a

method that allows to tell apart false from veridical intuitions.[81]

This lack of external standards makes it impossible find a common ground

everybody is obliged to accept. Intuitionism which started out as a defense of

objectivity in ethics ultimately collapses into subjectivism.[82]

On Ross' behalf, several arguments can be offered in response to these

worries.[83] To start with, if one rejects coherentism or an infinite regress and does

not believe that fundamental moral principles can derived from non-moral

premises, then justification will run out at some point. The disagreements that

might result are however few; they concern a narrowly circumscribed set of basic

principles such as the *prima facie* wrongness of lying or the fact that we have a

prima facie duty to promote beneficence.

Against the first argument, it Rossian intuitionists can respond that

widespread and persistent disagreement can be found in many areas of discourse

without raising the worry that there might be no knowledge in these areas.[84] Also,

the argument overlooks how difficult it is to exclude the presence of defeaters such

---

[81] Brandt (1959), 194-196 and 201. Similar worries are raised by Binkley (1961), 43f., Nowell-Smith (1954), 45, Hare (1998), 82-102, Korsgaard (1996), 32f., Warnock (1967) 14f. and Johnson (1957). In order to avoid redundancy, I will not discuss their arguments in detail.

[82] See Hare (1998), 92-95.

[83] For an excellent response, see Stratton-Lake (2002).

[84] The list of examples is long and includes e.g. disagreement about the correct interpretation of quantum mechanics, about the effects of taxation, the cause of illnesses etc. See Huemer (2006), 134f.

as biases, emotional arousal, errors of reasoning etc. Those factors, it can be assumed, might be responsible for a large number of disagreements.[85]

Against the second argument that intuitionists have no method of resolving those disagreements, intuitionists can reply that most of those cases are based on disputes over non-moral issues.[86] Furthermore, the fact that some moral beliefs are foundational does not mean that they are irrational. As seen above, Ross believes that only mature agents who have reflected on a proposition and who have considered what the tradition and other experienced agents have concluded are in a position to hold them as self-evident. Based on Ross' own writings, it is not clear whether he allows for epistemic overdetermination, but it would certainly not contradict any of his arguments.[87] If that's the case, intuitionists do not have to reject any argument that can help to solve disagreements.

Hence, a lot can be said in defense against arguments from disagreement. This, however, does not mean that the worry is without merit. Even if the disagreements that might occur in Ross' framework are few, they are of fundamental significance and he owes us a more explicit account of how to handle them.

2.3.2 Intuitionism explains moral cognition by appeal to a mysterious faculty

Another commonly made criticism against intuitionist moral epistemology is that philosophers such as Ross rely on a "mysterious faculty" in order to explain how we are able to grasp self-evident moral truths. Other than the first worry, this

---

[85] See Huemer (2006), 137-139.
[86] See Huemer (2006), 143.
[87] This is one of the main points of Audi's Kantian intuitionism. See his (2004), 80-119.

kind of argument is often not informed by the writings of intuitionists. I will limit

my discussion to the most widely known defender of the claim, John L. Mackie. In

his *Ethics – Inventing Right and Wrong*, he connects a metaphysical with an

epistemological criticism against intuitionism (and, as he believes, any other

ethical view which defends objective values):[88]

> If there were objective values, then they would be entities or
> qualities or relations of a very strange sort, utterly different from
> anything else in the universe. Correspondingly, if we were aware
> of them, it would have to be by some special faculty of moral
> perception of intuition, utterly different from our ordinary ways of
> knowing anything else.[89]

The strangeness of "objective values" lies in the fact that they not only objective,

but also prescriptive.[90] There is, Mackie claims, no other kind of object that is both

truth-apt and has "to-be-pursuedness somehow built into it."[91] No other epistemic

faculty has the ability to pick out those kinds of entities, and if intuitionists rely on

such a faculty, they face an unredeemed and probably unredeemable burden for

making a case for them.[92]

It would be easy to dismiss this criticism insofar as it applies to Ross, but in

light of our guiding question of what challenges recent intuitionists needed to

address in order to make intuitionism a contender among contemporary ethical

---

[88] see Mackie (1977), 37.

[89] Mackie (1977), 38.

[90] I assume that by using the term "value," Mackie does not intend to exclude statements about what is right from the scope of his argument. As the name of the book indicates ("Inventing right and wrong"), his target is the objectivity of ethics, including statements about what is objectively right.

[91] Mackie (1977), 40. Mackie is not explicit on what he means by "objective," but I assume it includes that ethical values are truth-apt. Nothing in my argument depends on it.

[92] In all fairness, Mackie does not limit the account the intuitionist would have to provide to a perceptual model: "[N]one of our ordinary accounts of sensory perception, or introspection or the framing and confirming of explanatory hypotheses or inference or logical construction or conceptual analysis, or any combination of these, will provide a satisfactory answer." (1977), 39.

theories, it make sense to give Mackie a more charitable reading. Ross is immune against the "argument from queerness" as stated since he does not accept an assumption that Mackie makes. For him, the motive to act (what Mackie calls "to be pursuedness") and the rightness of an act are separate.[93] If both fall apart, moral values lose their "queerness." We would still need an account of how we come to come to possess self-evident moral knowledge, but the task looks less daunting.

Still, the worry that intuitionists like Ross rely on an unexplained moral faculty should however still be taken seriously. One way to understand it is to argue that Ross does not give us a detailed enough account of how the process by which we come to hold moral knowledge is supposed to work. As seen above, he repeatedly likens moral to mathematical knowledge. But how much work is this analogy able to shoulder? No matter how much reflection, is the conceptual claim that we have a duty to promote justice ever as self-evident as the statement that "2+2=4"? I will extend these arguments in my own criticisms of Ross below, but the point is already clear: Ross does not give us a detailed and convincing positive argument for his account of moral knowledge. Even if the challenge he is facing is not as big as Mackie makes it sound, self-evident knowledge of moral principles is still something that needs argument. Less mysterious is still mysterious.

---

[93] Ross (1930), 158: 'Human nature being what it is, the latter thought [that this action would be pleasant]arouses an impulse to action much more constantly than the former [that this action would be right]', quoted after Darwall (2002), 265. For more on Ross' position regarding moral motivation, see Darwall (2002), 264-268.

2.3.3 Summary of criticisms

In summary, critics of intuitionism have raised a number of questions that any re-formulation of the theory would have to take into account. These are the main ones:

1. If moral intuitions provide us with self-evident knowledge and if moral facts are objectively true or false, how can there be moral disagreements?

2. How can we distinguish between genuine and pseudo moral intuitions?

3. Does knowing a self-evident proposition imply that we are aware that it is necessary?[94]

4. How does the process of reflection which results in self-evident moral knowledge work? How does it differ from reflection that results in non-self-evident knowledge?

2.4 My own criticisms

Besides presenting the criticisms against Ross' moral epistemology that have found wider echo in the philosophical literature, I also want to add a few worries of my own. They center around three topics: conceptual reflection, intuitive induction and the significance of general knowledge for moral judgment.

2.4.1 Conceptual reflection

Belief in a self-evident propositions stands apart from other kinds of belief insofar as it does not require external premises in order to be justified. If such basic beliefs exist and also confer justification onto other beliefs they stand in a relation

---

[94] This question has not been explicitly discussed above, but Brandt raises it in the context of whether it is possible to distinguish genuine and false instances of moral intuitions. See his (1959), 201.

with, this would offer an attractive response to a serious philosophical problem: can we justify our beliefs without falling into an infinite regress or resorting to coherentism? Foundationalism would be very attractive for *moral* beliefs since justification for normative claims is considered an especially thorny issue. However, the idea that some beliefs are – unlike all other beliefs – able to bootstrap themselves into existence is a proposal that requires good reasons in order to be convincing. Does Ross deliver those arguments and make a plausible case that beliefs concerning *prima facie* duties are justified just in virtue of their content? I will raise two skeptical points.

First, the criterion for self-evidence is that a proposition passes the "test of reflection." What does this test consist in? We are unable to believe its denial because we so firmly believe the proposition upon reflecting on it.[95] Is it a worry that some people hold with unshakeable confidence all kinds of beliefs that an observer would not consider as self-evident, such as convictions about the superiority of their race, sport clubs or horoscopy? To deal with this problem, Ross excludes defeaters such as intellectual immaturity, lack of reflection and inconsistency with other beliefs from affecting self-evident beliefs. But even assuming that this strategy works in excluding convictions such as those above, he does not give us a *positive story* about why our firm convictions in other beliefs grounds their self-evidence. Why is it that mature, reflected and consistent agents come to have strong confidence in beliefs about principles of *prima facie* duties and why does their confidence confer a special epistemic status on those beliefs?

---

[95] See Ross, (1930), 39f.

Looking for companions in guilt, Ross relies on two analogies, one between moral perception and natural science and the other with mathematical knowledge. I will deal with the second case below and address the first here.

Natural science, Ross argues, should not rely on "'what people really think'" because this would be prone to misinterpretations of sense-experience.[96] Rather, sense-experience itself should be the "real data."[97] In the case of moral knowledge, sense-experience does not provide us with direct access to moral facts and we have to resort to idealized reflections. He concludes his comparison between natural science and morality by claiming that "the moral convictions of thoughtful and well-educated people are the data of ethics just as sense-perceptions are the data of a natural science."[98] There is an obvious gap in this argument: if natural science would be unreliable if it was based on "what reasonably thoughtful and well-educated people think,"[99] does the same limitation not also apply to moral knowledge? Ross would need an additional premise showing what makes reflection an adequate means of gaining knowledge in one area but not in the other. This brings out just what is missing in Ross' picture: a positive theory of what it is about reflection that gives us access to moral facts.

A second issue concerns the demarcation of self-evident from non-self-evident moral beliefs. Even if some moral beliefs are self-evident, it would go too far to claim that all of them enjoy this privileged status. But where and how should

---

[96] See Ross (1930), 40f.
[97] I will leave aside Ross' simplified view of how science operates; what counts here is his view on moral perception.
[98] Ross (1930), 41.
[99] Ross (1930), 40.

the line be drawn? For Ross, only the fact that certain considerations invariably

count in favor or against an action, but not their strength or rank-ordering in

comparison to other duties, are beyond doubt. Although he does not believe that

there are general rules indicating which *prima facie* duties carry more weight than

others, he admits that duties of perfect obligation such as the keeping of promises

or of reparation are normally more stringent than their competitors.[100] What makes

Ross so confident that those kinds of beliefs cannot be as self-evident as beliefs in

prima facie duties? Perhaps he assumes that beliefs which are self-evident must

also be exceptionless, but this is certainly no conceptual necessity.

2.4.2 Intuitive induction

Is it possible to generalize from moral knowledge in several individual

instances to a universal rule? As argued above, the concept of "generalization" can

be understood in two ways. Either, we encounter repeated instances where an x is

y; after having encountered a sufficient number of instances, we explain our

observations with a general rule that all x are y and assume that the high number of

observed instances justifies our expectation that in all likelihood, other instances

instantiate the same rule. Or, those repeated instances serve as examples which

illuminate the nature of the object being observed; eventually, the rule that all x are

y becomes self-evident to us. Ross has reason to prefer the second interpretation.[101]

According to him, possessing self-evident knowledge that x is y entails that we –

---

[100] See Ross (1930), 41f.

[101] The same applies to abduction: as argued above, Ross has strong requirements for moral beliefs to count as knowledge, and an abductive argument would – in virtue of the fact that it does not guarantee the truth of the conclusion – fall short of that standard. If contemporary intuitionists are willing to lower those standards (and there are good reasons for doing so), this would allow them to also use enumerative induction and abduction.

as seen above – cannot even imagine the claim being false. Further instances

where x turns out to be y could therefore not increase our confidence in the

generalization, as much as encountering another case where two and two matches

add up to four matches won't augment the strength of our trust in the general rule

once we have grasped it. In the mathematical example, this explanation seems

indeed convincing, but does this plausibility also carry over to the moral case? As

seen above, Ross himself stresses an important disanalogy. Mathematical facts,

e.g. that "[a] triangle which is isosceles necessarily has two of its angles equal,

whatever other characteristics the triangle may have,"[102] are context-independent:

the size of the triangle or the presence of other geometrical figures it intersects

with cannot change the facts about its angles. Moral features are different. If we

want to know what difference a lie makes in a certain situation, all sorts of facts

can play a role. Even if in many cases, lying counts against an action, there are

disablers, enabler, disablers of disabler, disablers of enablers, fortifiers etc. which

would all have to be taken into account in a principle specifying the role of

lying.[103] What we learn from each case is hence not about the behavior of lying per

se, but about the difference lying makes in situation x, y, or z (for a detailed

discussion of this issue, see appendix 1). Even if we assume that eventually, all

conditions that affect it can be specified,[104] it is clear that it would be very hard to

---

[102] Ross (1930), 33.

[103] It is important to keep in mind that the principles Ross is after do not take the form "other things being equal, lying is a bad-making feature" but "lying always counts *prima facie* against an action." It is therefore not possible to reply to the present problem by saying that Ross is only interested in *prima facie* duties which allow that a duty can be defeated – in question here is whether the duty in question counts in favor or against an action to begin with, and that's an information that a *prima facie* duty contains.

[104] For further discussion on the topic, see McKeever and Ridge (2006), 25-45. The possibility that the general rule is open-ended raises serious problems for the project of formulating a list of *prima facie* duties**.**

distill the nature of lying by considering separate situations which have this feature.[105] This makes it questionable whether normal agents would have the resources to go through the process of intuitive induction which might result in generalizations so complicated that they are possibly useless.

A further challenge for Ross arises from his characterization of how we come to apprehend *prima facie* rightness or wrongness in concrete situations. As discussed above, this process cannot be the same as the one that we employ in conceptual reflection. What happens is that we consider a concrete situation and emerge with a conclusion concerning whether the features present in this situation count in favor or against doing it. In a further step, we weigh them against other *prima facie* duties and come to a conclusion about our duty all things considered. Do we make this judgment based on an inference (which would have to be different from those employed in conceptual reflection) or via perception?[106] How do we bridge the gap between a situation composed of non-moral facts to a moral conclusion? (Note that in the mathematical example to which the moral case supposedly relates "[i]n a precisely similar way,"[107] this problem does not arise). The plausibility of Ross' account depends on those details, but unfortunately, he does not spell them out.

---

[105] A version of this idea is the main argument for Jonathan Dancy's particularism. For the whole argument, see his (2006).

[106] In the *Foundations of Ethics*, Ross argues against the direct perception of moral properties: "[Rightness] is not an attribute that its subject is just directly perceived in experience to have, as I perceive a particular extended patch to be yellow, or that noise to be loud […] [I]t is only by knowing or thinking my act to have a particular character, out of the many that it in fact has, that I know or think it to be right." Ross (1939), 168.

[107] Ross (1930), 33.

2.4.3 The significance of general principles for moral judgment

In Ross' framework, there is moral knowledge on three levels, and it is an important question how these are related. The most abstract level deals with general principles of *prima facie* duty. Second, there is the knowledge that a certain aspect of a concrete instance – such as lying or the chance to benefit somebody – counts against or in favor of an action. This is the kind of knowledge that intuitive induction starts with. Last, there is our duty all things considered about which, as Ross argues in both in *The Right and the Good* and in *The Foundations of Ethics*, we can never reach a level of justification sufficient for knowledge.[108]

While in *The Right and the Good*, Ross develops no explicit account of the significance that general knowledge has for moral judgments on the concrete level, the context of his discussion on moral epistemology suggests – as argued above – that it plays an important role. Things are different in *The Foundations of Ethics*. There, he argues that only in rare circumstances, judgments about duties all things considered relies on abstract principles. Does this turn his overall account into a reconstruction of common sense morality that is mainly interesting for philosophers, or can our moral life benefit from a systematic account of the principles which stand behind our decision-making?

Even in *The Foundations of Ethics*, Ross' position is probably not as extreme as it sounds. Besides claiming that we have "direct insight" into what

---

[108] The claim that we cannot have knowledge about our duty all things considered depends, as has been argued above, on Ross strong requirements on the notion of moral knowledge.

duties are relevant in concrete situations,[109] he also maintains that "the final judgment is not a logical conclusion, but yet is something that presupposes preliminary mental acts, in which we study the situation in detail, till the morally significant features of it become clear to us."[110] If, as Ross certainly accepts, there is a list of *prima facie* duties to which we have epistemic access, it is hard to imagine why we wouldn't make use of it while figuring out which features in a given situation make a normative difference regarding our duty all things considered. Going beyond Ross, it is not difficult to come up with scenarios in which appeal to general principles plays an important role in our moral decision making, for example when facing especially difficult decision on a personal or societal level (e.g. "Should we unplug Aunt Susan?" type of scenarios often evoke discussions of whether considerations of beneficence trump our respect for life) or when trying to figure out whether our various ethical decisions follow coherent patterns and we therefore test whether we are consistent in applying the same principles. Tom Beauchamp and James Childress have even built a whole approach to biomedical ethics on the idea that ethically challenging situations can be fruitfully analyzed by means of a small set of ethical principles.[111]

This discussion suggests that the question of whether general moral principles are of use in everyday moral deliberation has no single answer. In some

---

[109] Still stronger, Ross even says that "I never seem to be in the in the position of not seeing directly the rightness of a particular act of kindness, for instance, and of having to read this off from a general principle" Ross (1930), 171; see also on p173.
[110] Ross (1939), 186.
[111] See Beauchamp and Childress (2012).

cases – such as, to use Ross' example, when helping an old man over the street[112] –
we have direct insight into the duties that are at stake,[113] while in other places,
without principles our ethical thinking would be impoverished.

2.4.4 Summary of criticisms

To sum up, the discussion in this section has brought to light a number of
ways in which Ross' account of moral epistemology is problematic or
insufficiently elaborated. The following were the main points:

1. We lack a positive account of what the "test of reflection" which determines
   whether a belief counts as self-evident consists in.

2. The method of intuitive induction has shown to be problematic: how can we
   draw general conclusions from instances of morally relevant factors that are
   context-dependent?

3. In concrete instances, it is not clear how we come to self-evident moral
   knowledge from non-moral premises.

4. Ross seems to differ in his writings over the role that general principles play
   for moral decision-making. The more radical position he supports in *The
   Foundations of Ethics* is not plausible.

In the following chapters, I will suggest a path for an alternative account of
intuitionist moral epistemology which explains how we come to have knowledge

---

[112] See Ross (1939), 168.
[113] One can also adopt a stronger position. As James Nickel has suggested to me, the phenomenology that
sometimes, we see directly what we should do without taking recourse to moral principles might be an
illusion based on long habituation in following the principle of being helpful to others when the costs of
doing so are low.

of *prima facie* duties; the hope is that this theory gives the positive story that Ross is lacking. If successful, this account would also put the present worries to rest.

2.5 Conclusion

In this section, I have presented and criticized Ross' moral epistemology. The resources and challenges that his view presents form the background for recent intuitionists who have tried to reformulate the theory with the aim of making it again a plausible contender among the main normative accounts. Their success depends among other things on how well they are able to deal with the objections raised in this chapter. Next, we turn to a philosopher in the center of this project: Robert Audi.

Chapter 3: Audi on the Self-Evidence of *Prima facie* Duties

3.1 Introductory Remarks

Soon after Ross, intuitionism lost its standing as a widely respected ethical theory. In the period from 1940 to 1980, there were few defenders of the view.[114] The reasons for this shift were manifold. The ones that were inherent to the views of Ross and his fellow intuitionists were mentioned in the last chapter, often focusing on the notion of self-evidence or the problem of moral disagreement. The fact that most of those criticisms don't withstand critical scrutiny or a close reading of the philosophers they attack (as discussed in the last chapter) suggests that they emerged in a philosophical climate that was already hostile to intuitionism. As a matter of fact, intuitionism stands at odds with two dominant currents of its time.[115] Logical positivists, one the one hand, denied that moral discourse is cognitively meaningful and rejected moral realism in favor of expressivist views which claim that moral judgments express emotive attitudes.[116] In general, logical positivists argued that philosophy was a second-order discipline, describing the principles on which first-order discourse relies without taking a stance in that discourse.[117] In this context, the idea that self-evident beliefs provide us with access to mind-independent moral truths is not a claim that many philosophers took seriously. Ordinary language philosophy, the second dominant current of the period, was receptive to common sense, but believed that theoretical systemization of areas of discourse

---

[114] One of the few notable exceptions was A.C. Ewing.
[115] For a more detailed historical account, see Hurka (2004), 248f.
[116] See Ayer (1936) and Stevenson (1944).
[117] See Hurka (2004), 249.

violates the way that language is used. This contradicts the intuitionist method of deriving general principles from common sense morality.

Hence, ethical intuitionism is at odds with both views. Viewed from the perspective of those two paradigms, it is therefore not surprising that intuitionism had little appeal.[118] Even after the influence of logical positivism and ordinary language philosophy had diminished, the dismissive attitude against ethical intuitionism remained in place. As an illustration, Rawls on the one hand recognizes intuitionism as one of utilitarianism's main historical opponents,[119] but in his own discussion of the view, he subscribes to an understanding of the theory that is impoverished compared to the traditional view. In his account, there is no recognition of intuitionism's main virtue, its close relation to common sense morality. Rawls further believes that the notion of self-evidence can be 'left aside'.[120] What remains is Rawls' claim that there is a plurality of potentially conflicting first principles and the denial "that there exists any useful and explicit solution to the priority problem" (i.e. the issue of how to rank principles such that conflicts among them can be resolved).[121] Thus understood, intuitionism is 'half a theory' that is a serious contender only in case all other theories that assign weight to moral principles fail.

If today, we see a revival of ethical intuitionism,[122] this is to a significant extent the merit of Robert Audi's work.[123] Starting in the 1980s until today, Audi has worked

---

[118] As Hurka fittingly summarizes, "[o]n the one side was a view friendly to abstract principles but hostile to common sense; on the other was a view friendly to common sense but hostile to abstract principles." Hurka (2004), 249.

[119] See Rawls (1999), xvii.

[120] See Rawls (1999), 31.

[121] See Rawls (1999), 36.

[122] This revival is reflected by an increasing volume of academic work devoted to ethical intuitionism: Ross (2004) [prefaced by a substantial discussion of his theory by Philip Stratton-Lake], Stratton-Lake (2002),

out various aspects of a comprehensive version of intuitionism that takes its main inspiration from Ross. The challenges his efforts are facing are two-fold. First, as seen in Rawls' discussion of intuitionism, the prevalent picture of intuitionism does not have enough theoretical structure to compete with other theories. Second, many of the epistemological and metaphysical views of earlier intuitionists such as Ross' are thought to be extravagant.[124]

In response, Audi develops a normative theory that takes seriously the arguments of traditional intuitionists by showing how they can avoid most of the charges raised against them. But beyond that, Audi also argues for his own version of intuitionism that differs in important respects from that of his predecessors. In order to show the scope of his theory, here are four of his arguments which set his work apart from that of earlier intuitionists:

1. In regards to the notion of self-evident moral knowledge, Audi offers a detailed account of the concept and argues that it commits us to less than has usually been assumed. For example, an agent can possess self-evident knowledge without being aware that her beliefs are self-evident; further, self-evident beliefs are not necessarily indefeasible; self-evidence is compatible with supporting premises as long as those are not external to the belief at stake; and a person can adequately understand a self-evident proposition without believing it.

2. It is often thought that if a proposition is self-evident, there cannot be additional independent justification for believing it. Audi shows that intuitionists are not

---

Hurka (2011), Huemer (2006), Roeser (2011) and Shafer-Landau (2003, ch. 6). For an overview of the extensive recent literature on intuitionism, see Cowan (2012).

[123] Other factors also contributed to this process, for example the revival of moral realism and cognitivism (e.g. in the work of Jonathan Dancy and David McNaughton).

[124] See Audi (2004), ix.

committed to this assumption; self-evident beliefs can be epistemically overdetermined and receive justification from various sources. In line with this view, he develops 'Kantian intuitionism', a view according to which Kantian ethics provides alternative support for the list of *prima facie* duties that intuitionists believe are also self-evident.

3. As Ross admits, his list of *prima facie* duties is only preliminary. Audi makes suggestions for a more plausible collection of morally relevant considerations that are more specific than Ross' and include further *prima facie* duties, such as a duty to respect persons.

4. Contra Ross and most traditional defenders of intuitionism, Audi argues that intuitionism is not committed to any particular theory in moral metaphysics or semantics.[125] While such a view is weaker than Ross', it is stronger than Rawls' reconstruction insofar as it includes a foundationalist epistemology as a core element of intuitionism.

I will focus my discussion on the first point. First, I introduce Audi's account of self-evidence and discuss some implications of his theory: the relation between adequately understanding and believing a self-evident proposition, Audi's notions of adequate understanding and reflection, his view on epistemic overdetermination and the distinction between self-evidence and second order knowledge. After clarifying Audi's account and showing its philosophical commitments, I move on to the question of why adequate understanding is connected to justification. I critically discuss two possible answers. The

---

[125] See Audi (2004), 2.

first response argues that *prima facie* duties express paradigmatic reasons, while the second reply makes the case that *prima facie* duties are connected to abstract entities.

3.2. Self-evidence: Audi's non-entailment strategy

Audi's account of self-evidence pursues a "non-entailment strategy": if the only function of self-evidence is to explain how we gain epistemic access in a non-inferential way to moral principles, features like infallibility or immediacy become inessential and can be left out. In order to mark the difference from the classical approach, Audi gives his weaker concept the name of "soft self-evidence."[126]

Relating his approach to the intuitionist' tradition, Audi believes that soft self-evidence can account for Ross' list of *prima facie* duties[127] (or at least for a version of the list – Audi himself proposes a slightly changed collection of first principles).[128] This spells hope that the problems raised against intuitionism in the previous chapter are not insurmountable. Since Audi offers a more detailed account of how we come to have knowledge of principles, traditional worries about a lack of such explanation in Ross might be laid aside.

3.2.1. Audi's definition of self-evidence

Turning now to the details of Audi's theory, his definition of self-evidence spells out the foundational role the concept plays for our acquisition of knowledge (moral and otherwise). A self-evident proposition is:

> a truth such that any adequate understanding of it meets two conditions: (a) in virtue of having that understanding, one is justified in believing the

---

[126] See Audi (2004), 53f.
[127] See his (1999), 225.
[128] See his (2004), ch. 5

proposition (i.e., one has justification for believing it, whether one in fact believes it or not); and (b) if one believes the proposition on the basis of that understanding, then one knows it.[129]

3.2.1.1. Understanding a self-evident proposition and believing it

A number of features of this definition are noteworthy.[130] First, Audi allows

for a gap between self-evidence and belief. One can, he claims, adequately

understand a self-evident proposition without believing it and still be rational.[131]

Prior to discussing what factors can cause that gap, it is important to note that

allowing for this possibility comes at a cost. Insofar as self-evident propositions

express facts about moral principles, those who understand but do not believe such

propositions hold wrong beliefs. If this happens frequently, skepticism regarding

our knowledge of *prima facie* duties follows since it cannot be guaranteed that

rational agents who are exposed to the relevant data get it right. The issue would

be less pressing if it could be shown that in typical cases, adequate understanding

of a self-evident proposition results in knowledge.[132]

---

[129] Audi (2008), 151.

[130] A third feature is that Audi distinguishes in his definition between the justified belief in and knowledge of a self-evident proposition – see his (1999), 210. I will not discuss this point since not much depends on it for present purposes.

[131] Audi (1999), 207: "Quite apart from how much of a deficiency in rationality may be indicated by certain persistent failures to form beliefs in such cases, there are surely instances in which one can see *what* a self-evident proposition says – and thus understand it – before seeing *that*, or how, it is true." As this quote makes clear, Audi's claim is not that it is psychologically possible to withhold belief in self-evident propositions; rather, his interest is in whether it can be rational to do so.

[132] Here is a reply against this line of argument: perception and memory are also fallible, and they are exposed to the same gap. Sometimes, we have the impression that we perceive something, but we still withhold belief (because, for example, we might have reason to assume that our vision is distorted). However, this does not commit to skepticism regarding those faculties. Why should the same issue then lead to skepticism regarding moral knowledge, as I suggest? The answer is that whether the gap between adequate understanding and belief is problematic depends on how often it occurs. If it is rare, as with perception and memory, the gap does not undermine the epistemic function of the faculty, but if it occurs to frequently, it can lead to a worry about skepticism. What is needed is an explanation as to why in the case of moral knowledge, the gap is rare.

One reading of Audi's distinction between understanding a self-evident proposition and believing it would easily solve the problem. According to it, the gap is explained by pointing out that the self-evident is sometimes not identical with the obvious; an agent can be aware of a self-evident proposition, but needs time in order to come to believe it. Audi's own writings suggest such an interpretation: "[O]ne can fail initially to "see" a self-evident truth yet later grasp it in just the way one grasps the truth of a paradigmatically self-evident proposition."[133] As an example, take the proposition that "the existence of great-grandchildren is impossible apart from that of four generations of people" which will – although the proposition is self-evident – for most people requires reflection before they are ready to believe it.[134]

However, this reading of Audi's distinction between understanding and believing a self-evident proposition would make his point trivial. After all, it is a psychological fact few people would deny that we need time proportionate to the complexity of a statement in order to process it properly; only after that period can we be expected to have an appropriate cognitive response. Ross also stresses that it might need repeated reflection before embracing the truth of a *prima facie* duty.[135] Second, even without having analyzed Audi's notion of "adequate understanding," any reasonable interpretation of the term should imply that we have finished the process of understanding what the proposition in question means. Hence, any sensible definition of self-evidence should allow for cases where we are temporarily hesitant whether we should believe a proposition based on unclarity about its content.

---

[133] Audi (2004), 49. See also his (1999), 206.
[134] Audi (2004), 49. See also his (1999), 207. Another example Audi uses is the proposition that "the mother-in-law of a spouse of a person's youngest sibling is that person's own mother" Audi (2013), 94.
[135] See Ross (1930), 24.

Even granting that situations where a gap between understanding and believing a self-evident proposition due to temporary lack of clarity are possible, Audi could still preserve the entailment by excluding those situations from his definition of self-evidence, e.g. by adding that a clause such as "adequate understanding of a self-evident proposition entails believing it as soon as temporary doubts about its content have been resolved."

There is however a second and more interesting interpretation of Audi. According to it, the gap between self-evidence and belief can be permanent. Besides temporary factors such as caution and slow uptake, skepticism – which might be a permanent feature of our belief-system – can explain our rational resistance to believe self-evident claims.[136] The idea seems to be that if I am skeptical about a certain area of discourse – be it about morality as a whole or only about the claim that some moral truths are self-evident – then I am entitled to withhold belief about self-evident propositions even if I adequately understand them. Details matter however. Simply taking a skeptical stance for no good reason will certainly not be enough to count as an excuse for withholding ones belief in a self-evident proposition. Otherwise, why should *anyone* who is in doubt about the truth of a *prima facie* duty (assuming *prima facie* duties are self-evident propositions) feel compelled to believe it? Worse, why should those who already believe a self-evident proposition not feel that their belief is based on weak grounds that cannot withstand scrutiny? In response to this worry, it needs to be the case that the skeptic has good reasons for his stance, reasons strong enough to undermine belief in adequately understood *a priori* propositions.[137]

---

[136] See Audi (1999), 207.
[137] As discussed below, Audi believes (as Ross does) that self-evident propositions are *a priori*.

An explanation of the gap between adequate understanding and belief can be found in Ross' discussion of the defeasibility of intuitions:[138] in cases where two propositions that seem self-evident to us conflict with each other, Ross argues that we have to ask ourselves which one is better supported by common sense and choose it over the other belief.[139] Applied to contemporary terminology, our beliefs ideally stand in a reflective equilibrium with other beliefs. If adopting a self-evident propositions would be hard to reconcile with other firmly held beliefs (which are often themselves be based on self-evident beliefs), we might have reason to reject them despite their attractiveness. This illustrates how there can be cases in which we fully understand a self-evident proposition, but do not adopt it as a belief.

3.2.1.2. Adequate understanding

The next feature that is relevant for Audi's account of self-evidence is his interpretation of the term "adequate understanding." It can be placed on a continuum between perfect and deficient understanding and contrasted with both endpoints. Perfect understanding ("comprehension completeness") includes awareness of all of a proposition's logical implications. Making it a requirement for belief or knowledge would raise the bar too high (under these conditions, few if any people would possess knowledge). On the other end, a deficient understanding of a self-evident proposition would – even if it resulted in believing a correct statement – make truth a matter of luck and therefore does not provide sufficient

---

[138] See Ross (1930), 41.
[139] If one assumes that self-evidence is a mark of truth and that true beliefs cannot contradict each other, the intuitionist should say that the belief that needs to be discarded only seems self-evident, but that this is shown to be false by the fact that there is another belief that has a higher credence and is incompatible with the first belief.

justification for believing it.[140] Even though it is hard to give a general account indicating what conditions need to be in place for a sufficient level of understanding (the level might vary from case to case, as Audi suggests), the concept can be made clearer by contrasting it with various ways in which understanding can be deficient.[141] Two are relevant here, misunderstanding and insufficient understanding. In the first case, the error consists in assigning a false meaning to the terms used in the proposition.[142] Insufficient understanding means that one does not grasp a certain aspect of the proposition; in Audi's example, an agent lacks knowledge of the correct modal status of the claim that knowledge entails true belief.[143] The difference between misunderstanding a proposition and having an insufficient understanding of it is a matter of degree: being mistaken about the modal status of a proposition is compatible with having overall a good grasp of a proposition whereas misunderstanding a proposition implies a more serious shortcoming.

Below, we will see that the example Audi uses in his characterization of insufficient understanding stands in tension with his later argument that the content of a self-evident proposition is independent of second-order knowledge about its modal status.

---

[140] See Audi (1999), 208.

[141] Audi himself believes that a general account of understanding might exist, but he does not attempt to formulate it. See his (1999), 208.

[142] In the example that Audi gives, somebody equates the four generations in the proposition about great grandchildren with four successive thirty-year spans. Audi (1999), 207.

[143] See Audi (1999), 207.

3.2.1.3. Reflection

Moving from clarifications concerning Audi's definition of self-evidence to a number of distinctions he introduces in order to either sharpen aspects he shares with other intuitionists or to stress that his theory is not committed to implausible commitments that other intuitionist's views carry, Audi discusses whether self-evidence is compatible with reflection. The underlying issue is how a proposition that is epistemically self-sufficient, i.e. which is not derived from other premises, can be justified or supported by a reflective process which involves inferences. This problem does not occur with all self-evident propositions, as some can be grasped without the need for reflection (some non-moral propositions such as "All sisters are female" belong into that category). Others however require further thought before a normal agent has acquired enough justification to belief them. Moral principles of the sort that intuitionists are interested in (e.g. the *prima facie* duty not to tell lies) fall into the second category. In order not to undermine the foundational nature of self-evident propositions, the inferences that are invoked in order to support these claims must be *internal*.[144] As Audi defines it, internal inferences are contrasted with external inferences which 'entail epistemic dependence on one or more premises.'[145] An inference counts as internal, if it involves thinking through a logical statement that is supposed to be self-evident or

---

[144] The present focus is on what is required for self-evident beliefs to count as foundational. This does not exclude, as Audi explicitly argues in his discussion of epistemic overdetermination, that there are additional non-foundational justifications for self-evident propositions.

[145] Audi (2004), 52. See also his (1999), 214.

remembering the definitions of the terms used in a proposition.[146] Even if the

distinction between internal and external inferences has not been used by

traditional intuitionists, it is already implicit in the theories of philosophers such as

Prichard or Ross.[147] A merit of Audi's contrast between different forms of

inferences is that it explicitly addresses the popular worry that intuitionism relies

on a mysterious sense that gives us immediate insight into moral truth and does not

need or even allow for any form of reflection.

3.2.1.4. Epistemic overdetermination

One of Audi's most fruitful distinctions that allows him to develop ethical

intuitionism beyond its traditional form concerns the issue of whether self-evident

propositions can be epistemically overdetermined. Besides non-inferential

justifications based on self-evidence, Audi argues that self-evident propositions

can also be independently justified through inferences that rely on premises that go

beyond the proposition in question.[148]

For example, the *prima facie* duty of fidelity is according to intuitionism

self-evident and can be recognized without reliance on external inferences. At the

same time, it is possible to come to know it via reflection on Kantian ethics,

applying the categorical imperative to the problem of lying.[149] Using this method

---

[146] In Audi's example involving four generations, one might bring to one's mind the definition of a grandparent in order to understand the proposition. See Audi (2004), 52. In his (1999), 218, he calls these inferences "comprehensional."

[147] As seen in the last chapter, Ross explicitly allows for extended reflection on the meaning of a self-evident proposition.

[148] See Audi (2004), 41f.

[149] The details of how this might work and how to deal with differences between Rossian intuitionism and Kantian ethics (for example, Ross talks about *prima facie* duties while Kant believes we have a perfect duty to tell the truth) is not essential for the present debate. See Audi (2004), ch. 3.

allows Audi to systematize Ross' list of *prima facie* duties (or a variation thereof), showing how they relate to Kant's claim that we should treat humanity as an end, not as a means.[150] This answers a long-standing worry that intuitionism relies on an 'unconnected heap of duties'[151] without giving up the claim that all *prima facie* duties can be recognized separately and without reliance on premises.[152]

An further benefit of this move is that it offers additional resources in order to solve moral disagreements: if two agents differ in their moral beliefs and both consider those beliefs to be self-evident, they can introduce further considerations in support of their convictions.

3.2.1.5. Self-evidence and second-order knowledge

Another important contrast Audi draws in regards to self-evidence concerns our knowledge of the content of a self-evident proposition versus our knowledge of its self-evident status. Do we need to be aware of both in order to be justified in believing a self-evident proposition? At stake is how epistemically demanding (and consequently how accessible) self-evident beliefs are.

Audi argues for the weaker requirement according to which justifiably believing a self-evident proposition does not require awareness of its higher-order property concerning self-evidence:

> It is first-order moral propositions, such as the principle that there is
> a *prima facie* duty to keep promises, and not the second-order thesis
> that such principles are self-evident, which are the fundamental

---

[150] See Audi (2004), 90f.
[151] McNaughton and Rawling (2013), 291.
[152] The same method could also be used to permit pragmatic justifications or justifications that rely on reflective equilibrium.

kind of thing we must be able to know intuitively if a Rossian intuitionism is to succeed.[153]

The reason Audi offers for the independence of the two kinds of knowledge from each other is that judgments concerning the self-evidence of a proposition require "conceptual sophistication" and familiarity with technical concepts about "how [those propositions are] properly understood" or about their modal status.[154] However, in many instances people without such competencies can still be justified in their belief and have knowledge about the content of self-evident propositions. Therefore, such second-order knowledge cannot be a precondition for knowledge of first-order moral truths about self-evident principles.[155]

I want to make two critical remarks. First, a case can be made that adequately understanding a self-evident propositions requires both kinds of knowledge. If asked what justification is necessary for being justified in believing in a given proposition we believe to be true, knowing that it is self-evident reveals the right way to defend it (namely by elucidating its meaning instead of, say, searching for empirical considerations in its favor).[156] To put it another way, lack of awareness regarding the self-evidence of a

---

[153] Audi (2004), 43.

[154] See Audi (2004), 43.

[155] There is another objection against the claim that second-order knowledge is required in order to have adequate understanding of a self-evident proposition (raised by James Nickel): if one had to have higher-order knowledge of each proposition one knows, would this not lead to an infinite regress? In reply, one could limit the requirement that first- and second-order knowledge is required to self-evident propositions. Since second-order knowledge about self-evident propositions is itself not self-evident, there is no regress. The justification for the claim that second-order knowledge is required only for self-evident proposition is that for those propositions, we need assurance of their status as self-evident (since this status bestows justification), whereas there is no such need for non-self-evident propositions.

[156] It could be objected that since Audi allows for epistemic overdetermination, why should knowledge that a proposition is self-evident exclude any kind of additional justification that could be given in its support? As shown above, adequate understanding, which is a required for justification in a self-evident proposition, implies knowledge of the proposition's modal status. In the next section, it will be argued that propositions that are justifiable in virtue of their meaning (such as self-evident propositions) have to be true in all

proposition indicates that we are unclear about what kinds of reasons show it to be true. This displays a deficiency in our adequate understanding of the proposition. Hence, there is an important connection between first-order knowledge about the content of a self-evident proposition and second-order knowledge about its status as self-evident.

Audi's own examples can be used to illustrate this point. If I believe that the existence of great-grandchildren presupposes that there are at least four generations but am ignorant that the proposition is self-evident, I couldn't detect whether the wrong kinds of reasons were offered in its support and I therefore lack adequate understanding.[157, 158] For instance, somebody might try to support the claim by offering an inductive argument, recounting all the cases of great-grandchildren she had encountered and comparing whether in all cases, there are at least four generations. Without awareness about the status of this claim, I couldn't explain why those kinds of considerations cannot provide support for the proposition.[159]

But does not this set the bar too high, Audi might reply, depriving all those people who are unfamiliar with arcane philosophical terminology from having justified beliefs or knowledge of any proposition that counts as self-evident? My response is that even if the

---

worlds. Therefore, they are true by necessity. If that's the case, then in virtue of knowing that a certain proposition is self-evident, I also know that contingent empirical considerations cannot offer support for it.

[157] e.g. somebody might try to support the claim by offering an inductive argument, recounting all the cases of great-grandchildren she had encountered and comparing whether in all cases, there are at least four generations. While this wouldn't lead to false results, it wouldn't provide support for the proposition either.

[158] One could object that as argued above, adequate understanding lies on a continuum and does not require perfect understanding. Why not therefore assume that second order knowledge is beyond the requirement for adequate understanding? In reply, not knowing what kind of justification is able to vindicate a proposition displays a shortcoming that is in light of the discussion above (3.2.1.2) serious enough to qualify as an insufficient understanding of the proposition in question.

[159] See the previous footnote on the connection between second-order knowledge about the status of a self-evident proposition and necessity.

concept of modal status plays a role in understanding the concept of self-evidence,[160]

there are ways to express the idea in a non-technical vocabulary familiar to most

competent speakers in any given language. As long as it is understood that there are

propositions that are always true and could not be false and other propositions that can be

either true or false, people possess a sufficient sense of modal notions and should at least

in principle be able to detect into which group a propositions belongs.[161] Indeed, few

competent language users who haven't been explicitly trained in modal terms would

struggle to recognize that "all sisters are female" holds with necessity or that "Barack

Obama is the president of the United States" is contingently true. Presumably, intuitions

about the modal status of *prima facie* duties are more divided, but even here, it is not too

much of a stretch to assume that most competent moral agents would agree that e.g. the

statement "There is a *prima facie* moral reason against causing harm to other people"

could not turn out false.

In summary, Audi rightfully draws attention to the important distinction

between first- and second-order knowledge regarding self-evident propositions.

However, there is room for disagreement with him whether the first kind of

knowledge is sufficient for an adequate understanding of self-evident propositions.

3.2.1.6 Self-evidence and necessity

As the discussion in the previous section has shown, an account of self-

evident knowledge needs to take a stand on the relation between self-evidence and

---

[160] See Audi (2004), 43.

[161] Arguably, somebody who believes that statements such as "frozen water is ice" or "bachelors are unmarried man" might sometimes turn out false has significant deficits in their language competency and might be expected to also make mistakes when it comes to identifying self-evident moral propositions.

necessity. The relationship between these two notions can be investigated on three levels. First, it can be asked whether self-evident *propositions* are true by necessity. Second, it can be asked whether *believing a self-evident proposition* entails that we *immediately* see that it holds with necessity. Finally, it can be wondered whether *adequately understanding* and believing such proposition implies that we hold a view concerning its modal status *at all* (that is, even after careful consideration in situations that cannot be described as "immediate belief"). Audi takes a stand towards all of these issues:

> As I have characterized the self-evident, the notion is *non-modal*, at least in this sense. It is not defined explicitly in terms of a modal notion, and its application to a proposition does not immediately entail that the proposition is necessary. […] In denying an immediate entailment here I mean roughly that the claim that self-evident propositions are necessary is neither self-evident (in any preanalytic, intuitive sense), nor self-evidently entailed (in any such sense) by any philosophically uncontroversial account of self-evidence. Self-evidence, as I conceive it, is a matter of *how* a truth can be seen, not a matter of seeing what *kind* of truth it is. This absence of an immediate entailment from the self-evident to the necessary is a desirable result of my characterization. Not only is there dispute about whether every *a priori* proposition is necessary; it also seems possible to see the truth of a self-evident proposition without either seeing that it is necessary or even having the concept of necessity.[162]

All three of the worries raised above can be found here: the issue of whether "every *a priori* proposition is necessary," the question of whether this entailment must be seen immediately and the claim that the content of a claim can be grasped without awareness of its modal status.[163] Since the second and the third claim have already been dealt with above, I will not address them here again.

---

[162] Audi (1999), 211f.
[163] See also Audi (2008), 156f.

In response to the first issue, it is important to take into account that self-evident propositions are a priori (this is why an adequate understanding of such propositions is sufficient for their justification).[164] Statements that are *a priori* are often considered as necessary. What about self-evident moral propositions? Do they fall under the class of a priori propositions whose necessity is under dispute, as Audi seems to suggest (see his quote above)? In a footnote pertaining to that sentence, Audi references to Kripke's *Naming and Necessity*. There, Kripke uses the example of a standard meter bar in order to show that there are a priori propositions that are contingent.[165] His idea is that we can use a stick ("Stick *S*") to "fix the reference" of the term "one meter." The statement "The length of stick *S* at time t_0 is one meter" is, Kripke claims, contingent even though it is a priori since the length of the stick could have been different in another possible world.

Even if Kripke's arguments show that some a priori propositions are contingent, this wouldn't necessarily show that the kind of moral principles intuitionists are interested in (e.g. "It is *prima facie* wrong to break a promise") are also contingent. For in the meter-example, the term "meter" is defined by "fixing" it on a procedure which determines what the term refer to ("Whatever the length of Stick *S* is at time t_0").[166] Intuitionists however would reject the claim that the content of our common sense beliefs, i.e. what is self-evident to us, *fixes* what our *prima facie* duties are. In analogy with Kripke's meter-example, this would imply

---

[164] Audi (1999), 211: "It should be clear that if understanding a self-evident proposition provides justification for believing it, then self-evident propositions are justifiable *a priori* ." See also Audi (2008), 155.
[165] See Kripke (1972), 55.
[166] See Fitch (2004), 115f. "Fixing the reference" is contrasted by Kripke with normal definitions which "give the meaning" of a term.

that the content of morality is contingent on what seems self-evident to us at a certain time.[167]

This would squarely contradict Ross' views on the objective nature of morality (especially his views on moral progress), and there is no reason to think that Audi is committed to this view either.

Although at first, Audi considers the possibility that his account of self-evidence might leave room for the possibility that moral propositions might not hold with necessity, he ultimately rejects the idea. A contingent proposition *p* couldn't be justified merely by adequately understanding it since it is true in some worlds but not in others and the grasp of concepts alone cannot discriminate whether in any given world, a contingent proposition turns out to be true. Hence, self-evident propositions can only be statements that hold with necessity.[168]

3.2.2. What explains self-evidence?

So far, the discussion has focused on the implications of Audi's account of self-evidence: assuming that adequately understanding a self-evident proposition justifies our belief that the proposition is true, what else must be the case about the person doing the understanding and believing?

As seen above, Audi's view of self-evidence comes with minimal constraints. Adequate understanding, he argues, is compatible with withholding belief in a self-evident proposition. His notion of adequate understanding is not overly

---

[167] In other words, the issue is whether our common sense beliefs determine what is moral or whether what is moral determines what is self-evident.

[168] The present argument is not meant to demonstrate that basic moral propositions are necessary truths. Rather, my aim is to draw out what intuitionists such as Audi are committed to in regards to self-evidence and necessity given their overall views.

demanding. Self-evidence is compatible with certain forms of reflection. Self-evident propositions can be epistemically overdetermined. No second-order knowledge is required for adequate understanding of a self-evident proposition.

All this has allowed Audi to extend the appeal of intuitionist moral epistemology to philosophers who had formerly rejected self-evident moral knowledge because of its presumed philosophical baggage. But even if Audi's strategy takes care of this worry, there is a second issue that also needs to be addressed. What justifies the assumption that adequately understanding a self-evident proposition provides us with justification in the first place?

Audi offers two answers. The first explains the evidential force of intuitions in virtue of the fact that by understanding a self-evident proposition, we apprehend its relation to abstract entities. Audi also offers an alternative explanation for those who are skeptical of abstract objects. This second account focuses on the phenomenology of conceptualizing the notion of a moral reason. I will start by discussing the phenomenological account.

3.2.2.1. The phenomenological account of self-evidence

Self-evident propositions can either be analytic or synthetic. To take an analytic example first, thinking of a vixen entails the thought that vixens are female.[169] This *containment relation*[170] is not applicable to synthetic kinds of self-

---

[169] See Audi (2008), 479.
[170] Audi (2008), 478f. speaks of a "conceptual containment" or "containment relation."

evident propositions, for instance moral cases.[171] It is possible to adequately

understand the concept "moral reason" while being ignorant about any item on

Ross' list of *prima facie* duties.[172] This raises a problem. If the notion of a moral

reason does not entail any *prima facie* duties in the sense in which the concept of a

vixen entails the concept of being female, how are we to make sense of the idea

that moral principles are self-evident?[173]

What is needed is a weaker form of containment. To illustrate how this

might be possible, Audi offers an analogy to object perception. When looking from

the outside into a room, we can – depending on our standpoint – see certain items

of furniture, but not others. The same holds for reflection on moral concepts:

"[W]e can see that a certain concept is contained in another without seeing all that

the former contains."[174] When asked whether a certain type of act, e.g. "burning

someone with a red hot poker" is morally reprehensible, "naturally" some

paradigmatic reasons, among them the *prima facie* duty of non-maleficence, will

come to mind.[175] What happens here is that the situation we are confronted with

draws our attention to certain aspects of what counts as a moral reason. This

---

[171] As an example, according to the *prima facie* duty of non-maleficence, causing harm to another person is a moral reason against doing so. However, it is not obvious that the concept of a moral reason entails this duty.

[172] Audi (2008), 479: "The non-reductivist containment view that I suggest, then, does not presuppose something that at least Moorean intuitionism would imply is impossible: a naturalistic analysis of moral (*prima facie*) reasons. It does not presuppose even the weak kind of analysis that would be pluralistic, disjunctive, and non-hierarchical, for instance an analysis on which to be a moral reason is to be either an obligation of non-injury or an obligation of veracity or an obligation of beneficence... where the disjunction contains all the basic *prima facie* obligations."

[173] After all, if we take analytic self-evident propositions as the paradigm case, then reflecting on the notion of a moral reason should bring out all that is contained in the idea (as much as reflecting on the concept of a vixen entails the thought that vixens are female). Otherwise, how could the mere understanding of a proposition give us reason to believe that the terms in it are related?

[174] See Audi (2008), 479.

[175] See Audi (2008), 479.

explains how the *prima facie* duty of non-maleficence is conceptually related to the idea of a moral reason even though it does not follow analytically from the concept.[176] However, Audi grants that these "phenomenological points" don't amount to a full defense of the justificatory force of self-evidence.[177] Indeed, the analogy between object perception and the recognition what the notion of a moral reason entails helps us to understand how we come to recognize (non-analytic) self-evident propositions, but it does not tell us why that recognition bestows justification on our beliefs.[178]

This, Audi argues, can be achieved by recognizing that the question of what constitutes a moral reason can only be answered by pointing to examples where those reasons are instantiated. There is no need to further justify self-evident moral reasons such as the *prima facie* duty of non-maleficence. This is reflected by the fact that in normal everyday conversations – apart from discussions with a philosophical skeptic – it would be inappropriate to ask for a justification of certain moral reasons (such as: "You cannot just assume that to cause suffering to somebody else is bad. Do you have any reasons for believing this?").[179] Once we give up on the aspiration of vanquishing the skeptic, paradigmatic reasons can be "standardly presupposed" without providing additional inferential justification.[180]

---

[176] Given different concrete examples, intuitions about other *prima facie* duties could be elicited.

[177] See Audi (2008), 480.

[178] There are significant disanalogies between object perception and our linguistic intuitions, and even if we are justified in believing the results of the former, it does not follow that we can therefore trust the latter. For a detailed discussion, see Bedke (2008).

[179] It is important to keep in mind that we are only talking about *prima facie* reasons, not about our duty all things considered. Certainly, even on the level of normal language users who are uneager to question paradigmatic reasons, it is acceptable to require a justification for judgments all things considered.

[180] See Audi (2008), 480.

From a pragmatic standpoint, Audi's strategy is convincing. In order to get common sense morality off the ground, the intuitionist has to give up aspirations to take the moral skeptic on board. If somebody denies that there is a reason against causing suffering or lying, this should not put the burden of proof on the defender of common sense morality but mark the end of the dispute.

At places, Audi even makes his task seem more difficult than it is. There are two ways to interpret the containment-relation between the notion of a moral reason and the concrete *prima facie* duties.[181] First, one can wonder whether adequately understanding the notion of a moral reason entails belief in the list of *prima facie* duties.[182] This is a difficult challenge, and it is plausible to deny that there is a strict entailment. However, it is also possible to understand the entailment in the reverse direction. Here, understanding a description of the act-types depicted in the various *prima facie* duties implies the thought that there is moral reason against or in favor of doing them. It is the second, weaker form of entailment that intuitionists need in order to make the point that *prima facie* duties are self-evident. It would be interesting and provide further support for intuitionism if the inverse was also true – that the notion of moral reasons contained the list of all *prima facie* duties – but even if this cannot be established, this does not mean that those duties, taken in isolation, are not self-evident. This

---

[181] For an explanation why the containment-relation is important for self-evidence, see Fn. 173.

[182] For example at Audi (2008), 479: "Must the concept of a moral reason for action, then, contain, say, the notion of avoidance of causing pain if a Rossian principle of non-injury is genuinely self-evident?" He goes then on to argue that even though there is no strict entailment, it is natural that *prima facie* duties come to mind when reflecting on the notion of a moral reason (especially when confronted with concrete moral problems)

project is much less ambitious than Audi's attempt to show how paradigmatic reasons can be derived from the notion of a moral reason.

Some philosophers might consider this explanation of self-evident propositions in terms of paradigmatic reasons as too weak in order to provide epistemic justification. Moral realists, for example, who believe that correct moral judgments reflect stance-independent facts, would claim that in order to be justified in my moral beliefs, my beliefs have to express those facts.[183] It is possible that the inappropriateness of asking for a justification for basic moral reasons is a sign that those reasons express moral facts, but the present account does not provide an argument that this is the case. The fact that mature moral agents agree on basic moral standards and are furthermore unwilling to take challenges against those beliefs seriously, the moral realist might argue, does by itself not bestow justification on those beliefs and is compatible with their wrongness.

Many moral philosophers and epistemologists have good reasons to be doubtful about the idea of self-evident moral beliefs, and the fact that there is a level of moral discourse where raising those worries would be misplaced should not be able to silence those worries.[184] Instead, the apparent inappropriateness of asking for justification in everyday conversations could equally be explained as a convention aimed at facilitating a discourse that is fulfilling a practical purpose.

---

[183] Depending on their epistemic views, moral realists would however disagree over whether I need to be aware of that connection or whether I only need to employ a reliable mechanism of belief-formation.
[184] One does not need to be a moral realist in order to raise this objection against the paradigmatic reasons account.

To conclude, it might well be that we have reason to trust the paradigmatic reasons presupposed by common sense morality. However, many philosophers will reasonably seek a more robust explanation as to why this is the case.

3.2.2.2 Being in contact with abstract entities: the metaphysical account of self-evidence

The reason that Audi offers two explanations for knowledge based on self-evident propositions is that some, but not others are willing to accept the idea of abstract entities.[185] His first explanation was addressed at those who reject the idea, while his second account (which he himself prefers) is based on this commitment. According to it,

> [k]nowledge of the self-evident […] may be conceived as grounded in a kind of understanding. This kind of understanding both causally grounds beliefs constituting knowledge of the self-evident and characteristically *puts us in direct contact with elements that form a sufficient basis of the truth of the a priori propositions* it justifies us in believing and enables us to know. These elements are most plausibly conceived as abstract.[186]

At other places, Audi explains the fact that in self-evident propositions, there is a containment between a term and its definition in virtue of the "abstract entities and their relations" they refer to.[187]

---

[185] See Audi (2008), 478f.

[186] Audi (2008b), 167. Italics added.

[187] See Audi (2008), 478f. In this passage, he presents this view as an alternative to the "self-evidence as paradigmatic reasons"-view, but expresses preference for the first. The reason Audi argues for the viability of both views is that the second "paradigmatic reason"-view is "usable even by those who prefer to avoid positing abstract entities." In his (2013) which discusses moral perception, Audi compares moral intuitions with the act of seeing; the objects perceived in intuitions are "both abstract and in some way connected with linguistic usage, and they are surely accessible through understanding that usage;" Audi (2013), 88. Even though this passage does not directly address the issues of defining self-evidence, it illustrates his view that understanding moral propositions brings us in contact with abstract entities. Audi makes a similar point at (2013), 97.

The advantage of this theory is that – other than the paradigmatic-reason view – it offers resources for those who question that self-evident propositions are justified without further justification: by adequately understanding self-evident propositions, we grasp abstract entities. Not even a skeptic could ask for more.

As it stands, it is however hard to evaluate this view for the lack of a detailed explanation. How are those "abstract elements" or "entities" to be understood? Why is it that we can connect to them by understanding self-evident propositions? Recently, Terence Cuneo and Russ Shafer-Landau have further developed this line of thought. I will first briefly show that their theory shares many assumptions with Audi's view and then discuss how their contribution advances the discussion. Cuneo and Shafer-Landau's starting point is the assumption that there are a number of basic moral propositions ("moral fixed points") that are true in virtue of the concepts they are composed of.[188] There are four marks of those conceptual truths:[189] first, they are necessarily true. Second, the moral fixed points specify fundamental moral considerations and are in this respect similar to Ross' *prima facie* duties. Third, the denial of a proposition of this sort "would tend to evoke bewilderment among those competent with its constituent concepts – a response to the effect that its denial would be almost crazy."[190] Fourth, mature moral agents acquire knowledge of the moral fixed

---

[188] See Cuneo & Shafer-Landau (2014), 5.
[189] As Audi, Cuneo & Shafer-Landau do not believe that moral propositions are analytic. See their (2014), 12f.
[190] Cuneo & Shafer-Landau (2014), 9f.

points by adequately understanding them. Except for the first, these four criteria

bear strong similarity to Audi's theory.[191]

What Cuneo and Shafer-Landau add to the discussion is an account of the

nature of the moral fixed points and of the concepts they are composed of. In

regards to those concepts, Cuneo and Shafer-Landau adopt a view that is informed

by Frege. Three features characterize those concepts. First, they are "abstract,

shareable, mind-independent ways of thinking about objects and their

properties."[192] Hence, concepts are objective and exist independently of our

linguistic conventions. Second, propositions are constituted by concepts.[193] Third,

concepts have essences and "determine the range of entities that they are about."[194]

For example, moral terms such as "being wrong" can apply only to those things

that are objectively wrong:

> To see how this works for the moral fixed points, consider the
> proposition <that recreational slaughter of a fellow person is
> wrong>. This is a conceptual truth in case it belongs to the essence
> of the concept 'being wrong' that, necessarily, if anything satisfies
> 'recreational slaughter' […] it also satisfies 'being wrong' (in a
> world sufficiently similar to ours).[195]

Specifying the nature of the concepts that compose fundamental moral truths

(whether they are called "moral fixed points" or "*prima facie* duties") helps to

understand why agents who properly grasp those propositions are justified in

believing them. It addresses the worries of those who were skeptical about the

---

[191] As discussed above, it can be questioned whether Audi shouldn't adopt a view according to which self-evident moral propositions hold with necessity.
[192] Cuneo & Shafer-Landau (2014), 11.
[193] This contrasts with the view that propositions are constituted by the properties themselves.
[194] Cuneo & Shafer-Landau (2014), 12.
[195] Cuneo & Shafer-Landau (2014), 12.

paradigmatic reason account and its lack of explanation regarding the justificatory force of self-evident propositions. At the same time that Cuneo and Shafer-Landau help to develop Audi's original position, they also leave important questions open. Even if we accept the existence of concepts as objective and extra-mental items as Cuneo and Shafer-Landau suggest (they do not offer separate arguments for those claims), it is still not clear how human agents connect to them through their judgments about what seems self-evident. There is also no explanation of why the fact that rejecting moral fixed points causes bewilderment among most mature agents shows that we are justified in believing them.

Many philosophers will not see the need for such a robust account of justification for our moral judgments. Moral realists however – and it is important to keep in mind that traditionally, most intuitionists have subscribed to that view – are looking for answers to the questions raised in this section. For them, the attractiveness of the views of Audi, Cuneo and Shafer-Landau depends on the existence of such an account.

3.3. Conclusion

Audi's strategy is to present a view that carries less theoretical baggage but achieves the same goal as traditional intuitionists, namely showing how belief in moral principles expressing our *prima facie* duties can be justified without relying on external premises. As seen, Audi's view has many advantages. It offers a more detailed account of self-evidence and is minimal in its commitments.

      Despite its merits, Audi's theory lacks a fully worked out explanation of the connection between adequately understanding a self-evident propositions and being justified in believing it. The next two chapters attempt to address that challenge.

Chapter 4: Ethical Intuitionism and the Linguistic Analogy

4.1 Introductory Remarks

A moral epistemology which accounts for the possibility that we grasp fundamental moral principles in a non-inferential way is central to ethical intuitionism but yet, as argued in the preceding two chapters, the most prominent attempt to formulate such a view face problems. There are two strategies in response to this challenge: improving on existing accounts or developing an alternative theory. In this and the next chapter, I follow the second route and argue for an intuitionist moral epistemology which does not – as traditional intuitionist theories do – rely on the idea of self-evidence. This, however, is not meant to imply that self-evidence based accounts couldn't be fixed; instead, there might be more than one epistemological account supporting intuitionism.

I proceed by drawing an analogy between our ability to understand language and the intuitive judgments we make in response to moral problems. To make the case that our understanding of moral reasoning can be informed by linguistic theory, a number of steps have to be taken. First, I argue that there are two dimensions of a complete moral epistemology (4.2): an account that explains the content of our moral beliefs and a theory of justification. Then, forestalling some objections, I discuss how even within the descriptive project, empirical findings can be rejected as false (4.3) and I also consider whether reliance on people's judgment is necessarily conservative (4.4). After those preliminary remarks, I formulate what results should be expected if Rossian intuitionism is descriptively accurate (4.5). In order test those predictions, I turn to John Rawls for the Rawls'

idea that developments in modern linguistics can be used in order to offer an empirical model of moral psychology. After introducing some arguments from Noam Chomsky's theory of generative grammar (4.6),[196] I discuss in more detail how John Mikhail's work on moral grammar develops Rawls' idea (4.7-4.10). As it turns out, the empirical predictions that are implicit in Ross theory are confirmed to a large degree. In (4.11), I analyze some disanalogies between linguistics and moral psychology and in (4.12), I focus on how Mikhail's results compare to Ross's picture of moral deliberation. It turns out that there is an important omission in the empirical literature that involves situations where duties collide. I conclude that a more complete empirical framework should take seriously Ross suggestion that moral principles are to be understood as *prima facie* duties (4.13).

The first step towards such a new theory consists in asking what to expect from an intuitionist account of moral knowledge. This will enable us to disentangle various questions, to see their connections and to answer them separately. A feature that sets intuitionism apart from competing theories is that we discover the content of normative principles not by derivation from other considerations, but by reflection on our well-considered judgments. Assuming (as Rossian intuitionists do) that mature agents can acquire moral knowledge via this process, our moral faculties play a central role for the overall theory: they are both the source and

---

[196] There are important limitations to my exposition. First, I do not pretend to offer a comprehensive survey of modern linguistics; at the opposite, since my focus is a descriptive account of moral knowledge, I confine myself to a minimal number of distinctions and arguments in linguistics necessary to understand the context of recent work in empirical moral psychology. Further, my summary relies on the linguistic accounts of Mikhail and Hauser (See Mikhail (2011), 14-40 and Hauser 2006.) I am also assuming that Chomskian theory is the correct approach to linguistics and ignore criticisms against his accounts or competing theories. It follows that if Chomskian linguistics turns out to be flawed, this would significantly weaken my overall argument.

justification of moral knowledge. As seen in the previous two chapters, the notion of self-evidence ties those two claims together. In a nutshell, the criterion for admission to the list of *prima facie* duties is that the principle in question is self-evident. At the same time, those beliefs are justified in virtue of the fact that they are self-evident. Once the notion of self-evidence has been cast into doubt, the possibility needs to be considered that those two functions of moral cognition (providing access to moral principles and justifying them) come apart. It is one question to ask what principles the judgments of mature people rely on, and a second issue what justifies reliance on those principles. The answers to these two problems will be different because the first issue, to which I will refer as *descriptive adequacy*, requires an answer located in the field of empirical moral psychology that takes into account people's actual well considered beliefs, whereas the second question addresses the *justificatory adequacy* of the moral knowledge employed by those agents and calls for a conceptual, normative account.[197]

Before accepting this picture, a temptation to dismiss the problem of empirical adequacy as irrelevant has to be dealt with. Ross himself was doubtful about a school of sociologists who "seek to replace moral philosophy by the 'science of mœurs', the historical and comparative study of the moral beliefs and practices of mankind."[198] The problem Ross sees with this form of inquiry is that it reduces moral beliefs to societal and historical circumstances, suggests a skeptical attitude towards them and ignores that moral beliefs have truth-values.[199] Rather,

---

[197] Those terms are borrowed from Mikhail (2011).
[198] Ross (1930), 13.
[199] See Ross (1930), 13f.

he argues that the human mind has "in fact an insight into certain broad principles of morality, and it can distinguish between a more and a less adequate recognition of these principles."[200] Does this not suggest that an empirical account of the principles that govern our moral decision-making bears little relevance on intuitionist epistemology which is focused on an analysis of how our conceptual faculties perceive moral facts? In response, descriptive adequacy does not aspire to replace moral theory, to reduce it to its genealogy or to deny that our moral faculty is able to grasp mind-independent principles; it merely addresses an aspect that Ross himself thought was an essential part of his project. Throughout, he seeks to establish the superiority of his own view against his competitors by showing that it aligns closer with common sense, to what "seems clear," "is plain on reflection," "conforming to what most men (if not all) believe" etc.[201] In one of the most frequently quoted passages by Ross, he claims that "it is more important that our theory fit the facts than that it be simple," and for him, the way to find out about those facts is to consult "what we really think."[202] The reason that appeal to common sense carries evidentiary weight is that our convictions serve as "the data of ethics just as sense-perceptions are the data of a natural science."[203] It is this link between our well-considered judgments and the truth of the matter that justifies an investigation into the actual moral beliefs people hold. This makes it

---

[200] Ross (1930), 14.
[201] See among other references to his (1930), 2, 6, 8f, 11, 17, 18 and especially 39f.
[202] Ross (1930), 30.
[203] Ross (1930), 41.

clear that for Ross, there is an important connection between his normative ethical theory and an account of the principles that in fact govern our moral judgments.[204]

Apart from textual evidence found in Ross, there are independent reasons to accept descriptive adequacy alongside justificatory adequacy as part of a complete intuitionist moral epistemology. For intuitionists to present a unified ethical theory, it is not enough to show that mature agents possess a moral faculty that provides foundational insight into basic moral principles. On the other side, it is insufficient to offer a list of *prima facie* duties without explaining what makes their possessors justified in believing them. In order to avoid a theory that is either empty or lacks justification, the intuitionist needs to show (a) that moral faculties of mature agents yield knowledge and (b) what this knowledge consists in and how it comes about. Assuming the first task can be resolved, the well-considered judgments of mature agents are trustworthy indicators of moral truth.

If empirical moral psychology showed conclusively that under ideal situations, the judgments of educated agents reflected a system of moral principles that contradict the list of *prima facie* duties assumed by intuitionism (for example by rejecting some elements on his list or by accepting new principles such as the doctrine of double effect), there would be reason to change the normative theory accordingly.[205] The fact that traditional intuitionists have made their assumptions

---

[204] As is apparent from the discussion in chapter 2, intuitionists like Ross are not interested in moral convictions regarding concrete cases which involve the weighing of various ethically relevant considerations and judgments about contingent circumstances. Rather, he is concerned with the underlying structure of those decisions, analyzing the fundamental ethically relevant considerations that determine the rightness or wrongness of acts across situations. As has also been shown, Ross does not give equal credence to all intuitions. Only considered judgments by mature moral agents count.

[205] As a matter of fact, Ross explicitly allows for the possibility that there are *prima facie* duties that he has overseen. He says about his list that it is "*a prima facie classification of the duties which reflection on our*

about the contents of common sense beliefs from the armchair shouldn't indicate

that empirical considerations are alien to the intuitionist' project. Collecting

empirical data about the rules underlying moral judgments and extracting reliable

information from them poses a methodological challenge that hadn't been solved

in the days of Sidgwick, Ross or Prichard. At the time, relying on introspection

was the only available means to stipulate what common sense beliefs. If progress

in the social sciences has resulted in more reliable techniques, the intuitionist

should be happy to embrace them.[206]

4.2 Discounting flawed intuitions

Even though empirical data should – if the foregoing argument is correct –

play a role in determining the content of the moral principles intuitionism adopts, it

does not follow that every common sense belief has to be taken at face value. If

common sense adopts under idealized conditions incoherent beliefs, those mistakes

should be corrected. Ross himself, for example, notes ambiguities in the common

usage of the concept of right and sides with the interpretation that makes most

sense to him.[207] More recently, Joshua Knobe has demonstrated that for most

people, the perceived moral quality of the side-effects of an action affects whether

---

moral convictions seems actually to reveal. And if these convictions are, as I would claim that they are, of the nature of knowledge, and I have not misstated them, the list will be a list of authentic conditional duties, correct as far as it goes though not necessarily complete. The list of *goods* put forward by the rival theory is reached by exactly the same method – the only sound one in the circumstances – viz. *that of direct reflection on what we really think.*" Ross (1930), 23 (italics added).

[206] Railton (1986), 206 [quoted from Mikhail (2011), 314] and Zimmerman (2010), 96 also argue that the ethical intuitionist needs to provide an epistemological account which is "psychologically plausible" (Brandt, 1979) (Brandt, 1979).

[207] See Ross (1930), 2f. where he laments that some people assume that it is morally obligatory to act from a certain motive because they "fail to notice the distinction between 'right' and 'morally good'." He then argues that proper reflection on the meaning of those terms reveals their difference.

it is believed that those side-effects have been caused intentionally.[208] Insofar as it

is clear that the moral quality of a side-effect has nothing to do with the issue of

whether the side-effect was intended or not, we are justified in discounting those

beliefs.[209] Cases like those do not undermine the general Rossian claim that our

reflected judgments about the fundamental morally relevant considerations are

veridical, but single out specific cases where our reasoning abilities are led

astray.[210] From the standpoint of Rossian intuitionism, those exceptions are

nothing to be surprised about. The intuitions that Ross takes as inputs into his

theory stand in the middle between two extremes. On the one side, there are the

moral judgments of idealized abstract agents whose perfect moral reasoning

powers always get it right and on the other side there are the unfiltered opinions of

average citizens. Ross is interested in the judgments of mature agents that have

reflected on their beliefs, people whose moral insights is to be trusted, but who are

also not immune from the systematic irrationalities that form part of our

psychological setup. The difficult task in the search of descriptive adequacy will be

to replicate those conditions in experimental settings.

As a clarificatory note, when stating that empirical data should be taken into

account in developing a descriptively adequate moral epistemology, this does not

mean that demands for the reformulation of our list of duties can be derived

---

[208] See Knobe (2003).

[209] The example is non-moral, but moral cases are easy to come by. If, for example, it should turn out that the doing/allowing distinction people rely on when judging that switching a trolley from one track to another in order to save five workers while causing the death of one other worker is morally allowed while pushing a large man from a bridge in order to save five workers is not allowed are incoherent or based on confusion, those intuitions should also be discarded in the construction of an intuitionist moral theory.

[210] This assumes that absent good reasons that undermine an intuition, our common sense beliefs are justificatory (see next chapter).

directly from statistical results gained from the responses of certain agents. Since

the object of inquiry is the set of principles that underlies our moral judgments and

since judgments about concrete moral problems usually involve the weighing of a

number of morally relevant considerations, those judgments need to be interpreted

in order to extract the various *prima facie* duties that are involved. In other words,

what is of interest lies below the surface of everyday moral judgments. There will

be more than one way of interpreting people's intuitions about different cases, and

the challenge is to make an inference to the best explanation that can make sense

of the variety of intuitions people have. Hence, empirical data are only one part of

the task of empirical adequacy.

4.3 Descriptive adequacy and moral criticism

Connected to the intuitionist reliance on common sense comes another

challenge. Theories such as utilitarianism are able to criticize popular moral beliefs

in case they diverge from the utilitarian calculus. Ethical intuitionism on the other

hand has no independent criterion of moral rightness, and it therefore lacks the

ability to distance itself from common sense morality and is, so the criticism goes,

a prisoner of the *status quo*; it holds no promise of moral improvement over

present practices. In defense against this conservatism-worry,[211] intuitionists can

argue that even though they defend a reconstructive program which presupposes

---

[211] This is one way of spelling out the problem. See Nelson (1990) for alternative readings of the conservatism-worry. The present understanding of the issue has also been shared by Brandt and others. Brandt (1979) for example writes: "What we should aim to do is step outside our own tradition somehow, see it from the outside, and evaluate it, separating what is only the vestige of a possibly once useful moral tradition from what is justifiable at present. The method of intuitions in principle prohibits our doing this." (quoted from Nelson (1990), 287).

that the underlying moral rules people employ are correct, the theory does not contradict the best available explanations of moral progress or strong criticisms of political injustices. The reason lies in the limited project of Rossian intuitionism: the view tries to locate and to defend the most fundamental morally relevant considerations and has little to say about all-things considered judgments. The fact that the intuitionist tradition has no concrete insight into pressing issues such as how to treat animals, the permissibility of abortion or the rights of future generations does not mean that the theory supports the presently accepted view. Rather, their evaluation fall outside the scope of the theory. The reasons for this is that – as will be discussed below – in line with Ross' arguments, it can be argued that there is no way to codify in abstraction from the indefinite number of possible cases how various morally relevant considerations are going to interact with each other. Hence, whenever there is a conflict of duties (as with most applied ethical issues), the intuitionist will allow for moral judgment in order to determine how those duties combine in concrete instances.[212] Those judgments can then be criticized for failing to take certain considerations into account, for exposing inconsistencies with other beliefs etc. In conclusion, the charge of conservatism misunderstands the goal of intuitionism which is merely to identify basic moral principles; an intuitionist can still question others for failing to apply them well.

---

[212] Beauchamp and Childress have developed such a theory for biomedical ethics where they specify what rules might be relevant for an issue. See Beauchamp and Childress (2012)

4.4 What empirical claims does Rossian intuitionism make?

There are three possible outcomes from an empirical study into the principles that underlie the common sense judgments of mature agents. First, it is possible that there is no satisfactory answer to the problem: no universally shared set of rules can be found. This would weaken the intuitionist assumption that there is one rule-system on which agents rely in their moral judgments. A second, less pessimistic outcome would still be a burden for intuitionism: there is a list of *prima facie* duties which is stable across various cultures, but it differs significantly from the classical intuitionist picture. This would allow the intuitionist to correct her normative account, but might also undermine many of Ross arguments against his utilitarian competitors which often rely on claims regarding "what is firmly believed" by common sense. Third, a review of evidence collected by empirical moral psychology might also confirm many of Ross' armchair intuitions and consequently strengthen his theory.

In order to find out, we first need to formulate what results should be expected if Ross' assumptions concerning common sense morality turned out to be descriptively adequate. This will later help to analyze the empirically supported evidence and to judge which of the three possibilities just outlined is the most plausible. Here is a list of expectations that can be extracted from Ross' theory:

1. Despite disagreements about concrete issues, mature moral agents at least implicitly accept the same fundamental morally relevant features and apply them in their deliberations.

2. There is a plurality of basic morally relevant features, i.e. the features cannot be further reduced to a single unifying principle.

3. Those features can be expressed in the form of *prima facie* principles: the moral rules determine whether the morally relevant features count in favor or against an act, but not how various principles interact with each other.

4. The *prima facie* principles that underlie the judgments of morally mature agents bear resemblance to the items on Ross' list of seven *prima facie* duties (fidelity, reparation, gratitude, non-maleficence, justice, beneficence and self-improvement); minimally, the rules that actual mature agents follow should contain both consequentialist and deontological principles.

5. Mature agents (a) accept that the *prima facie* duties are not in need of further justification and (b) are strongly reluctant to give them up.

More items could be placed on this list, but those mentioned here capture the central expectations that a Rossian intuitionist should have in regards to the outcomes of an empirical study into people's moral decision making. Given the introspective method which traditional intuitionists have used, it would be surprising if all of their predictions would be vindicated from an empirically informed point of view. More realistically, even a partial match between their assumptions and an empirically adequate picture would testify to their highly refined sense of moral imagination.

4.5 Chomskian linguistics

Chomskian linguistics is the attempt to answer three distinct questions.[213] The first question asks for an account of what it means to know a language. In response, Chomsky offers a *Generative Grammar*, a set of rules possessed by competent speakers which allows judging whether a given sentence in a language is grammatical. A second question asks how this knowledge is acquired. Here, Chomsky's solution is given in terms of a theory of *Universal Grammar* which claims that language acquisition results from the interplay between a genetically endowed language faculty and external stimuli. Since the genetic component is the same for all normal human beings, all languages share certain structural features. Third, Chomsky is concerned with how the knowledge of language is being put to use. This problem is addressed in terms of a *Theory of Linguistic Performance*. For present purposes, only the first two questions are of relevance. Before turning to the arguments in support of a *Generative Grammar* and a *Universal Grammar*, two distinctions need to be introduced. The first differentiates between competence and performance.[214] *Competence* refers to the knowledge a speakers needs to have in order to master e.g. a specific language; this knowledge, Chomsky argues, is unconscious and inaccessible.[215] *Performance*, on the other hand, describes a speaker's actual linguistic behavior which is influenced by her knowledge, but also by other components such as "her memory structure, mode of organizing experience, perceptual mechanisms and attention span, and a wide range of

---

[213] See Mikhail (2011), 14f.
[214] See Mikhail (2011), 17-19.
[215] See Hauser et al. (2008), 110.

additional factors."[216] *Generative Grammar* is only about a speaker's competence, the cognitive system that contains the rules of language; it abstracts from all additional components that also affect linguistic behavior.[217] The second distinction concerns two kinds of principles explaining our judgments about language.[218] *Operative principles* are those principles that explain our language use. They do not need to be accessible by the speaker through introspection. Therefore, a competent language speaker can – and, as Chomsky argues, in fact does – make mistaken assumptions (*express principles*) about what principles govern the exercise of her abilities to speak a language. Generative grammar is about the operative, not the express principles.

What arguments does Chomsky use to establish his theory of Generative and Universal Grammar? The theory of Generative Grammar receives support from the *argument for linguistic grammar*, whereas the *poverty of stimulus argument* is supposed to establish the Universal Grammar. Both arguments are abductive, offering an inference to the best explanation; in principle, further evidence could prove them false. Let's first turn to the argument for linguistic grammar.[219] It starts with the observation that competent speakers have encountered throughout their lives only a limited set of linguistic utterances. However, based on this finite experience, they are able (a) to produce grammatically correct novel utterances that are unbounded in scope and (b) to

---

[216] Mikhail (2011), 18.
[217] This puts Chomsky's theory in contrast to behaviorism which is unable to distinguish such a cognitive component from other factors that also contribute to actual behavior.
[218] See Mikhail (2011), 19-21.
[219] See Mikhail (2011), 43-46.

judge the grammatical correctness of utterances from an infinitely large set (such

as in Chomsky's famous example "colorless green ideas sleep furiously")[220]. At

the same time, competent speakers are (unless they are linguists) unable to produce

the complicated grammatical rules which explain their judgments and utterances.

How can this linguistic knowledge be accounted for? One solution would be that a

list with all the possible utterances is stored in the brain, but given our limited

cognitive capacities, this is not plausible. A better explanation is that our mind

possesses a system of rules whose application unconsciously renders our linguistic

judgments and helps us form new grammatical statements. These operative

principles form the linguistic grammar. This theory would gain further plausibility

if we also had an answer to the second question outlined above: how do we acquire

such a rule system? The argument from poverty of stimulus starts with the

observation that even young children obey complex linguistic rules.[221] At the same

time, the stimuli that children receive are – as Chomsky argues – both degenerate

and impoverished. The input is degenerate since the child is faced with expressions

that are not well-formed, "including false starts, slips of the tongue, changes of

plan in mid-course, endings that do not match the beginnings, and other similar

departures from well-formedness."[222] Stimulus is impoverished insofar as it leaves

out evidence that would be necessary for understanding certain rules that children

nevertheless pick up. How can it be explained that despite these shortcomings,

---

[220] Chomsky (1957), 15.
[221] Here are two examples: Principles for verb contraction indicate that it is appropriate to say "Frank's foolish" instead of "Frank is foolish," but that we are not allowed to say: "Frank is more foolish than Joe's" (see Hauser et al. (2008), 111f). Another example is mentioned by Mikhail (2011), 70: The structure-dependence of rules regulates the sequence of words in well-formed sentences in a highly complex way.
[222] Mikhail (2011), 70.

children still turn into competent language users? The Universal Grammar, a genetically endowed rule system in combination with an appropriate (even if degenerate) external input enables language acquisition. Since the child's learning of a language is partly determined by those innate principles, the range of possible languages that it can acquire is limited.[223]

Both arguments are independent but not unrelated. The argument for linguistic grammar establishes the existence of a system of linguistic principles whose development is explained by the theory of Universal Grammar. For our purposes, it is important to stress that it is possible to develop a linguistic grammar without addressing how we came to learn it.

Chomsky's account, although it was originally developed as a linguistic theory, has broadened into a research program that has fruitfully been applied to a number of disciplines. Unconscious principles are postulated in order to understand our knowledge in areas such as mathematics, music and object perception but also our ability to attribute intentional states to others or to classify animals and plants.[224] This invites the question whether the same arguments can also provide an explanation of moral judgments.

4.6 Mikhail's linguistic analogy

The idea that the principles that guide our moral intuitions are structurally similar to the rules of grammar predates modern linguistics. Adam Smith, for

---

[223] See Hauser et al. (2008), 112.
[224] For a list of literature, see Hauser (2008), 109, Sripada (2008), 319 and Mikhail (2011).

example, wrote in an often quoted passage from *The Theory of Moral Sentiments*

that

> The rules of justice may be compared to the rules of grammar; the
> rules of the other virtues, to the rules which critics lay down for the
> attainment of what is sublime and elegant in composition. […] A
> man may learn to write grammatically by rule, with the most
> absolute infallibility; and so, perhaps, he may be taught to act
> justly.[225]

Although the general belief in the analogy between the spheres of language and

morality has been around for a long time, it is only after the Chomskian turn in

linguistics that philosophers have started to pursue the project in detail. John Rawls

deserves credit for being the first to outline what has come to be called the

"linguistic analogy." In his *Theory of Justice*, he argues that many of Chomsky's

arguments and distinctions – e.g. the argument for linguistic grammar, the

distinction between operative and express principles and between operative and

express principles – can be used to characterize our "sense of justice."[226] Since

Rawls' remarks are terse and take many assumptions for granted,[227] I will not

engage in an interpretation of his text, but turn instead to more recent defenders of

the linguistic analogy.[228] Initially, Rawls' remarks in paragraph 9 of his *Theory of*

Justice didn't receive the attention they deserve, but over the last decade, more

theorists have taken inspiration from Rawls' application of Chomsky's theory to

---

[225] Smith (1817/2002), 205. The quote (or parts of it) is used by Harman & Roeder (forthcoming), Hauser et al (2008), 107 and Mikhail (2011), 1. I will return to Smith's argument below.
[226] See Rawls (1971), 40-42.
[227] At places, his argument for the linguistic analogy is not only terse, but also incomplete. See for example Mikhail's reconstruction of Rawls' argument for moral grammar in Mikhail (2011), 46-48.
[228] Besides its shortness, it is not obvious how Rawls' discussion is to be interpreted. Assuming he meant to defend a 'linguistic analogy', how does this fit with his discussion of the reflective equilibrium or with his remarks on moral psychology in part III of the *Theory of Justice*? It seems advisable to turn to self-standing accounts of the moral analogy that don't raise further complications.

morality.[229] Mikhail's *Elements of Moral Cognition – Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment* offers the most comprehensive and detailed account. He advances the linguistic analogy in two respects: by spelling out in detail Rawls' suggested application of Chomsky's arguments to morality and by conducting empirical studies in moral psychology with the aim of showing the structural similarities between linguistics and morality and in order to unveil the main traits of our 'moral grammar'.[230]

Of the three questions that Chomskian linguistics can be divided into, the linguistic analogy primarily addresses the first issue about the rule system that underlies our moral judgments. The central idea is that

> the normal individual's moral knowledge consists in part in […] a complex and largely unconscious system of moral rules, concepts, and principles that generates and relates mental representations of various types.[231]

Officially, the theory[232] remains neutral regarding the question of nativism which would be addressed by a theory of Universal Moral Grammar. At the present stage, this is a necessary limitation. On the one hand, in case the results of a nativist account of our moral faculties coincided with the conclusions drawn from the analysis of people's well-considered judgments about moral problems, this ontogenetic explanation could offer additional support for a solution to the

---

[229] See for example Harman & Roeder (forthcoming), Hauser (2006) & Hauser at al (2008), Mikhail (2007), (2008) and (2011) and Dwyer (2006).
[230] See Mikhail (2011), 3, 11 and 308.
[231] Mikhail (2011), 16.
[232] This is at least the scope of the version of the linguistic analogy that I am considering. For a more ambitious interpretation of the linguistic analogy that includes ontogenetic questions, see Hauser (2006) and Hauser et al. (2008).

descriptive problem.[233] On the other side, as of now, research on moral nativism is less advanced[234] than empirical studies on the rules underlying moral decision making and it seems advisable not to built one's theory on empirical research that is still in early stages and where there is little consent.[235] It is – as demonstrated below – possible to argue that there is a moral analogue to generative grammar without also having to take a stance on whether there is a moral analogue to universal grammar:[236] for descriptive adequacy as defined above, it is sufficient to give an empirically defensible account of the rules that guide the moral judgments of mature agents.[237] For present purposes, it seems therefore best to focus on developing a moral grammar and to remain officially neutral on questions about nativism.[238]

---

[233] Support goes into the other direction as well. If it can be established that there is a moral grammar that is not part of the conscious knowledge of competent moral agents, this would add plausibility (even though it would not offer conclusive evidence) to an explanation of the origins of our moral faculty which appealed to nativism (and not to explicit instruction, for example, which would let us expect that we have explicit awareness of the moral rules). See also Mikhail (2011), 120f and 179.

[234] Besides the lack of empirical data, the terms of the debate are also not settled yet. As Joyce (2013) convincingly shows, the present discussion regarding nativism suffers from conceptual ambiguities about what is meant by the term "innateness": some assume a evolutionary conception according to which, as Prinz ((2009), 168, quoted from Joyce (2013), 551) puts it, "morality is an evolved capacity;" others adopt a developmental conception which holds that "the emergence of the trait is buffered against variation in the developmental environment" (Joyce (2013), 552). Joyce argues that the two understandings are however not co-extensive: some traits might be "a specific adaptation but may nonetheless require particular structured environmental inputs in order to become manifest. […] The reverse is also true: Human morality may be developmentally canalized while not being an adaptation" (553).

[235] Even if it can be shown that there is a close analogy between moral and linguistic grammars (more on that topic below), it cannot be inferred from this fact that the analogy also extends to the question of nativism.

[236] See James (2011), 110-112 and Mikhail (2011), 35f. and 103f.

[237] Ross, for example, was concerned about our moral faculties which, he thought, could recognize an objective moral truth; he was not interested in an explanation of how this faculty had developed (other than making general remarks that morally competent agents have to be mature) and whether it was native in any sense of the word. However, as will become clear below, the origins of our moral faculties will play a role in the justificatory part of intuitionist moral epistemology.

[238] There is a further preliminary question to be addressed: how tight is the analogy between moral and linguistic grammar to be construed? I will leave this question open for now and discuss it after presenting the arguments for the linguistic grammar. This will put us in a better position to see in which aspects the analogy should be rejected.

4.7 The moral grammar hypothesis as a possible solution to the problem of descriptive adequacy

The goal of the linguistic analogy is to apply to the moral realm the same arguments that Chomsky uses in his argument for the linguistic grammar; as in the linguistic case, the aim is to (a) make the case for and (b) give a substantive account of the rule system that guides our moral judgments. Whereas in the case of linguistics, the "unit of appraisal"[239] is judgments about well-formed sentences, the moral philosopher relies on people's judgments about moral problems. Again in close analogy to modern linguistics, two distinctions have to be introduced in order to describe where the rules system is located.[240] The first distinction describes the difference between moral performance and competence. Often, moral judgments are affected by factors such as fear, mistaken beliefs, time pressure etc. In order to isolate our moral grammar from those distorting influences, moral psychologists try to eliminate this "noise" by eliciting responses under ideal circumstances. The well-considered judgments which result from this process of idealization are the basis from which the moral grammar can be derived. The second distinction differentiates operative from express principles. After having isolated a subset of undistorted moral judgments, there are two ways to uncover the rule system that explains them. First, we could ask subjects to give reasons that justify their responses. This is how Kohlberg and Piaget conducted their studies.[241] A second option consists in ignoring those self-reports and in investigating independently which operative moral principles have to be stipulated as (possibly unconscious)

---

[239] Hauser et al (2008), 116.
[240] See Mikhail (2011), 50-55.
[241] See Kohlberg (1981) and (1984); Piaget (1932/1964).

knowledge in the agent.[242] There is good reason to prefer operative over express

principles. As will become clear in the discussion of Mikhail's experiments below,

agents are often mistaken about the actual principles that guide their moral

judgments.[243]

The argument for moral grammar starts parallel to the argument for

linguistic grammar with the observation that mature moral agents are able to judge

the deontic status of a potentially unlimited number of actions. At the same time,

those agents have only been exposed to a finite number of cases, creating a

"projection problem":[244] how can this finite experience – presented as a collection

of individual judgments – be projected onto an infinite number of situations and

offer a solution for them? The conclusion that Chomsky draws from this problem

and which also applies to the moral case is that the knowledge of how to make

judgments about instances of the domain in question cannot consist of a list of

cases agents have been exposed to.[245] As an alternative explanation to the

---

[242] As discussed below, the distinction between operative and express principles might be given a stronger interpretation in linguistics than in moral psychology.

[243] Mikhail argues that Kohlberg's reliance on self-reported justifications and his emphasis on their variability is unable to account for a number of features that occur in Mikhail's own extensive empirical studies: "Unlike Kohlberg's (1981, 1984) dilemmas, the moral judgments occasioned by the trolley problems […] are spontaneous, stable, rapid, intuitive, involuntary, and stringent, and made with a high degree of certitude […]. Moreover, the judgments appear to be widely shared among demographically diverse populations, including young children; even in large cross-cultural samples, participants' responses to these problems cannot be predicted by variables such as age, sex, race, religion or education. Furthermore, individuals typically have difficulty producing compelling justifications for these judgments; trolley problem intuitions thus exhibit an apparent dissociation between judgments and justifications, thereby illustrating the distinction between operative and express principles." Mikhail (2011), 309. It is important to notice that Mikhail is not criticizing Kohlberg and Piaget for collecting people's beliefs about their actions, but for giving those beliefs a privileged status over other kinds of data such as the rules that can be inferred from people's responses (and which might contradict people's self-ascribed justifications). See especially Mikhail (2013), 40, but also Hauser et al. (2008), 114 and 125.

[244] See Chomsky (1957), 15, quoted from Mikhail (2011), 47.

[245] See Mikhail (2011), 46f and 94-96. Neither is it plausible to assume that we are genetically endowed with a list of all possible cases since the brain's cognitive capacity is limited.

projection problem, people could in the course of their upbringing either directly be taught a set of basic moral rules or be shown paradigmatic cases and then infer moral rules from those examples[246] or model their judgments after the paradigmatic cases (which cover most ground and can usually be compared to any number of new cases). This would explain how they are able to address any novel moral problem. However, if our moral judgments relied on principles that are derived from a process of deliberation or from explicit moral teaching, we should be conscious about their content. As argued in the discussion about express versus operative principles, defenders of the moral grammar hypothesis (which argue that – as in the linguistic case – the principles are not usually consciously known to the agent) point out that people's self-reported justifications for their actions turn up principles that are not only too coarse-grained to account for people's changing judgments when the scenarios they are asked to assess are changed in subtle ways, those self-reported principles also fail to predict how people are going to judge related cases. If, as the moral grammar hypothesis assumes, an alternative solution exists which does not have this problem, we should prefer it.

The projection problem relies on an unstated premise which is more in need of justification in the case of morality than it is in linguistics: the judgments under scrutiny need to follow a recognizable pattern and be analyzable in terms of common principles. If each judgment were to follow its own rules (or no rules at all), there would be no projection problem. In the case of language, the very fact that we are able to communicate demonstrates that this is not the case; if each

---

[246] This is the model that Ross was trying to articulate in his discussion of intuitive induction (see chapter 2).

speaker decided on a case-to-case basis what constituted a grammatical expression,

people would have no means to interpret each other's signs. The same does not

follow for the case of morality: it might be the case that we solve ethical problems

without reliance on a system of principles. Even if we sometimes compare our

judgments to other cases, this does not imply that our judgments follow a

comprehensible pattern. While this scenario describes a logical possibility, Mikhail

argues in his discussion of moral particularism that it is false on empirical grounds:

> The novelty and unboundedness of moral judgment, together with its
> frequent predictability, stability, and other systematic properties,
> implies the existence of a moral grammar, because to explain how an
> individual is able to project her finite experience to entirely new
> cases, we must assume that she is guided, implicitly, by a system of
> principles or rules.[247]

If the negative lesson from the argument for moral grammar excludes our limited

past experience, a genetically endowed list of cases, explicit teachings of rules and

the deliberate derivation of principles from examples as explanations of our ability

to judge an unbounded set of novel moral problems, it serves on the positive side

as test which can identify a solution that avoids the shortcomings of those theories.

---

[247] Mikhail (2011), 72. His argument against particularism is flawed. Particularists like Jonathan Dancy (the only representative of the view that Mikhail considers) do not attempt to describe people's actual moral judgments. See e.g. Dancy (2004), 1-3 where he explicitly acknowledges and rejects (on philosophical grounds) the fact that people's judgments rely on principles. Here is how he characterizes his theory: "There are two sorts of approach that a particularist can take in trying to make a persuasive case against any essential link between morality and principles. The first is to try to show that no suggested principles are anything like flexible enough to cover the ground and do the job we require of them. Moral life, it can be said, is just too messy, and the situations we encounter differ from each other in subtle ways that no panoply of principles could ever manage to capture. Principles deal in samenesses, and there just aren't enough samenesses to go round. I think that this is true, but it is not the way in which I tend to argue […]. The book I have not written would really be an investigation of the subtleties of our moral thought and the actual complexities of life. The book I have written is about how to understand the way in which reasons work, and deals largely with *theories* about reasons rather than with life." Dancy (2004), 2.
Hence, Mikhail's argument that Dancy's theory does not fit with people's actual moral judgments misses the mark. However, this poses no threat to the theory of moral grammar: both theories play on different fields.

A moral grammar, Rawls, Mikhail, Hauser and others have suggested, passes this test and solves the projection problem. A limited set of deontic rules combined with a theory of structural act descriptions and an account of how "the stimulus is converted into an appropriate structural description" can, defenders of the moral grammar-hypothesis argue, offer a descriptively adequate account of the data.[248] Since this knowledge is unconscious and might bear no relation to express principles (see below), the fact that agents are unable to explain at least some of their judgments is no objection to the theory of moral grammar.

However, as it stands the theory of moral grammar has merely the status of a *possible* solution to the problem of formulating an empirically adequate account of the knowledge that underlies the moral judgments of mature agents. It gains some plausibility from the fact that in linguistics, the theory of generative grammar is a taken seriously on conceptual and empirical grounds, but this mere fact will not suffice to convince the skeptics. There is no contradiction in accepting Chomskian linguistics, but doubting the theory of moral grammar. After all, defenders of the theory have made a number of contingent assumptions that might turn out to be wrong (e.g. that the explanations that agents offer for their judgments are inaccurate, that moral judgments share a common rule system or that the rules that are part of the moral grammar are applicable to all possible moral problems). The proof is in the pudding, and in order to support those stipulations, the defender of the moral grammar-hypothesis needs to offer empirical evidence.

---

[248] Mikhail (2008), 84.

4.8 Mikhail's empirical case for the moral grammar hypothesis

In his experimental work,[249] Mikhail presents test subjects with hypothetical scenarios modeled after so called 'trolley cases', a set of thought experiments developed by Philippa Foot.[250] The first main scenario describes a bystander observing a driverless train whose brakes have failed while heading towards five workers; the workers have no time to escape and would die unless the bystander uses a lever to turn the train onto an alternative track. However, there is one person on the alternative track who would die in case the train is diverted. In the second main scenario, the bystander has no lever that would allow her to change the path of the train, but stands on a footbridge over the train tracks. The only way to save the five workers is to push a heavy man who is also standing on the bridge onto the track in order to stop the train in time. All of the other cases that Mikhail uses are derived from those two main ones. The judgments people offer in response to those problems are – as judgments concerning the grammaticality of a given expression in Chomskian linguistics – analyzed in search for an underlying rule system. Reactions to trolley cases and their variations promise insight into the process of moral decision making in virtue of the fact that beneath a fairly simple scenario, they instantiate a number of central moral concepts such as the significance of consequences, respect for rights, the means/end-distinction and the difference between intended and foreseen consequences. Furthermore, in trolley cases test subjects are required to judge the deontic status of possible courses of

---

[249] See especially Mikhail (2011), 77-181.
[250] See Foot (1967).

action, limiting the options to unambiguous responses (allowed/forbidden) that allow to compare similar cases in order to see whether people's responses are consistent across cases and for the purpose of finding out which changes in the scenarios are responsible for differences in judgments. Hence, the data gained from collecting judgments about trolley cases provides suitable data to approach the question of moral grammar. This turns the abstract problem of formulating a descriptively adequate account of moral knowledge into the tractable task of finding rules that make sense of people's judgment about trolley cases.[251]

Before we can take seriously the claim that the principles explaining mature agents' responses to trolley cases allow insight into the problem of moral grammar, we must ask whether they conform with the requirements laid out above. Regarding the distinction between operative and express principles, the explanations that the test subjects would offer for their judgments are not considered in the examination of their responses; therefore, the analysis of trolley cases results in operative principles. More challenging is the question of whether responses to trolley cases reveal moral competence or performance. As argued above, only judgments under ideal circumstances reveal competence. Mikhail argues that his data fulfills this requirement: the experiments "elicit considered judgments in Rawls' sense"[252] since they the test subjects are "disinterested third

---

[251] Problems with that inference are discussed below.
[252] Mikhail (2011), 82 Fn.2.

parties […] in relaxed settings where they have plenty of time to rely on all of their moral capacities."[253]

Here is how the data is collected: test subjects from various backgrounds are presented with 12 variations of trolley cases and asked to evaluate their deontic status. The study has been conducted on subjects of various nationalities, economic groups, levels of education etc.;[254] in total, over 200.000 individuals from over 120 countries have participated. The results show that a significant majority of participants agrees in their assessment of which courses of action are morally permissible and which ones are forbidden.[255] Furthermore, in line with assumptions introduced above, test subjects are – regardless of their age, gender, and level of education – unable to offer explanations that are consistent with their judgments.[256] Of crucial importance is that the *input stimuli* alone (the act description) is insufficient to explain the output, the deontic judgment.[257] Something else has to be postulated in order to understand the "mental

---

[253] Mikhail (2011), 96. In addition, the assessments of trolley cases are "stable, stringent, and highly predictable" (Mikhail (2011), 83). This provides at least indirect evidence that the data is free from distortions: in case the presence of distortions was pervasive, it would be plausible to assume that some of them – fatigue, states of emotional arousal etc. – would be present in some test subjects but not in others. The test results would reflect this fact and be 'noisier' than they are. Hence, it seems that the test settings Mikhail uses are able to control reasonably well for those factors and that his results approximate well-considered judgments.

[254] See Mikhail (2011), 106-110.

[255] See Mikhail (2011), 111 for a table with the results. I will discuss later how to interpret the diverging view of the (sometimes significant) minority. It would diverge from the present task to present all of the trolley variations Mikhail used in his studies or to discuss statistical details of the results; in the course of the further discussion, relevant scenarios will however be introduced.
It is true that my overall argument would be strengthened if I discussed the empirical details of Mikhail's study in order to confirm that his theory can indeed offer an empirically adequate theory of moral cognition, but for reasons of space (and time!) I have to forego this advantage and assume that the setup of his study is sound, the results correct and that the derivation of the deontic rules from his data is correct. Instead, I will focus on arguing – among other things – that his theory misses important dimensions of morality which would render it more akin to Ross's presumptions about morality.

[256] See Mikhail (2011), 110 and Edmonds (2013), 113.

[257] See Mikhail (2008), 83 and 87 and also his (2011), 111 and 114.

computations underlying the ordinary exercise of moral judgment."[258] This

becomes clear when looking at descriptions of trolley cases which superficially

resemble each other even though the majority of test subjects assign them a

different deontic status. Take two such scenarios, "Bystander" and "Loop Track."

These are the descriptions that test subjects receive:

> Bystander: Hank is taking his daily walk near the train tracks when he notices that the train that is approaching is out of control. Hank sees what has happened: the driver of the train saw five men walking across the tracks and slammed on the brakes, but the brakes failed and the driver fainted. The train is now rushing toward the five men. It is moving so fast that they will not be able to get off the track in time. Hank is standing next to a switch, which he can throw, that will turn the train onto a side track, thereby preventing it from killing the men. There is a man standing on the side track with his back turned. Hank can throw the switch, killing him; or he can refrain from doing this, letting the five die. Is it morally permissible for Hank to throw the switch?[259]

> Loop Track: Ned is taking his daily walk near the train tracks when he notices that the train that is approaching is out of control. Ned sees what has happened: the driver of the train saw five men walking across the tracks and slammed on the brakes, but the brakes failed and the driver fainted. The train is now rushing toward the five men. It is moving so fast that they will not be able to get off the track in time. Ned is standing next to a switch, which he can throw, that will temporarily turn the train onto a side track. There is a heavy object on the side track. If the train hits the object, the object will slow the train down, giving the men time to escape. The heavy object is a man, standing on the track with his back turned. Ned can throw the switch, preventing the train from killing the men, but killing the man. Or he can refrain from doing this, letting the five die. Is it morally permissible for Ned to throw the switch?[260]

In the bystander case, a large majority believed that it was morally permissible to

throw the switch; in the loop track case, most test subjects have contrary intuitions.

---

[258] Mikhail (2011), 106.
[259] Mikhail (2011), 106.
[260] Mikhail (2011), 107f.

To make sense of those judgments, the immediate content of the case descriptions

is not sufficient. In both situations, the agent has the choice between either letting

five men be killed or preventing their killing by turning the train on a side track

which kills one man.[261] The relevant difference between Bystander and Loop

Track becomes however intelligible as soon as we introduce additional concepts

such as the distinction between directly versus indirectly intended actions, the idea

of an "otherwise prohibited action," good and bad effects, the relation of

"outweighing" and the concepts of a morally preferable alternative.[262] Given this

conceptual apparatus, it is possible to form a hypothesis about the different mental

representation that the two cases trigger. In Bystander, the death of the one worker

is the result of an indirectly intended action that would otherwise be prohibited; the

alternative between saving five men versus killing one can be understood as a

situation where good effects outweigh bad ones. Things are different in Loop

Track: here, the death of the one worker must be directly intended since otherwise,

the train wouldn't slow down and the remaining five workers couldn't escape.

Once the initial act description is analyzed with the help of those distinctions,[263] a

moral principle that is sensitive to them can be applied and output different deontic

judgments for both cases. In the scenarios under discussion, the moral rule that

explains the intuitions is a version of the principle of double effect:

---

[261] See the act descriptions Mikhail provides in his (2011), 115.
[262] See Mikhail (2011), 149.
[263] This is done by treating the act description as "a body of information" that is unconsciously matched against a number of questions akin to the classical questions provided Cicero: "*Quis? Quid? Ubi? Quibus auxiliis? Cur? Quomodo? Quando?* Who? What? Where? By what aids? Why? How? When?" Mikhail (2011), 126.

> an otherwise prohibited action, such as battery or homicide, which has both good and bad effects may be permissible if the prohibited act itself is not directly intended, the good but not the bad effects are directly intended, the good effects outweigh the bad effects, and no morally preferable alternative is available.[264]

Given – as argued above – that agents are not consciously aware of those rules and unable to provide a fitting alternative explanation for their judgments, there is reason to believe that rules such as the doctrine of double effect are part of the moral grammar of mature agents.[265]

Generalizing this insight, the mental process that occurs between receiving the stimulus and assigning a deontic judgment can be described as an operation consisting of three elements:[266] (1) conversion rules "transform verbal or visual stimuli into a structural description"[267] and bring out an "implicit geometry"[268] that is hidden in the act description. This presupposes (2) a "lexicon of structural description,"[269] which entails concepts used by the conversion rules (such as the distinction between directly and indirectly intending an action, good and bad

---

[264] Mikhail (2011), 149.

[265] As the extensive literature about Trolley cases shows, there are scenarios which suggest that the doctrine of double effect does not cover all possible variations. In defense of Mikhail's research program, his main aim is to demonstrate that there is an unconscious rule system which guides our intuitions; as discussed below, he is not committed to the claim that there cannot be refinements of the rules he suggests or exceptions as long as they can be accounted for in terms that can be incorporated in a system of moral grammar.

[266] See Mikhail (2011), 116; see also p120f. for a more detailed description which analyzes the process as a sequence of six mechanical steps.

[267] Phelan (2012).

[268] Mikhail (2011), 120.

[269] Phelan (2012).

effects etc.). Once the structural description has been created as a mental representation, (3) deontic rules can be applied and yield judgments that describe the action in question as morally permissible or forbidden.[270]

4.9 Deontic rules

In order to make sense of the intuitions about the variations of trolley cases that he has collected from his test subjects, Mikhail proposes a list of deontic rules. He starts with the "Principle of Natural Liberty" or "Residual Permission Principle" according to which an act is forbidden if it "has features F1…Fn […]; otherwise, it is permissible."[271] In other words, if a particular action falls under no rule specified by the moral grammar, the action counts as morally permissible.

Two criticisms:[272] it is not clear how surveys about a small sample of cases can exclude the existence of rules that are not part of the moral grammar. Culture or individual upbringing, for example, might account for those additional principles. In order to draw such a conclusion, intuitions about all possible moral problems would have to be evaluated, a requirement impossible to fulfill. Second, rather than increasing confidence in the completeness of Mikhail's list of rules, the principle of natural liberty serves as an invitation to find examples which illustrate how much his theory leaves out (there is no word in Mikhail's theory about the institution of promise, respect for autonomy, parental duties etc., even though these

---

[270] See Mikhail (2011), 111-122, especially figure 5.4 at p114.
[271] Mikhail (2011), 133.
[272] I raise my worries about the closure rule here and my more general criticisms about Mikhail's project as a whole in a separate section below because my present criticism relates to whether the closure rule can be derived from the empirical data; my more general criticisms are of a different kind and it might be confusing to mix them together.

are commonly recognized grounds of duty and would have to be accounted for in a complete system) and consequently how difficult it is to formulate a complete moral grammar.[273]

The other principles are less problematic. As a second deontic rule, Mikhail introduces the prohibition of battery and homicide.[274] Homicide is understood as "causing the death of a person"[275] and can be committed purposeful or mere knowingly. To commit battery is defined as "causing harmful contact with a person without her consent"[276] and can also be done in a way that is either purposeful or knowingly. The principle can explain a range of intuitions about variations of trolley cases.[277]

Third is the self-preservation principle. Its purpose is to define under what circumstances it can be assumed that person would not consent to an action. In Mikhail's formulation it states that "if a harmful contact with a person necessitates killing her, then she would not consent to it."[278] The principle helps to explain different intuitions towards a trolley case variation ("Implied Consent") where a bystander has the chance to throw a man off the track, saving him from death through a collision with the trolley, but hurting him in the process from the already described Footbridge case where the large man is also thrown, but for the purpose

---

[273] In fairness to Mikhail, this does not show his approach to be flawed, but merely points attention to the challenge the development of a fully worked out theory would face. He repeatedly recognizes that at the present stage, his theory of moral grammar "remains incomplete in many aspects" (Mikhail (2011), 178).

[274] See Mikhail (2011), 133-137.

[275] Mikhail (2011), 134.

[276] Mikhail (2011), 134.

[277] For the purposes of keeping the presentation of Mikhail's theory as short as necessary, I abstain from describing all of the relevant cases and from showing how the principle applies to them. Instead, I refer to Mikhail's reconstruction which I believe is correct: see his (2011), 154-156.

[278] Mikhail (2011), 137. Obviously, this is not an exhaustive definition of consent, but it is all that Mikhail needs in order to explain the trolley cases he is analyzing.

of sacrificing his life in order to save the five workers. With the self-preservation

principle in mind, it is clear why the prohibition of battery and homicide applies to

the second case but not to the first (where the person would consent to being

thrown and consequently being hurt).

Next, Mikhail turns to the moral calculus of risk.[279] Good and bad effect

must be specified; but as seen in the discussion about the doctrine of double effect

(for more applications, see below), agents also need to be able to rank and compare

them. Mikhail postulates three primary bad effects: the death of a person, bodily

harm to a person and the destruction of a valuable thing.[280] Second, those bad

effects are ranked in an ordinal way and connected through a "morally-worse-than

relation."[281] The worst effect is the death of a person, followed by bodily harm to a

person. The bad effect with the least relative weight is the destruction of a valuable

thing. Furthermore, values can be aggregated: e.g. the death of more people is

worse than the death of less people.[282] More complicated is the comparison

between bad effects. Here, the ordinal ranking of bad effects contributes only one

aspect to the ranking; probabilities also enter the picture. In order to determine "the

reasonableness and hence justifiability of a given risk," five variables are of

importance:

> (i) the magnitude of the risk, Rm; (ii) the value of the principal
> object, Vp, which may be thought of as the life, safety, or property

---

[279] See Mikhail (2011), 137-144.

[280] Mikhail (2011), 137. I leave out a number of complications that Mikhail discusses, but which don't matter for the our purpose such as the definition of good effects and the badness of preventing a good effect.

[281] See Mikhail (2011), 138.

[282] Mikhail does not specify how different kinds of bad effects compare in case their quantity varies (e.g. is the death of one person worse than severe bodily harm to ten other people?).

interests of the individual in question; (iii) the utility of the risk, Ru; (iv) the necessity of the risk, Rn; and (v) the value of the collateral object, Vc, which may be thought of as the actor's own purpose in imposing the given risk. […] [J]ustifiability depends on whether Rm multiplied by Vp is greater than (i.e. morally worse than) the combined product of Ru, Vc, and Rn: $(Rm)(Vp) > (Ru)(Vc)(Rn)$.[283]

The reason for quoting this formula derived from a legal context which plays no further role in the remaining discussion is that it shows the idealized nature of (at least parts of) Mikhail's reconstruction of our moral grammar.[284] After all, the empirical evidence collected from the responses to the variations of trolley cases that Mikhail relies on provides no conclusive grounds for supporting this specific principle. This will be helpful to keep in mind for the critical discussion below.

Besides playing the role of an auxiliary principle for the doctrine of double effect where the good effect must outweigh the bad effect,[285] the moral calculus of risk (or a variation thereof) might also play a more direct role in explaining our intuitions. As discussed above, in Bystander most test subjects agree that it is morally permissible to throw the switch, whereas in Footbridge, the majority believes it to be forbidden to push the large man onto the tracks. The already presented deontic rules allow us to explain the difference in judgment: according to the principle of double effect, the death of the one worker in Bystander is not directly intended, but in the Footbridge case, it is. The moral calculus of risk can provide the same explanation and could therefore replace the doctrine of double

---

[283] Mikhail (2011), 140.

[284] Other rules, such as the doctrine of double effect, the prohibition of battery and homicide or the self-preservation principle are however better confirmed by the available data (and also intuitively more plausible), even if they might have to be refined in order to provide accurate predictions about all cases they cover.

[285] Mikhail does not consider this function of the moral calculus of risk, but I take it that it partly justifies its place on the list of principles that are part of the moral grammar.

effect and potentially Mikhail's other deontic principles as well.[286] If test subjects

assume in Footbridge that the chances that the large man can bring the trolley to a

halt in time are below 0.2, "the expected costs would exceed discounted expected

benefits, and the conduct would be unjustifiable on that basis alone."[287] In

Bystander however, it is plausible to assume that the likelihood that diverting the

train saves the five workers is 1; based on the moral calculus of risk, it is therefore

allowed.[288] However, Mikhail argues that nevertheless, cost-benefit analysis falls

short as a theory of moral grammar. Apart from arguments from developmental

psychology observing how small children arrive at normative judgments (not by

conducting a cost-benefit analysis)[289] and other considerations that cannot be

evaluated in the present context,[290] there are variations of trolley cases which

cannot be explained by this rule.[291] Furthermore, on a phenomenological basis, it

seems implausible to believe that the fact that somebody is being used as a means

plays no role in the formation of our judgments. However, even though the moral

calculus of risk cannot replace the other deontic rules, Mikhail argues that it "may

play a subordinate operative role in these problems."[292] He does not elaborate on

what role this could be, but it could help him solve a major problem his view is

---

[286] This alternative raises to the descriptive level the fundamental question as to whether consequentialism can replace right-based theories.

[287] Mikhail (2011), 142. This interpretation of Footbridge would contradict the explicit description handed to the test subjects where the implicit assumption is that the likelihood that the large man saves the five workers is 1, but it is not implausible that test subjects believe that this stipulation is not realistic and "correct" the mistake.

[288] Mikhail (2011), 142 adds further cases where the moral calculus of risk could also explain people's intuitions, but for the sake of briefness, I will not discuss those scenarios.

[289] See Mikhail (2011), 143.

[290] See Mikhail (2011), 143.

[291] Again, for reasons of space, I will not discuss those scenarios and refer to Mikhail (2011), 143f.

[292] Mikhail (2011), 144.

facing: the variance in intuitions. For many of his trolley cases, a significant

portion of test subjects disagrees with the majority regarding the permissibility of

the action in question. How is this possible if all mature agents are supposed to

possess an identical moral grammar? The moral calculus of risk could help explain

this situation. Even though all participants follow the same deontic principles, they

diverge in their perception of the probabilities that are relevant in the scenario, and

this affects their overall judgments.[293]

This discussion shows two things. First, the fundamental and for

intuitionists central normative debate over whether forms of consequentialism can

replace non-consequentialist right-based accounts gets replicated at the descriptive

level. Second, Mikhail's unclear account of how judgments about probability

interact with other deontic rules raises the question of whether a fully worked out

theory of moral grammar can be executed through the deontic schema presented

above.

Mikhail's next deontic rule consists in a rescue principle.[294] It specifies a

prohibition of failing to rescue an endangered person, however not at all cost: the

duty is lexically ordered and subordinate to other principles, such as the

prohibition of "purposeful homicide, purposeful battery, or knowing homicide

whose (discounted) expected benefits do not exceed its expected costs."[295] This is

---

[293] Caution has to be exercised with this argument: it cannot be that any possible test result is explained by pointing out that the underlying principles all test subjects use are the same, while their assumptions about probability differ. Refined studies would be necessary, testing people's actual probability assumptions alongside their intuitions (such information was not collected in Mikhail's experiments). This would allow to determine whether despite different overall judgments, test subjects use the same deontic rules.
[294] See Mikhail (2011), 144-148.
[295] Mikhail (2011), 147.

in line with intuitions about trolley cases where the five workers can only be saved by e.g. pushing a large man from a bridge, while it corresponds to the permissibility judgment in scenarios such as Costless Rescue. Mikhail concludes his list of deontic rules by introducing the doctrine of double effect which has already been discussed above.

Now that the *content* of our moral grammar is on the table (that is, not all the rules required to cover the complete moral terrain, but the principles necessary to deal with trolley cases), what *form* does the process of moral reasoning leading to deontic judgments take? In answer to this question, Mikhail argues that the deontic rules provide "necessary and sufficient conditions for assigning a deontic status to a given act or omission"[296] and take the form of the following schema:

A has deontic status D = A has features F1…Fn
A has features F1…Fn
A has deontic status D[297]

Even if everyday moral reasoning does not follow this schema (because agents misapply the rules, ignore some features and invent others, allow irrelevant factors such as emotions or wishful thinking to interfere with the conclusion etc.), an idealized mature agent in possession of a moral grammar and in absence of defeaters would unconsciously derive his or her intuitions by using Mikhail's schema. Since the schema mechanically matches the features present in a given situation with the respective deontic rule, the agent executing the schema does not

---

[296] Mikhail (2011), 124.
[297] Mikhail (2011), 124. It should be emphasized that the features Mikhail refers to are not limited to natural features; the fact that an action is directly intended, that its effects are worse or better than another action or that it implies implicit consent would all count as features.

need to be a human being, but could also be a machine such as a robot.[298]

Comparing moral deliberation to the calculation of a chemical reaction, trolley cases (and presumably any other moral problem as well) can be placed on a "periodic table"[299] which indicates what elements (such as homicide as an end or means or a side effect, battery as an end or means or a side effect etc.) a scenario instantiates. Applying deontic rules, it can be read off the table whether an action is morally permissible or forbidden.

It is important not to confuse the two ways in which Mikhail considers his account closed or complete since it is possible to agree with one sort of closure and disagree with the other (and vice versa). In accordance with the Residual Permission Principle, it can be argued that an ideal complete moral grammar contains all of the deontic rules are sufficient for account for all well-considered moral judgments; at the same time, one can doubt that Mikhail's "computational account of moral cognition"[300] offers a complete theory of how moral deliberation works. Before turning to that possibility, I want to discuss (a) the analogies and disanalogies between moral and linguistic grammar and (b) the similarities and dissimilarities between Mikhail's theory of moral grammar and Rossian intuitionism.

---

[298] This is not a side-issue for Mikhail, but one of the main motivations behind his theory. See Mikhail (2011), 39, 90, 124 and 153.
[299] See Mikhail (2011), 153-162.
[300] Mikhail (2011), 124.

4.10 The relation between moral and linguistic grammar

Why bother with the relation between Mikhail and Chomsky's theories when our concern is about the epistemology of Rossian intuitionism? Since the overall aim is to use Mikhail's approach in order to answer the intuitionist' quest for a descriptively adequate account of moral knowledge, a more complete understanding of what Mikhail is committed to would be of interest. By considering where the theory of moral grammar breaks off from its linguistic counterpart, some important features of Mikhail's theory that he does not focus on come to light. Hence, this means that an intuitionist whose epistemology is informed by Mikhail's theory should be interested in the relationship between the moral grammar hypothesis and modern linguistics. Although there is no formal way to measure the closeness of the analogy, it is apparent that a number of distinctions and arguments from modern linguistics are central to Mikhail's case for a moral grammar:[301] the distinction between operative and express principles, between descriptive and explanatory adequacy, the competence-performance distinction, the argument for linguistic grammar and the argument from the poverty of stimulus.[302] Even if defenders of a theory of moral grammar reject a

---

[301] There is a fundamental commonality between linguistics and the moral grammar hypothesis that should not be overlooked and which sets Mikhail's theory apart from most other theories in moral psychology: in both linguistics and in his approach, contingent empirical evidence serves as a premise to the argument. Mikhail (2011), 68: "both the argument for moral grammar and the argument for linguistic grammar are non demonstrative. [...] The arguments are an attempt at scientific explanation; we are not compelled to accept them by means of conceptual analysis or logic. Alternative explanations are logically possible."

[302] There are even more distinctions that Mikhail borrows from linguistics, but that I have omitted them for reasons of space and because they are not essential for present purposes. For a list of all features, see Mikhail (2011), 42f.

The actual rules governing linguistics and morality are obviously different (for once, morality lacks the syntax/semantics distinction), but the closure rule seems to be an exception: a complete list of linguistic rules should be sufficient to determine the grammaticalness of any sentence to the same extend that

commitment to a "strong version" of the analogy according to which "language and morality work in much the same way,"[303] the parallels between moral and linguistic arguments are significant.[304] For present purposes however, more interesting than the numerous commonalities are the differences. I want to mention two of them.

First, in the linguistic case the rules of grammar "are assumed to lie beyond [the] actual and even potential consciousness" of competent speakers.[305] Therefore, hypothesis these speakers form about the rules that underlie their linguistic competency (express principles) are no reliable guide for discovering the actual operative principles of their grammar. Instead, linguists have to employ scientific methods in order to unveil the grammatical code. Does the same hold for morality? Mikhail's answer is ambiguous. In his book, he argues that analogous to the linguistic case, moral agents have no direct cognitive access to the content of their moral grammar.[306] In later publications however, he backs away his this claim and

---

(according to Mikhail) a complete list of deontic rules suffices for determining whether an action is morally forbidden or allowed.

[303] Hauser et al., (2008), 139.

[304] For this reason, I believe that the claim that the linguistic analogy "is merely a heuristic for posing the right sorts of questions about the nature of our moral competence" (Hauser et al (2008), 139) is a strategic move (and not necessitated by the argument) supposed to preempt criticisms that the analogy isn't carried far enough (for example, it might be asked whether Chomsky's notion of idealization has an adequate counterpart in the argument for moral grammar). Mikhail describes the relation between the two disciplines by emphasizing that the "'the theory of moral cognition is usefully modeled on aspects of linguistic theory'' (42, emphasis added) [Mikhail is referring to his (2011)]. These qualifications are important. With only limited exceptions, the primary comparisons between linguistics and ethics I draw in EMC [=Elements of Moral Cognition, his (2011)] are theoretical and methodological, not substantive." Mikhail (2013), 24. Since there is no doubt that Mikhail is using many of Chomsky's distinctions of arguments in order to argue for a moral grammar, it is not clear what the distinction between methodological and substantive arguments means in this context.

[305] Mikhail (2011), 19.

[306] Mikhail (2011), 20: "[M]oral theorists should not assume that the normal individual is aware of the operative principles that constitute her moral knowledge, or that she can become aware of them through introspection, or that her statements about them are necessarily accurate."

takes an agnostic stand.[307] Mikhail does not indicate the motivation for his change

of mind, but he has good reasons to be more cautious. In the case of linguistics as

well as in other areas such as "the study of […] vision, musical cognition, face

recognition, and other cognitive domains,"[308] the fact that we cannot have

knowledge of the principles that guide our behavior does not infringe on the

practices in which those abilities are embedded. In linguistic for example, under

normal circumstances competent speakers don't face the need to directly appeal to

the principles of grammar in order to justify their linguistic judgments to others or

in order to figure out for themselves whether an expression is grammatical or not.

Things are different for the moral case. Even though for a wide range of cases,

mature agents have – as the responses to the trolley cases indicate – no difficulty in

dealing with moral problems, there are issues on the individual and societal level

where even under idealized circumstances, people are unclear about their

judgments or have firm convictions, but disagree among each other. In those

situations, progress can often be made by considering what the moral principles are

that underwrite our judgments in less controversial cases and to apply those

already accepted principles to the problem at stake; a similar method consists in

bringing out what principles the various options in a difficult case presuppose and

to hope that there is more agreement concerning what abstract principles should or

---

[307] Inaccurately (compare the quote in the previous footnote) he claims that this was already his position in his (2011). Mikhail (2013), 5: "I do not assume (nor do I deny) that these principles are introspectible or accessible to consciousness in principle. For example, I do not assume that an individual who is unaware of the theoretical claims of moral psychology will recognize certain deontic rules as the rules she has been following all along. It may turn out that some components of moral knowledge are unconscious, yet accessible by ordinary means of reflection, introspection, or other methods. On the other hand, researchers might discover that some aspects of moral knowledge are both unconscious and inaccessible. *In EMC, I did not take any firm position on these questions*" (emphasis added).
[308] Mikhail (2011), 21.

shouldn't be the base of our deontic judgments. To believe that agents have no

cognitive access to the principles that guide their moral deliberations would require

an error theory about this moral practice. Any moral reflection that either infers

what principles agents have relied on or which prescribes which principles should

guide our judgments would be an exercise in dumbfounding.[309] While few would

deny that often, we rationalize our judgments by making them superficially

consistent with certain principles when as a matter of fact, other considerations are

responsible for our intuitions, it is also part of our common understanding of moral

deliberation that detached reflection and argument can bring us closer to the actual

reasons behind our moral beliefs and that the methods of moral deliberation

described above can be conductive to finding correct answers. Sometimes,

appearances cannot be saved, but strong reasons have to be offered (and Mikhail

has not provided them) in order to make the case that firmly held convictions are

false. At the opposite, some evidence points into the opposite direction. If the rules

of our moral grammar were inaccessible to mature agents, it would be

unexplainable why Ross' list of *prima facie* duties – which are supposed to be self-

evident, i.e. derived from "what we really think" – bears a remarkable resemblance

to Mikhail's list of deontic principles (more on the comparison between the two

below).[310] In fact, even where Mikhail's moral grammar goes beyond Ross'

theory, those principles are either familiar to common sense or can be found in

---

[309] The problem raised here concerns reasons for forming a judgment based on certain principles. The distinction between operative and express principles leaves untouched reasons for action (unless those reasons imply that we should act from a certain motive).

[310] Even though Ross was a moral philosopher and trained in systematic reflection on moral theories, he didn't use any of the scientific methods that Mikhail claims are needed in order to unveil operative principles.

other philosophers. The doctrine of double effect, for example, was not discovered

by analyzing empirical data and forming structural descriptions that match mental

representations of act descriptions, but when Aquinas reflected on the

permissibility of self-defense.[311] This does however not imply that there is no

difference between operative and express principles; it merely suggests a weaker

version of the distinction. It is likely the case that only in few cases, carefully

examined explanations of our moral judgments based on self-reports are accurate;

in most other instances, they should not be relied upon. Hence, as a generalization,

express principles should be regarded with suspicion, but there is no need for the

stronger claim that even under ideal circumstances, it is impossible that they

approximate the rules operative in our decision making.[312] This weaker version of

the distinction between operative and express principles is still sufficient to support

Mikhail's argument that Kohlberg's approach of giving credence to self-reports is

flawed and that instead, people's judgments should be analyzed by means of

scientific methods introduced by Chomsky. At the same time, this weaker

distinction is closer to our moral practice and makes Mikhail's theory a more

attractive candidate for a descriptively adequate moral psychology.

The second disanalogy concerns the fact that test subjects vary in responses

to trolley cases. As a look at the data shows, some of the test cases elicit significant

---

[311] See Aquinas (2012), II-II, Qu. 64, Art.7

[312] Even a stronger point could be made against Mikhail: in many natural languages such as in English, there are useful express rules (such as subject verb agreement) which offer useful guidance. Hence, not even in linguistics is the distinction between operative and express principles as sharp as Mikhail makes it sound like.

disagreement.[313] This conforms with common sense experience: it is rare to find an

ethical issue of significance that is undisputed.[314] Those disagreements occur not

only between different cultures, but also within them. In linguistics however,

within a community of speakers of a natural language, there is much less variance;

if most grammatical assessments were disputed, communication would be more

difficult than it is.[315] What explains this significant disanalogy between linguistics

and moral psychology? In order to get the argument for moral grammar off the

ground, this problem needs to be addressed. Otherwise, opponents of the moral

grammar hypothesis will argue that if people differ in their moral judgments, the

best explanation for this fact is that they don't share a common moral code.[316] I

cannot fully answer to that problem (one of the reasons is that this would require

further empirical studies which is something beyond my means and expertise), but

I will discuss and reject some of the replies that have been suggested and then

propose my own possible solution.[317]

---

[313] See Mikhail (2011), 319-350.

[314] It might be objected that Mikhail is interested in operative principles and that in everyday life, defeaters distort the application of our moral grammar, but as argued above, the distinction between operative and express principles appears to be less sharp than in linguistics. Hence, we should expect that in everyday life, well-reflected opinions should express (even if imperfectly) our moral grammar and hence converge, but that does not seem to be the case.

[315] Different languages, while following different surface rules, still share the same structural features (e.g. the difference between verb and subjects); Chomskian linguistics is concerned with the structural features, not with the surface grammar.

[316] For a good reply to that worry, see Huemer (2005), 128-155.

[317] Mikhail presents Rawls solution to the problem of moral diversity: we should focus on the principles that guide one person and then "take for granted that these principles are either approximately the same for persons whose judgments are in reflective equilibrium, or if not, that their judgments divide along a few main lines represented by the family of traditional doctrines that I shall discuss." Rawls (1999, 50), quoted from Mikhail (2011), 56. I shall not consider this solution here: with no reliable data at hand, it might be a good idea to simply assume that the moral grammar of one person can be extended to other people as well, but since Mikhail's experiments have changed that situation, we not have reason to question Rawls' idea.

Mikhail suggests that by paying attention to the structural descriptions of various scenarios, we can explain why in some cases, there is a higher likelihood than in other scenarios that test subjects judge a scenario to be ethically permissible: "Other things equal, acts are more likely to be judged permissible as counts of battery committed as a [sic] means decrease from three (Footbridge) to two (Drop Man) to one (Loop Track), and as these violations become side effects and additional features come into play."[318] In the Footbridge scenario which only 10% of test subjects judge as permissible, the large man needs to be touched (first battery) and thrown at the track (second battery); therefore, the bystander causes the train to hit the large man (third battery).[319] In contrast, in Drop Man, a scenario identical to Footbridge apart from the stipulation that the bystander pushes a button which opens a door beneath the large man and causes him to fall onto the tracks, 35% judge the action to be permissible. Here, we have one less battery since the large man does not need to be touched. In Loop Track, almost half of all test subjects judge the action to be permissible. The explanation, Mikhail suggests, is that there is only one battery involved (causing the train to hit the man). This solution, which Mikhail himself labels as "tentative"[320] and which he does not use to explain variance in later publications,[321] has problems. Relying on the number of batteries that are committed as means presupposes that there are well-justified ways of distinguishing how to count them. It is however not obvious why for

---

[318] Mikhail (2011), 126.
[319] See Mikhail (2011), 118.
[320] Mikhail (2011), 126
[321] See Mikhail (2013). However, he does not present an alternative explanation either, leaving it open how we would defend his theory against this significant objection.

example (as in Footbridge) touching and throwing the large man counts as different instances of battery. More severely, the method of counting numbers of batteries seems to be ad-hoc. While it might explain the variations in the trolley cases that Mikhail uses for his experiments, the method will not work for other scenarios where the count of batteries is increased, but their severity is reduced (a trolley case where the bystander needs to push the large man ten times with moderate force but against his will in order to convince him to throw a rock onto the track will in all likelihood be judged by more test subjects as permissible than cases that Mikhail uses even though they require a lower number of batteries). It is also not clear how Mikhail's talk of different degrees of likelihood of permissibility judgments is compatible with the idea that all test subjects have the same moral grammar. What about the responses of the test subjects in the minority who have the intuition that even though the number of batteries is high, the action is still permissible (or vice versa)? Are all those responses to be discounted? On what ground?

As an alternative account of variability in permissibility judgment, it could be argued that based on an analogy to Chomskian linguistics, we shouldn't even expect that all test subjects obey the same rules. After all, the linguistic grammar only determines structural features and allows individual languages to differ in their surface rules. Moral grammar might behave in the same way, delineating only

boundaries, but not determining the exact nature of specific moral rules.[322] Sripada

argues that our moral grammar consists in

> *high-level* themes that one sees in the contents of moral norms in
> virtually all human groups – themes such as harms, incest, helping
> and sharing, social justice, and group defense. However, the *specific
> rules* that fall under these high-level themes exhibit enormous
> variability.[323]

If this was correct, the cost for accommodating the problem of variability would be

to give up the specific account of Mikhail's deontic rules and to assume that the

moral grammar is divided into less specific clusters of areas of moral relevance.

The problem with this suggestion is that it does not fit well with Mikhail's data.

Since the specification of the high-level themes occurs at the level of human

groups, we should based on Sripada's account expect that the differences in

permissibility judgments coincide with the cultural background of the test subject.

This, however, is not the case. Response patterns are very similar for test subjects

from various geographical, educational and economical backgrounds.[324] Hence, the

variety does not stop at the group-level, but extends to individuals.[325]

In order to respond to the problem from variability without making major

changes to his theory, Mikhail needs to show how the differences in people's

judgments can be accounted for by appeal to factors other than the deontic rules. In

order to avoid the criticism against Sripada's theory, those differences must be able

to occur within societies, not only between groups (although it should be possible

---

[322] For a suggestion along those lines, see Hauser et al. (2008), 118-120.
[323] Sripada (2008), 330.
[324] See Mikhail (2011), 110.
[325] This is not to say that Sripada's account is false; rather, she seems to be focusing on thick concepts that carry descriptive baggage which can be defined differently in different contexts (e.g. notions of honor or social justice). On this level, cultural variation is more plausible.

that groups also play a role in shaping how moral judgments are being made). At three steps in the process of moral deliberation, those factors might be located: first, test subjects can interpret in various ways the facts that have to be established before a moral rule comes to bear (e.g. how serious is the harm being caused? How likely are certain outcomes judged to be? Who counts as a moral agent?); second, in scenarios where duties conflict, individuals might vary in the weight they assign to them; third, nothing in the moral grammar hypothesis excludes that agents adopt through their culture or by personal choice of additional moral rules that can play an additional role in their moral deliberation. This would however require a change in Mikhail's closure principle. Those three factors can explain moral disagreement while keeping the deontic rules that are part of the moral grammar fixed.[326] In order to count as a solution for the concrete problem of how to account for variability in deontic judgments in concrete cases, it is however not enough to present a hypothesis about what test subjects disagree about. Rather, it would be necessary to test this assumption empirically. If it turned out that test subjects with different intuitions follow different underlying rules, this would imply a significant disanalogy between linguistics and moral psychology and raise serious issues about the viability of Mikhail's research program.

In conclusion, the discussion of how deep the analogy between the two discipline goes has revealed one aspect (the distinction between operative and express principles) where the moral grammar hypothesis can be made more

---

[326] It would be an interesting exercise to go through each of Mikhail's variations of the trolley problem (see Mikhail (2011), 106-109) and to consider which of those factors might be responsible for the disagreement among test subjects. For reasons of space (and because it wouldn't help the main question pursued here), I am not going to engage with that question.

convincing by stressing the difference between linguistics and moral psychology; in another regard (the issue of moral disagreement), the difference spells potential trouble for the argument.

4.11 The relation between Mikhail's deontic principles and Ross' list of prima facie duties

If the ultimate goal is to explain claims made by Rossian intuitionists in terms of the moral grammar hypothesis, it must be shown that these two theories converge in the account they offer of what is in Ross' and Audi's terms self-evident and for Mikhail the best interpretation of our well-considered moral judgments. I will first consider what both have in common and then discuss two disagreements between them.

Even though Mikhail and Rossian intuitionists employ different methodologies – for Mikhail, empirical considerations play an essential role while Rossians offer a conceptual account – both theories are seeking to uncover the fundamental moral rules that govern the moral deliberation of mature agents.[327] At first sight, an overlap between the respective principles exists, but is far from complete:

---

[327] Rossian intuitionists believe in addition that these rules are made true by a mind-independent moral reality. This second claim obviously does not follow from a descriptive account and needs a separate defense; see the next chapter.

Ross	Mikhail[328]
Fidelity	
Reparation	
Gratitude	
Non-maleficence	Prohibition of intentional battery and homicide
Justice	The life of one person has as much moral worth as the life of any other[329]
Beneficence	Rescue principle
Self-improvement	
	Principle of self-preservation
	Moral calculus of risk
	Doctrine of double effect

Table 4.1

Mikhail also defends the closure rule, but since it is a second-order rule

specifying the completeness of his list of deontic rules, I have not listed it in the

table. Ross shares the same belief. For him, the list of *prima facie* duties specifies

all morally relevant rules.

The fact that the overlap between deontic rules and *prima facie* duties is

only partial does not imply that Mikhail would reject rules concerning fidelity,

reparation, gratitude or self-improvement. His aim is to explain a narrow set of

thought experiments that deal with moral obligations in the context of harming or

---

[328] As Mikhail stresses in his (2013), 25f., there are further "elements of moral cognition" that must be presupposed as part of the moral grammar such as the basic principles of deontic logic, of first-order logic, causal concepts etc. Since the present focus is on deontic rules/*prima facie* duties, I will omit those additional assumptions. Even though there is no mention of them in Ross, it is plausible that he also needs to presuppose them.

[329] See Mikhail (2011), 167. This principle, which is not given by Mikhail the status of a deontic rule, does not directly correspond to Ross' *prima facie* duty of justice which is concerned with the distribution of pleasure in accordance with merit, but both rules can be taken to be motivated by the same concern: human beings are equal and merit the same treatment unless an unequal treatment is justified by a relevant difference.

saving other people; fidelity, reparation etc. play no role here. However,

concerning the areas both Ross and Mikhail consider (non-maleficence,

beneficence and to a smaller extend also justice), both have similar rules. Had

Mikhail included a broader range of scenarios – e.g. the ultimatum-game or

variations of the prisoner's dilemma – it stands to be expected that he would have

found rules regulating those areas of life as well, increasing the overlap between

him and Ross.

Ross, on the other hand, has no explicit principle of self-preservation, pays

no attention to the moral calculus of risk or the doctrine of double effect. The

reason, as with Mikhail, is not that he would necessarily oppose those principles.

His aim is to develop a moral theory that entails principles which are not reducible

to either consequentialist or deontologist theories. If it turned out that mature

people who had reflected on those principles were unwilling to give them up (and

Mikhail's data suggest this), Ross would have reason to add them to his list of

prima facie duties.

As another similarity, Ross and Mikhail are both skeptical about the

unfiltered moral judgments of everyday people and rely on idealization – Ross by

stressing the need for repeated and detached reflection and Mikhail by defining

situations which elicit well-considered judgments.[330]

Two differences between intuitionism and the moral grammar hypothesis

deserve special attention. First, the two theories use different methods in order to

---

[330] In his description of the kinds of moral judgments he is trying to isolate, Mikhail even refers to Ross. See Mikhail (2011), 83. There, he agrees with Ross that "the judgments are typically made with a high degree of certitude and, indeed, are often experienced as being "objective" rather than a matter of subjective taste or preference from a phenomenological perspective."

determine the moral principles that guide moral deliberation. Mikhail investigates the moral grammar from a third-person standpoint, whereas intuitionists such as Ross adopt a first-person perspective. For him, the status of a belief as self-evident is determined by the degree of certainty with which it is held after due reflection; our mode of access is introspection.[331] Awareness of inner states does not matter to Mikhail; in order to count as part of the moral grammar, a rule needs to have explanatory power in regards to our moral judgments. As a consequence, Mikhail puts more trust in judgments about concrete cases than into his abstract principles. The latter are – at least in principle – open to challenge and reformulation should new data about concrete judgments emerge and be incompatible with the old principles. Ross in contrast is skeptical about concrete judgments and assigns a higher epistemic status to the outcome of reflection. This aspect of their difference about the method of determining moral principles is however not fundamental. Mikhail assumes that by choosing appropriate test settings and by presenting test subjects with well-suited scenarios such as trolley cases, he is able to elicit judgments that are well-considered and don't require test subjects to weigh moral principles against each other, a process which would introduce a further source of error (according to Mikhail, one rule is sufficient to solve the trolley problems). Ross on the other hand believes that in most cases, a conflict of duties is unavoidable and this reduces his trust in the reliability of concrete moral

---

[331] This refers to Ross' reflective method. The same could however be said about intuitive induction: I need to figure what in a number of instances that what makes a certain action wrong is that it involves e.g. lying. From there, I can infer that lying is bad everywhere. In order to establish what it is that caused my judgment in a certain case, I have to introspect my beliefs.

judgments.[332] Further, the concrete judgments Ross has in mind haven't been collected in controlled test settings, but are the result of an encounter with everyday moral problems. Faced with this data, he needs to correct for possible errors in the process of reflection; this is why he puts more trust in the abstract level than in concrete judgments. Nothing in Ross' theory would however forbid him to accept that under Mikhail's assumptions, concrete judgments deserve a more elevated epistemic status. The more important question is whether the intuitionist is able to accept that descriptive adequacy – which is, as argued above, an important element in Ross' moral epistemology – can also be achieved through the third-person perspective, i.e. by analyzing the responses test subjects offer to moral problems under idealized conditions. Again, the answer seems yes. What counts for Rossian intuitionism is that we identify the principles which are in line with well-considered moral judgments. Ross is not married to introspection; for him, it only serves as a means to an end. If a better methodology becomes available, he should be flexible and adopt it. As a further reason to be open to alternative accounts of descriptive adequacy, I have argued in the previous two chapters that the idea of self-evidence faces problems. Hence, there are two good reasons for the intuitionist to stop talking about self-evidence and to adopt an empirical account such as the moral grammar hypothesis: not only does the moral grammar hypothesis offer a more reliable way to determine the rules underlying

---

[332] As discussed in Ch. 2, Ross believes that more than one duty usually bears on a situation not only because he believes that most cases are too complex to be reduced to a single *prima facie* duty, but he also believes that good and bad consequences that are not intended and occur in the far future impact the deontic status of an action. I take it that it is possible to be a Rossian intuitionist, but to reject such a view as implausible; it also makes it easier to accept Mikhail's assumption that there are cases that can be explained by one moral principles.

our moral judgments, it also helps the intuitionist to replace an element in her

theory which makes the whole view less plausible than it could be.[333]

To argue that Mikhail's theory can be put to serve on behalf of intuitionism

requires however that it preserves its other main insights.[334] This is where the

second difference between Ross and Mikhail comes into play. It concerns the way

that rules interact in cases of conflict. As seen above, in Mikhail's picture of moral

deliberation, there is no room for weighing between competing moral principles:

rules specify which features in an action correspond to which deontic status. If the

presence of those features has been established, the deontic status automatically

follows.[335] Any "machine capable of computing judgments" can apply them;

Mikhail compares the task of the moral psychologist to that of an "engineer faced

with the task of designing [such] a machine."[336] Trolley cases fit this model. Even

though in most of Mikhail's variations, the action has good and bad

consequences,[337] there is a single principle such as the doctrine of double effect

which negotiates between them in a way that does not require the weighing of two

deontic principles against each other. In contrast, it is a central part of Ross' theory

that conflicts of duties are a frequent feature of morality and that there is no

---

[333] And in case both the self-evidence account and the moral grammar hypothesis converge in their results, intuitionism has two independent accounts of moral principles.

[334] It is certainly *possible* that the most accurate theory of descriptive adequacy is incompatible with central features of intuitionism and that intuitionists have to choose between (a) giving those features up in order to satisfy descriptive adequacy or (b) giving up descriptive adequacy itself. Before taking those drastic steps, intuitionists should however try to show that a descriptively adequate theory is able to accommodate their theory as fully as possible.

[335] See Mikhail (2011), 91-93 and 124f.

[336] See Mikhail (2013), 27. This implies that the deontic judgments Mikhail has in mind assign all-things-considered oughts – otherwise, the rules wouldn't by themselves be sufficient to arrive at a final statement as to whether an action is morally allowed or forbidden.

[337] Exceptions are Costless Rescue and Better Alternative (see Mikhail (2011), 108).

general rule that decides which principle trumps the other; even though the same

morally relevant features always count in the same direction (i.e. in favor or

against doing the action), their weight cannot be predicted through a general rule

and depends irreducibly on context.[338] To mark this feature, Ross calls his moral

principles '*prima facie* duties.'[339] The ability to weigh conflicting *prima facie*

duties against each other is an irreducible part of his theory. Therefore, a device

able to receive a situational description as input and to apply it to a deontic rule

would in most situations (i.e. wherever more than one *prima facie* duty is relevant)

be unable to determine properly the deontic status of an action. Which picture of

moral deliberation is descriptively more accurate – Mikhail's or Ross'? I will

argue that the adaptation of *prima facie* duties is a natural extension of Mikhail's

theory once the selection of cases to be explained by the moral grammar has been

extended.[340]

---

[338] This is yet another disanalogy between moral psychology and linguistics where a conflict between grammatical rules is not a frequent feature of natural languages.

[339] At several places throughout his (2011), Mikhail uses the notion of *prima facie* wrongs (39, 120, 143, 149f, 167, 173, 311, 337 and 340), but his application of the term differs from Ross'. At p150, he explains that some *prima facie* wrongs (knowing homicide and knowing battery) but not others (purposeful homicide and purposeful battery) can under certain circumstances (those defined in the doctrine of double effect) be morally permissible. Ross, in contrast, uses the term in order to indicate that there is *no* codified way to determine priori to encountering a specific situation whether an act to which a *prima facie* duty applies is permissible.

[340] It is important to separate the present issue from questions about the correctness of the closure rule or the variability of moral judgments among agents. Ross himself believes that his list of *prima facie* duties (or a complete version thereof) identifies all morally relevant features. Therefore, he implicitly endorses Mikhail's closure rule. At the same time, Ross' model of moral deliberation might provide a further explanation for the variability of moral judgments (see below), but as much as Mikhail's theory, it is vulnerable to variations that escape an explanation in terms of a set of common moral rules. Hence, both face the same fundamental problem. Instead, what is at stake here is the understanding of moral principles: do we need judgment in order to apply and weigh them against other duties, or is it the case that like in a game of chess, only one duty applies to each case (weaker, it might be possible that two rules compete, but that those rules contain provisions regulating which of them takes precedence)? In other words, the question is whether moral deliberation needs *prima facie* duties.

On the face of it, Mikhail's understanding of how we come to form moral

judgments looks unattractive. We don't value wisdom, moral expertise or a sharp

eye for what is right because those traits indicate that their bearers possess a more

complete and accurate set of moral principles or because they are better at

matching the features of the case with the right deontic rules. A failure in those

respects disqualifies for counting as a moral exemplar, but their presence is usually

not a reason for praise. Rather, we applaud people who know how to handle

difficult cases where competing moral reasons pull into opposite directions, agents

able to spot when a discussion gives short thrift to an argument at the expense of

other claims or individuals who make us see that an alternative way of balancing

moral considerations is as valid as our own.[341] All this is possible only if more than

one rule can make a legitimate claim on a particular situation and if the solution of

such conflicts has often no obvious solution, but requires skill and experience.[342]

How can this seeming incompatibility between Mikhail's empirically informed

view of moral deliberation and these deep-seated views about the nature of

morality be reconciled? Two replies are available to Mikhail, the first dismissing

the criticism by limiting the nature of the project to the explanation of a narrowly

---

[341] Those activities can reflect the grounds for our judgments (instead of being a mere exercise in moral dumbfounding) only if the distinction between express and operative reasons is – as argued above – more fluid than Mikhail suggests.

[342] Interestingly, even philosophers who are otherwise sympathetic towards Mikhail's research program reject this aspect of his theory. Adam Smith for example, whom defenders of the moral grammar hypothesis consider as a precursor of their view (see quote above), believes that rules alone are insufficient for arriving at moral judgments. After suggesting in the above quoted passage an analogy between the rules of grammar and justice, he claims that "there are no rules by the knowledge of which we can infallibly be taught to act upon all occasions with prudence, with just magnanimity, or proper beneficence: though there are some which may enable us to correct and ascertain, in several respects, the imperfect ideas which we might otherwise have entertained of those virtues" (Smith, 1759/2002, 205). Even though Smith does not further elaborate on this assertion, one way of developing his thought would be to argue – in line with my criticism of Mikhail – that rules alone lack the capacity to apply themselves to different contexts in a way that does justice to context and to the other duties that are also present.

circumscribed set of data, and the second suggesting how more complicated cases

could be handled without giving up the mechanistic character of his view.[343]

The first reply is straightforward, but does not solve the issue. For the

interpretation of the scenarios that Mikhail considers, no appeal to a conflict of

duties and consequently no description of our ability to weigh them against each

other are required. Mikhail is aware that he is merely offering a "fragment of a

moral grammar"[344] and that the analysis of scenarios outside the scope of his

limited model "is not necessary for our purposes."[345] This is not an oversight, but

results from a deliberate focus on a range of simple cases. In line with the

scientific approach Mikhail is adopting, he first considers scenarios that minimize

the complications he has to deal with. Once the fundamental problem – the

demonstration that mature moral agents possess a moral grammar – has been

addressed, further research can expand the theory to more complicated cases. As a

result, even though there might be aspects of moral deliberation which his account

leaves out, Mikhail can reply that this criticism is based on a misunderstanding of

his project. This response might preserve the letter of his theory, it leaves in the

dark those who want to know whether the account preserves the phenomena

without waiting until eventually, additional empirical research arrives and allows

to settle this question. Moreover, even if Mikhail's reply works against the charge

that his list of deontic rules is incomplete, but is less convincing against the worry

---

[343] Mikhail does not explicitly organize the discussion this way, but comparing his various responses to the problem, those two strands of argument emerge.

[344] Mikhail (2013), 25.

[345] Mikhail (2011), 140. For comments to the same effect, see also pp7, 85f, 103f, 137 and 149 Fn8; see also his (2013), 17.

that his account lacks the ability to account for a wide range of other cases. After all, adding more duties to his theory leaves the framework intact, but if its basic assumptions about the nature of moral reasoning cannot be generalized to include other cases, then the whole approach becomes far less interesting. Mikhail's second reply addresses this issue. In response to a criticism by Zimmerman, he attempts to show how scenarios with apparently conflicting duties can be accommodated by his theory. Zimmerman's challenge consists in presenting two kinds of cases to which the doctrine of double effect applies, but where it is trumped by a competing moral consideration.[346] First are cases of self-defense in which the attacker is being killed as a means to save one's life (and there is no alternative solution available). According to the doctrine of double effect, this act is impermissible since the bad effect is the means for achieving the end; most people would however judge the act as morally permissible. Second, Zimmerman suggests that if the killing of a person were necessary in order to save a very large number of other people, the majority of agents would judge this act as permissible, even though, as in the previous case, it would be forbidden according to the doctrine of double effect. In order to accommodate those cases, the doctrine of double effect (and in extension the rest of Mikhail's deontic rules as well) should be given a "comparative form,"[347] weighing it against other relevant normative considerations before arriving at an all-things considered judgment. This, though, does not fit with Mikhail's account of moral deliberation: "[I]f the doctrine is

---

[346] See Zimmerman (2013), 12f.
[347] Zimmerman (2013), 13.

given a comparative form, it cannot serve as the major premise in a modus ponens inference."[348]

As a point of agreement between him and Zimmerman, Mikhail concurs that the theoretical costs for his approach would be substantial if principles were to be assigned variable weights.[349] In arguing against this possibility, he discusses the cases of self-defense and extreme emergencies separately, but pursues a common strategy for both cases: counter-examples to a deontic principle should be treated as exceptions to moral principles.[350] By incorporating those situations into the formulation of the deontic rule, their occurrence becomes an instantiation of the rule and ceases to be a counter-example.[351] For extreme cases, Mikhail suggests alternatively a "supreme emergency exemption" which suspends under appropriate circumstances the normal operation of the doctrine of double effect and replaces it by a consequentialist reasoning. A similar proposal argues that under extreme circumstances, our moral compass is under "shock and awe" and commits performance errors.[352]

Overall, this approach reduces Mikhail's problem that there are *kinds of rules* his account leaves out to the already familiar and solvable issue that in order to do justice to a larger set of cases, *further and more-fine grained rules* have to be added to the moral grammar hypothesis (the supreme emergency exemption will

---

[348] Zimmerman (2013), 13. This, I take it, is just an alternative formulation of my criticism above that Mikhail's schema of moral deliberation does not allow for a judgment about the comparative weight of a the various duties involved in a situation.
[349] Mikhail (2013), 19.
[350] See Mikhail (2013), 16-19.
[351] Mikhail (2013), 18: "[O]ne can always supplement a theory with one or more ''unless clauses'' or auxiliary hypotheses to handle putative counterexamples."
[352] Mikhail (2013), 19.

be considered separately). This move comes however at a cost. In order to solve

the problem, it has to be repeated for every possible case that could be construed as

a conflict between a deontic rule and other relevant considerations. And a simple

lexical or cardinal ranking of rules wouldn't be enough: depending on the

circumstances, deontic rules have to be assigned different degrees of strength,

giving raise to an enormous, perhaps infinite amount of configurations to be

incorporated in the rule-system. If it is the case that most choices we are

confronted with in everyday life involve more than one moral obligation and if the

exact weight of those obligations differs from case to case, then the complexity of

a moral grammar accounting for all of those complications would resemble a

linguistic grammar that had a different rule for almost every possible sentence.

Thus, the complexity of the moral grammar would increase dramatically from

Mikhail's original proposal, making it doubtful that such a system could be stored

by the finite capacity of the human brain.[353] The point of this is not that the most

accurate formulation of the moral grammar shouldn't include specifications which

limit or describe the application of deontic rules in specific circumstances; the case

of self-defense is a good example where a specification might solve the issue.[354]

Rather, the lesson is that this strategy cannot solve the problem in its totality.

---

[353] This would also raise problems for nativism which is probably the most plausible explanation of how humans have come to possess a moral grammar: how could a process of natural selection result in such an extremely fine-grained system? After all, each part of the rule would have to be selected for separately, resulting in an extremely cumbersome selective process.

[354] It does not seem that in the case of self-defense, the fact that I am allowed to kill the attacker if necessary as a means to saving my life is the result of a weighing of the doctrine of double effect against my right to self-defense. Rather, the fact that I am attacked silences competing considerations that count against my right of self-defense.

It might be tempting to reply on behalf of Mikhail that he could limit the role of the moral grammar by deferring the assignment of specific weights to duties to the agent making the moral judgment, but unfortunately, this reply is not available to him. The aim of his "traditional deductive model of moral judgment" is that "case-specific deontic verdicts can be generated by covering laws,"[355] and this implies that the whole process of moral reasoning needs to be codified in the deontic rules without remainder left over to the agent (or the machine executing the rules of grammar).[356]

What about Mikhail's suggestion that under extreme circumstances, we are often willing to suspend rules that explain our intuitions in normal situations? Even this solution for a small set of exceptions does not seem convincing. There are situations where the violation of a duty that is minor compared to the huge and significant benefit it enables does not trigger consequentialist intuitions, for example in Le Guin's famous story of a city whose happiness depends on the suffering of an innocent child.[357] Hence, even if some extreme cases can be explained in the way Mikhail proposes, his idea cannot serve as a generalizable rule.

---

[355] Mikhail (2013), 19.

[356] To be exact, the first step in the process that leads to a deontic judgment is left to the agent: the establishment of which morally relevant features are present in a given situation. The rest – which rule(s) apply, how they have to be weighed, and what the final judgment is – can be executed by following the rules of moral grammar.

[357] "They all know it is there, the people of Omelas […]. They all know it has to be there […] [T]hey all understand that their happiness, the beauty of their city, the tenderness of their friendships, the health of their children, […] even the abundance of their harvest and the kindly weathers of their skies, depend wholly on this child's abominable misery. […] If the child were brought into the sunlight out of the vile place, if it were cleaned and fed and comforted, that would be a good thing, indeed; but if it were done, in that day and hour all the prosperity and beauty and delight of Omelas would wither and be destroyed. Those are the terms." Le Guin (2000), 38f., quoted from Sandel (2011), 27. Anti-consequentialist intuitions in extreme situations can also be invoked by considering surgeon cases where a large number of people could be saved by cutting up a healthy patient.

The challenge of explaining conflicts of duties brings out a trilemma in Mikhail's theory. These are the three claims his account needs to be in line with:

1. The "traditional deductive model of moral judgment"[358] expressed in his scheme of moral reasoning[359]

2. A moral grammar whose content is simple enough to be stored by the finite cognitive capacities of the human brain

3. A system of deontic rules that is complete, i.e. there are no idealized moral judgments it does not cover[360]

The first claim is a commitment Mikhail himself makes (see above). The second claim is an assumption that he is implicitly committed to: he repeatedly argues that his "provisional account"[361] could successfully be extended to other cases, and this is only possible if a complete theory lies within the boundaries of human capacity. This also motivates the third claim.

If, as argued above, the traditional deductive model of moral judgment entails that no further mental operations are necessary in order to reach a moral verdict (claim 1) and a moral grammar covers all possible complications (claim 3), then the second claim cannot be upheld since the resulting system of rules would exceed our mental limitations. This, however, would defeat the very purpose of the moral grammar hypothesis. Instead, either the first or the third claim has to be rejected. Rejecting the third claim would save the theory, but also at a very high

---

[358] Mikhail (2013), 19.
[359] See Mikhail (2011), 124.
[360] The term "coverage" used in 3. is deliberately ambiguous: it can either mean "providing all the information needed to explain a moral judgment" or "providing all the rules that judgment needs to apply and potentially weigh against other considerations in order to arrive at a moral judgment."
[361] Mikhail (2011), 140.

cost: if only a fairly limited range of cases would be covered by the hypothesis, this limited explanatory power wouldn't suffice for a general account of moral intuitions. It is this promise, however, whose fulfillment would advance empirical psychology. The best option, it seems, is to consider dismissing the first claim. Mikhail is firmly committed to it, seeing himself as executing a research program envisioned by a long tradition of thinkers.[362] There are however, as outlined above in the context of Adam Smith's quote, deep-seated beliefs about the nature of morality that are hard to reconcile with the idea that our moral intuitions could be generated (possibly in a superior way) by machines carrying out a moral code. I will therefore pursue the option of rejecting the first claim and consider what follows from this for the nature of moral deliberation that the moral grammar hypothesis has to assume.

The reason for the overbearing complexity of the moral grammar in case that the first and third claim are taken to be correct is that the system of rules needs to provide two kinds of information at once: given a description of the situation, an agent in possession of the moral grammar knows (a) what rules count in favor or against the action and (b) what their relative strength is. In comparison to the formulation of the rules themselves, the codification of how they interact with

---

[362] The passage where Mikhail explains his commitment is worth quoting: "The question that concerned me in this part of the book can be put this way: what rules and representations must be presupposed in order to explain ordinary, intuitive moral judgment? If one were tasked with programming a machine, say an autonomous robot, so that it is capable of producing the very same moral judgments that ordinary individuals make in response to trolley problems (taking the particular set of trolley scenarios as input and delivering a particular set of deontic judgments as output), what background information, instructions, and inference rules would the machine need to possess? Can one, in effect, show that Locke, Leibniz, and other Enlightenment philosophers were correct to assume that moral judgment is computational?" Mikhail (2013), 28.
Bernard Williams on the other hand is more skeptical about historical support for a position like Mikhail's. See his (1995), 184.

other moral principles is far more complex.[363] Being relieved from the obligation

to subsume all elements of moral deliberation under deontic rules while having to

reduce the size of the moral grammar, a natural move is to give up the goal of

finding a rule system that can determine in advance the weight each morally

relevant consideration has. Instead, it is a possibility that is – given the apparent

failure of a rule-guided system – worth taking seriously to assume that this task is

performed by a moral faculty whose operation, making a judgment about the

relative weight of *prima facie* duties, cannot be captured by a codified system. In

his contribution to the discussion about the computational theory of mind, Hubert

Dreyfus has argued that parts of our mental operation, especially those having to

do with expert knowledge, cannot be captured by an algorithmic procedure.[364]

Moral judgment, especially in the idealized form Mikhail is interested in, would be

a paradigmatic case of such a cognition that cannot be computed. While it is not

feasible here to engage with the relevant discussions in the philosophy of mind,

even a superficial look at the relevant literature reveals that it is not implausible to

assume that there are higher cognitive functions which cannot be reduced to an

algorithm.

     As a result of this changed understanding of moral rules, Mikhail's scheme

for describing the process of moral deliberation stops being a "logically closed

---

[363] As an example of how relatively simple the formulation of a deontic rule can be, consider Mikhail's discussion of his various principles. See his (2011), 132-152.

[364] See Dreyfus (1972).

system"[365] and needs to make room for the faculty of moral judgment. It has be revised along the following lines:

Prima facie reason[366] *P applies to A = A has features F1…Fn*

Prima face reason Q applies to A = A has features F1x…Fnx

Etc.

Agent X assigns a relative weight to prima facie reasons P, Q…

Agent X determines the all-things-considered judgment

——

A has deontic status D

In case only one *prima facie* reason bears on a situation, there is no need for the exercise of moral judgment, as the all-things-considered judgment about the moral permissibility of the action follows directly. This model offers distinct advantages over Mikhail's pictures. If Ross is right at least in principle and the number of irreducible *prima facie* duties is low,[367] the complete moral grammar does not face the challenge of being too complicated to be accommodated by a limited cognitive system or to have developed in a selective process that usually settles for economic solutions. In addition, once all fundamental moral considerations (more about their existence below) have been added to the moral grammar, this relatively sparse technical apparatus can explain a much wider range of cases than Mikhail's original theory (which was only designed to account for

---

[365] Mikhail (2011), 124.
[366] The rules that form the content of this minimal moral grammar bear close resemblance to Ross' notion of *prima facie* duties. For present purposes, they can be treated as identical.
[367] This allows for the actual content of the moral grammar to vary in content and to a certain extent in number from Ross' own theory.

some trolley-cases). And pace Mikhail, this theory does not support the implausible claim that complex and context-sensitive moral judgments can be reduced to an algorithm executable by a machine. Finally, the addition of *prima facie* duties to the moral grammar hypothesis provides a further explanation for moral disagreement: even though all test subjects abide by the same moral grammar, they differ in their assignments of strengths to the *prima facie* duties at stake.[368]

Mikhail's theory is only convincing to the degree that it is supported by empirical data, and the same is true for the modified moral grammar hypothesis. But how can the existence of a faculty whose operation is uncodifiable be shown to exist by analyzing response patterns to moral problems? If any degree of disagreement can be explained by reference to a difference in judgment between people, even a data sample with a random distribution of responses to moral problems serves as evidence for the modified moral grammar hypothesis. A theory that is compatible with everything explains nothing. However, there is indirect empirical evidence that can be appealed to. For simple cases that involve only one duty, Mikhail's theory is still valid and predicts the intuitions of idealized subjects.[369] In line with those findings, there are two ways of expanding this view to intuitions regarding more complex scenarios where more than one moral duty is relevant: either, Mikhail's claim that deontic rules contain all information necessary to deal with conflicts of duty is correct, or it turns out that this additional

---

[368] This does not replace the other possible reasons for disagreement discussed above, but only adds another source of explanation.

[369] This assumes that Mikhail is able to address the worry about the variability of moral judgments.

information is not contained and an additional capacity of judgment is necessary in order to deal with those cases. Mikhail's claim, it has been argued, is flawed. This leaves the second and less ambitious option as the remaining alternative.[370] In addition to resulting from a process of elimination, it is also in line with our common sense intuitions about moral deliberation (as argued above). Hence, even if there is no direct empirical evidence for the theory, there are still good reasons that point towards its plausibility.

Before concluding this chapter, I want to quickly point out how the present theory is able to solve a problem that Bernard Williams has raised for intuitionism:[371] what is the connection between what he calls the methodological and the epistemological doctrine of intuitionism, the first being the view that there is an irreducible plurality of first moral principles and the second a theory about how those rules are being known? Williams considers two ways of spelling the epistemology of ethical intuitionism, one in analogy to mathematical perception and the other in analogy to sense perception. For both cases, he argues that no close relation between the methodological and epistemological aspects of intuitionism can be established. If the modified moral grammar hypothesis is

---

[370] Should it turn out that there are further possibilities available, the modified moral grammar hypothesis would have to compete with them.
[371] See Williams (1995).

correct, there is a close connection between these two views: it is only by analyzing moral judgment that we can discover the structure of the rules that explains how we come to know them.[372]

4.12 Conclusion

Returning to the empirical predictions that follow from Rossian intuitionism, many of them have at least partially been vindicated by Mikhail's work on the linguistic analogy. To recall, here is the list again:

1. Despite disagreements about concrete issues, mature moral agents at least implicitly accept the same fundamental morally relevant features and apply them in their deliberations.

2. There is a plurality of basic morally relevant features, i.e. the features cannot be further reduced to a single unifying principle.

3. Those features can be expressed in the form of *prima facie* principles: the moral rules determine whether the morally relevant features count in favor or against an act, but not how various principles interact with each other.

4. The *prima facie* principles that underlie the judgments of morally mature agents bear resemblance to the items on Ross' list of seven *prima facie* duties (fidelity, reparation, gratitude, non-maleficence, justice, beneficence and self-improvement); minimally, the rules

---

[372] This presupposes that we also have an independent account of justification for our moral grammar; otherwise, it is possible, that we believe our moral judgment, but we do not know them since knowledge presupposes truth.

that actual mature agents follow should contain both consequentialist and deontological principles.

5. Mature agents (a) accept that the *prima facie* duties are not in need of further justification and (b) are strongly reluctant to give them up.

The first three and the fifth claim are, if the foregoing analysis is correct, confirmed by the evidence. The forth item is partially in line with Mikhail's finding, and the remaining lack of resemblance between the empirical findings on Ross' *prima facie* duties might be explained by the limited selection of the test cases. Future research, it can be hoped, might close the gap. It can be concluded that Ross' view of the rules that govern moral deliberation is largely correct.

Even though Mikhail's moral grammar hypothesis and Ross's ethical intuitionism could hardly differ more in their methodological assumptions, both share the same goal: the construction of a descriptively adequate account of moral judgments. The purpose of this chapter was to bring together those two theories and to show that they supplement each other. The combination of a naturalistic approach to moral epistemology with Ross' understanding of moral obligations as *prima facie* duties might increase the appeal of ethical intuitionism to those skeptical of foundationalism and the notion of self-evidence, but it can also serve supporters of Mikhail's project critical of his mechanistic picture of moral deliberation.

As argued at the beginning of this chapter, descriptive adequacy is only one

side of the coin for ethical intuitionism; in order to uphold the claim that our well-considered moral judgments reflect a moral reality, Rossian intuitionists also need to provide an account of what makes us justified in believing them.

Chapter 5: The justification of moral intuitions

5.1 Introductory Remarks

A complete moral epistemology, I have argued in the last chapter, consists of two parts, one providing a descriptive account of how our moral faculties work and the other justifying its ability to gain knowledge. It is this second task that is the subject of the present chapter.

The justificatory project differentiates linguistic from moral knowledge: once we have given an account of how people become competent speakers in a certain language and can describe what the rules of language consist in, there is no further question to be asked. Grammatical rules do not have truth values (although our descriptions of them do). In moral theory however, we face the additional task of validating our moral intuitions: in virtue of what are we justified in holding those judgments? Moral realists, among them most ethical intuitionists, claim that moral judgments are made true or false by stance-independent moral facts. Do we have good reason to believe that our moral judgments grasp them? That is the question of justificatory adequacy.

The descriptive and the justificatory part of moral epistemology pose different questions, but their answers are not unrelated. As Ross had already discussed in *The Right and the Good*, the etiology of our ethical beliefs can spell trouble for their claim to truth. On the one hand, he argues that if a belief is derived from other beliefs that are true, and the process of derivation can be shown to be valid, this establishes the validity of the belief in question. However, if a belief

such as an ethical judgment is being held for reasons that are not truth-conductive, for instance

> if any one can show that *A* holds actions of type *B* to be wrong
> simply because […] he knows such actions to be forbidden by the
> society he lives in, he shows that A has no real reason for believing
> that such actions have the specific quality of wrongness.[373]

While such a fact does not establish that the belief is wrong, Ross continues to argue, it undermines our justification for believing it. This, in a nutshell, presents a debunking-argument against moral realism that has been at the forefront of debates in moral epistemology over the last decade and that needs to be addressed by intuitionists who claim that many of our moral judgments correspond to a mind-independent realm of moral facts.

Ross himself was not too worried about this challenge. In one paragraph, he dismisses the problem. We are justified in our judgment, he argues, that some moral codes come closer to the truth than others. The reason for our confidence is that we have "an *a priori* insight into certain broad principles of morality, and it [the human mind] can distinguish between a more or less adequate recognition of these principles."[374] Again, Ross is anticipating an important argument in the contemporary debate (to which we will return later), even if his way of spelling it out is, as has been seen in chapter 1, open to serious objections.

Many would disagree with Ross. It is, critics worry, harder than he believes to make a convincing case that common sense moral convictions – such as those captured in the list of *prima facie* duties – are formed by mechanisms which are

---

[373] Ross (1930), 14.
[374] Ross (1930), 14.

truth-conductive. Absent such mechanisms and given the incredibly large number of possible beliefs that a non truth-tracking process could have lead us to adopt, it would be a cosmic coincidence if our moral beliefs matched the stance-independent moral truth that moral realists postulate. Since it is an essential component of ethical intuitionism that our well-considered moral judgments give us access to the moral facts, skepticism about the reliability of our intuitions threatens the viability of the intuitionist project as a whole. Before a positive account can justify our moral intuitions, debunking worries need to be defeated.

In what follows, I will present two types of debunking arguments, consider and criticize various responses made by ethical intuitionists and present what I believe are the most promising replies. It would be beyond the scope of this chapter to present conclusive arguments against the debunking worry; all I can do is to point towards some options on behalf of the intuitionist. In my opinion, debunking worries offer the most serious contemporary challenge for ethical intuitionism. If the case can be made that the situation is not hopeless for Rossians, this would already be an achievement. As will become clear during the discussion, most of the replies can also contribute to formulate a positive account of justification for the reliability of moral intuitions.

After giving a general account of debunking worries, I will first discuss and criticize an argument by Walter Sinnott-Armstrong to the effect that psychological defeaters stand between the exercise of moral judgment and moral truths. Then, I turn to evolutionary debunking arguments, advanced by Sharon Street and others.

5.2 Two Forms of debunking Arguments

The purpose of debunking arguments is to show that a causal explanation of how we came to hold certain beliefs undermines their justification. Guy Kahane helpfully analysis the basic structure of those arguments in the following way:

> *Causal premise.* S's belief that p is explained by X.
> *Epistemic premise.* X is an off-track process.
> Therefore
> S's belief that p is unjustified.[375]

The mere existence of a causal explanation, he emphasizes, is unsurprising and does not cast doubt on a belief. The challenge only arises if the causal process does not track the truth.[376] To illustrate this point, imagine a group of math students. At the end of the year, all have to take a quiz that tests their knowledge in the subject. Here is how most students arrived at their answers: they went to class, followed the teacher's explanations, did their homework and prepared for the quiz by participating in a study group. During the quiz, they solved the equations by applying their knowledge of the rules to the questions. This causal history does not defeat the students' beliefs about the answers; after all, their understanding of the mathematical principles involved explains why most of their answers should be correct. One student however never showed up to class, didn't do the homework nor prepared he for the quiz. Instead, he writes random numbers and symbols onto

---

[375] Kahane (2011), 106. For a sketch of a similar reconstruction of debunking worries, see also Street (2006), 155.

[376] Kahane claims that the causal process has to be such that "it leaves *no space* for the contribution of processes that would, in this context, track the truth" (2011, 106, my emphasis). This is too strong. Even a process that makes it merely unlikely or offers good but inconclusive reasons to doubt that the process is truth tracking generates a defeater for the belief. If, for example, a medical screening device produced the right results 55% of the time, a diagnosis based on the device expressing whether somebody had a certain disease (and not whether it is likely that he has it) would be unjustified, even if the device is more reliable than chance (i.e. truth-tracking to a slight degree).

his sheet. Here, the causal origin of the student's behavior reveals that his beliefs (assuming he takes his answer to be correct) are not based on a truth-tracking method. This defeats any justification he might claim to have even if by chance, his solutions to the math-problems were correct. Given all the possible answers, we could call it a cosmic coincident if he was lucky. According to debunking arguments, moral beliefs resemble in this respect the answers of the lazy student. Due to their unreliable process of belief-formation, we are in the dark whether they assert the moral truth or not.

There are numerous causal factors – historical, sociological, psychological, economical, evolutionary and others – which impact the formation of our moral beliefs, but not all of them make good candidates for debunking arguments. In order to fulfill the epistemic premise ("X is an off-track process"), the causal premise needs to be such that it undermines any truth-tracking mechanism. The reminder of this chapter discusses two promising candidates. Walter Sinnott-Armstrong advocates the first; in the formation of moral intuitions, he argues, various psychological factors are at work which render them unreliable as guides to the truth. The second debunking argument considers moral beliefs as the outcome of an evolutionary process. Since this process tracks survival but not stance-independent facts, we have no reason to assume that our moral beliefs correspond to the truth as the moral realist conceives of them.

While psychological debunking arguments make the case that there are factors (e.g. partiality or emotional arousal) which *interfere* with any truth-tracking mechanism, evolutionary debunking arguments make the stronger claim that *the*

belief-forming process itself is corrupted from a truth-tracking standpoint. Whether the defeater is external or internal to the belief-forming process, a main strategy on behalf of intuitionism is to identify beliefs that are immune from the defeater. Those beliefs can then be used to assess the veracity of the judgments under scrutiny and to correct them if necessary. This way, the moral realist (including the intuitionist) can defeat the defeater and restore justification in her beliefs.

5.3 Sinnott-Armstrong's psychological debunking argument

Walter Sinnott-Armstrong starts his criticism of intuitionism by locating the theory on the epistemological map. Instead of justifying moral beliefs on the basis of non-moral beliefs or by appealing to other moral terms, intuitionists claim that at least some moral beliefs are justified non-inferentially, i.e. "independently of whether the believer is able to infer those moral beliefs from any other beliefs."[377] For a belief to count as non-inferentially justified in the sense required by the intuitionist, it must not depend on the possession of a set of propositions which supports the belief.[378] All other moral beliefs derive their justification from non-inferentially justified core moral beliefs (for Ross, *prima facie* duties). Let's call these core intuitionist beliefs. They avoid the regress problem that those who justify moral beliefs in terms of other moral beliefs or rely on a weak notion of

---

[377] Sinnott-Armstrong (2006), 185. See also his (2011), 12.
[378] This contrasts with a weak notion of non-inferentiality where a believer possesses supporting beliefs, but does not need to go through them in order to be justified. See Sinnott-Armstrong (2011), 13.

non-inferential justification face:[379] if moral beliefs are justified in terms of other

moral beliefs, what in turn justifies those beliefs?[380]

Having defined the target of his argument, Sinnott-Armstrong introduces a

number of empirical premises about the process of forming moral beliefs designed

to cast doubt on the veracity of our moral judgments.[381] First, people disagree

about most moral issues. Assuming with the intuitionist that moral statements are

truth-apt, at least one of the competing standpoints in a moral disagreement must

be wrong. One reason for diverging moral opinions lies in the fact that human

beings are often partial, leading to judgments that favor one's side. Further,

emotions often infer with moral deliberation; depending on whether we feel e.g.

disgust, the likelihood that we morally condemn an action varies. Other non-truth-

tracking factors that affect our moral responses include order effects (our moral

judgments about a list of cases varies with the order in which we are presented

with the questions), the wording of questions, priming effects (reading e.g. words

that are associated with social power lets people adopt different moral judgments)

or sleep deprivation. In its totality, these and other similar effects provide evidence

that a large percentage of our moral judgments are distorted by non-truth-tracking

factors.

Can't the intuitionist accept with Sinnott-Armstrong that we have reason to

doubt many of our moral beliefs, but that the ethical judgments intuitionism relies

on are the outcome of a more reliable process of moral deliberation (such as

---

[379] See previous footnote.
[380] Sinnott-Armstrong also rejects the justification of moral beliefs in term of non-moral beliefs, but for other reasons; see his (2006), ch. 7.
[381] See Sinnott-Armstrong (2011), 15-17.

repeated careful deliberation, the absence of emotional arousal and sleep deprivation etc.) and are therefore more trustworthy? In order to block that move, Sinnott-Armstrong introduces two epistemic principles. First, he argues that if a belief falls into a certain group of beliefs who all possess a property with a given likelihood, this belief has the same chance of having this property. If one wants to assume that the belief falls into a different group, one needs reason to make that claim.

> It is reasonable to assign a probability that a case has a property based on knowledge that the case falls in a first class and that a percentage of cases in that first class has that property until one gains access to new information that the case in question falls within a second class that differs from the first class in the percentage of cases in that second class that has that property.[382]

It follows that if moral beliefs in general have a high chance of being wrong, any individual moral belief also has a high chance of being wrong.

The second epistemic principle holds if a person has reason to believe that a certain belief is probably wrong, that person is unjustified in maintaining that belief.[383]

Together with the empirical premises, the epistemic principles and intuitionist views about the non-inferential justification of basic moral beliefs, Sinnott-Armstrong formulates an argument that is immune against the intuitionist reply suggested above ("the moral judgments intuitionist are formed by processes that are more reliable than most other moral beliefs"). The argument can be reconstructed along the following lines:

---

[382] Sinnott-Armstrong (2011), 18.
[383] Sinnott-Armstrong (2011), 20: "If it is reasonable for a person to assign a large probability that a certain belief is false, then that person is not epistemically justified in holding that belief."

P1: Given the amount of non truth-tracking psychological processes that impact my moral deliberations, there is a high likelihood that set of moral beliefs is wrong. [Empirical premise]

P2: I need justification in order to claim that a certain moral belief(s) is part of a sub-set of moral beliefs that is less likely to be wrong. [First epistemological principle]

P3: Any belief which stands in need of justification is not non-inferentially justified. [Intuitionist' core belief]

Subconclusion: Therefore, I cannot be justified in believing that (according to intuitionists) non-inferentially justified moral beliefs are part of a set of beliefs which are less likely to be wrong than all other moral beliefs.

P4: If I know that a certain group of beliefs has a high likelihood of being wrong, I am not justified in upholding any of those beliefs. [Second epistemic principle]

Conclusion: I am not justified in believing any of my moral beliefs, including those which are according to intuitionists non-inferentially justified.

The very definition of what it takes for a belief to be non-inferentially justified prevents the intuitionist from arguing that it is not more reliable than the set of all other moral beliefs. This is a debunking argument in Kahane's sense insofar as it is the non-truth tracking nature of the causal process of moral belief-formation that causes trouble for the intuitionist.

5.4. A criticism of psychological debunking arguments

In response to this argument, the intuitionist can argue as follows: what does the work for Sinnott-Armstrong is his claim that nothing can be done to distinguish some kinds of moral beliefs from others in terms of their reliability without using a "justificatory inferential structure."[384] The absence of such a structure characterizes non-inferentially justified beliefs:

> For a believer to be non-inferentially justified in a strong way […] requires that the believer would still be justified even if the believer did not have as much as a disposition to accept any propositions that are, entail, or support any propositions that provide epistemic support for what is believed.[385]

The decisive question is whether evidence which establishes that some moral beliefs are not subject to the psychological defeaters Sinnott-Armstrong introduces in his empirical premise counts as considerations that 'are, entail, or support any propositions that provide epistemic support for what is believed'. In order to decide this issue, it helps to remind ourselves what the advantage of adopting non-inferential justification for moral beliefs are. Unlike competing theories, foundationalism is not at risk of being circular (like coherentist theories of justification) or to face the difficult task of justifying moral beliefs in terms of non-moral premises (like naturalism).[386] Instead, non-inferential justification serves as a regress-stopper: no further support is required in order to be justified.

It is important to note the scope of the skeptical arguments advanced by moral skeptics such as Sinnott-Armstrong. His arguments target various accounts

---

[384] Sinnott-Armstrong (2011), 24.
[385] Sinnott-Armstrong (2011), 13.
[386] For a detailed argument laying out the problems of the various forms of moral justification, see Sinnott-Armstrong (2006), chapters 7-10.

of moral knowledge, but he stops short of advancing a full-blown skepticism.

Whether there is reason to be skeptical about non-moral beliefs as well is a

separate issue and Sinnott-Armstrong's arguments against moral knowledge

presuppose the wrongness of such views. Otherwise, the empirical premise his

argument against intuitionist moral epistemology relies on would also lack

justification. But if empirical beliefs are allowed, why is it problematic to use them

in order to assess the reliability of a set of moral beliefs, as the intuitionist would

like to in his reply against the skeptic? Two reasons come to mind: either, this

move would justify moral beliefs in terms of non-moral beliefs, or it would start a

regress (since the empirical belief stands itself in need of justification etc.).[387] If

this were the case, intuitionism wouldn't have any advantage over naturalism or

coherentism. However, the intuitionist has a reply to both worries. Since Sinnott-

Armstrong himself even relies on empirical premises, he cannot believe that they

are unjustified. It follows that whatever their mode of justification is, he cannot

assume that they fall prey to an infinite regress either.[388] Second, the intuitionist

derives no normative conclusion from the empirical information about the

presence or absence of psychological defeaters. All it does is to provide the

information required by the first epistemic principle in order to distinguish a sub-

group of moral beliefs from other moral beliefs in regards to presence of

psychological defeaters. Taken by themselves, no moral belief follows from them;

rather, they protect the status of normative beliefs by establishing the absence of

---

[387] These reasons are not exclusive: empirical premises in support of intuitions could both involve a problematic inference from the natural to the normative *and* start a regress.
[388] For present purposes, it is not necessary to consider how non-moral beliefs can be justified, nor does Sinnott-Armstrong (2011) offer a theory.

psychological defeaters (which are themselves non-moral). Hence, the role these auxiliary beliefs play are different from potentially problematic justifications of moral through non-moral beliefs.

Holding those beliefs and using them in the process of justifying moral beliefs leaves intact what is attractive about non-inferential justification. Assuming that if left alone by psychological defeaters, we are able to make correct moral judgments (an assumption that Sinnott-Armstrong's argument does not question), and assuming that the intuitionist can make the case that the judgments she is concerned with are immune from those defeaters, it follows that those beliefs together with information about their reliability are sufficient for being completely justified in holding those moral beliefs; there is no regress. As seen in the discussion of Ross and Audi, there is indeed good reason for believing that the class of moral intuitions intuitionists talk about are free from the defeaters Sinnott-Armstrong introduces in his empirical premise: mature agents have carefully thought about the content of *prima facie* duties; there is little disagreement about them; detached and repeated reflection over time takes care of distortions caused by emotions, wording- or priming-effects and occasional sleep-deprivation.

This casts into doubt the third premise of Sinnott-Armstrong's reconstructed argument ("P3: Any belief which stands in need of justification is not non-inferentially justified.") What the intuitionist attacks is the undifferentiated use of the term "justification." While there is agreement that a moral belief that needs to be supplemented by considerations which "are or entail" the belief in question is not non-inferentially justified, Sinnott-Armstrong's inclusion of considerations that

"support"[389] a belief covers too much. Non-normative premises which are not used to derive the normative belief, but only clarify whether certain defeaters are present or not can play a role in the process of justification and they do not threaten the status of a justification as non-inferentially in the sense intended by the intuitionist since they don't change the fact that the justification serves as a regress stopper (which is what matters for the intuitionist).

What does this strategy in responding to Sinnott-Armstrong's argument suggest for the overall dialectic between debunkers and their opponents? While debunkers try to discredit moral judgments by showing that they are the result of untrustworthy causal origin, moral realists argue that even if these distorting processes exist, we can identify some moral beliefs which are not formed by corrupted processes; those beliefs can then be used to assess the other potentially misguided judgments. At least, they prevent debunkers from arguing for a complete skepticism in regards to moral beliefs. We will re-encounter this strategy in our discussion of evolutionary debunking worries.

5.5 Sharon Street's evolutionary debunking argument

Compared to psychological debunking arguments, evolutionary debunking arguments do not only focus on a different aspect of the causal origins of our moral beliefs, they also target more directly the heart of the moral realist's theory. While the psychological version of the debunking worry accepts (if only for the sake of argument) that our moral faculties are able to distinguish true from false moral

---

[389] Sinnott-Armstrong (2011), 13.

beliefs and undermines their justification by bringing out distorting factors that cloud our process of moral deliberation, evolutionary debunking arguments go a step further and attack the reliability of our moral faculty itself. The central idea is that the moral judgments we hold have been shaped by evolutionary forces and that those forces track survival and reproduction, but not stand-independent truths as the moral realists claims. Therefore, the realist (including the intuitionist) should be suspicious of those judgments. Before moving to a number of criticisms against this attack, I will develop Sharon Streets version of the argument.[390]

Street claims that natural selection has had a significant impact on our moral beliefs, arguing that "our system of evaluative judgments is thoroughly saturated with evolutionary influence."[391] After all, the evaluative judgments we hold can be detrimental or advantageous to our reproductive fitness and therefore, it is a natural assumption that selective pressures have preferred some belief systems over others.[392] In support of this assumption, Street observes that out of all the possible evaluative beliefs, human beings have adopted across time and cultures those that promote survival. For example, human beings commonly accept that we should do what promotes one's survival, that we have reason to promote the interests of family members, that "we have a greater obligations to help our own children than we do to help complete strangers"[393] to name only a few.[394] Why is it that we

---

[390] See Sharon Street (2006). Later, the discussion of Shafer-Landau's criticism of Streets' argument will bring to light some ambiguities in her formulation.

[391] Street (2006), 114. She also mentions rational reflection as a factor that shapes our beliefs, but I reserve discussion of this point for later.

[392] To be precise, Street only claims that "proto-beliefs," understood as "basic evaluative tendencies" have been selected for, not full-fledged beliefs (see 114 and 119f.). I will return to this point in a later section.

[393] Street (2006), 115.

[394] See Street (2006), 115.

haven't adopted the opposite beliefs? It is not contradictory to hold that we have

no reason to promote our own survival, that we have no reason to promote the

interests of family members etc., but no culture has ever sincerely believed those

judgments. A plausible explanation of this fact points towards evolutionary

pressures as selection process. Since these pressures are not sensitive to moral truth

but to survival, had the conditions under which this process had occurred been

drastically different, other kinds of evaluative attitudes would have been more

fitness enhancing and would have been selected for. The number of different sets

of evaluative attitudes that we could have been chosen under different

circumstances is almost indefinite. Street imagines that if we had developed under

circumstances that lions have encountered, we might see nothing wrong with

killing the offspring of our female partners that are not their own.[395]

This raises a question for moral realism: what is the relation between

evolutionary forces and stance-independent moral truths? There are two options,

but neither one looks good.[396] First, the realist can deny that a relation exists.

Darwinian pressures do not select for moral truth. This does not exclude the

possibility that our moral judgments pick out the truth, but there is no reason to

expect that they do. To use Streets' illustrative example, the situation is analogous

to a sailor who wants to go to Bermuda, but lets the winds and the tide determine

to course of the boat.[397] He should be skeptical about getting there, and to the same

degree, the moral realist should doubt the veracity of all of our moral judgments,

---

[395] For this and other examples, see Street (2006), 120f.
[396] See Street (2006), 121-135.
[397] See Street (2006), 121f.

even if stance-independent moral truths exist and could possibly coincide with our moral beliefs

The second horn of the dilemma asserts a relation between Darwinian pressures and stand-independent moral truth. According to this view, evolutionary selection is truth tracking. The reason, the realist could argue, is that it is advantageous to survival and reproduction to recognize the truth.[398] The advantage of his view is that it advances a testable scientific hypothesis. The disadvantage is that if fares poorly compared to other hypothesis, such as the adaptive link account[399] according to which certain evaluative judgments have proved to increase reproductive success. Such an account is able to explain many evolutionary adaptations, and considering the survival enhancing evolutionary judgments human beings have adopted, we have good reason to assume that it works in this case as well.[400] For the adaptive link account to work, it is not necessary that our evaluative judgments track the truth.

In conclusion, the evolutionary origins of our evaluative beliefs put pressure on the moral realist. It is hard to combine, so the argument goes, those selective pressures with the expectation that our moral beliefs are any more than tools to enhance our fitness. Even if possible, it would be a *cosmic coincidence* if our actual judgments actually lined up with the stance-independent moral facts the moral realist postulates. In order to avoid epistemic moral skepticism, the realist

---

[398] See Street (2006), 125f.
[399] See Street (2006), 127-129.
[400] Street lists three advantages adaptive link accounts have over truth-tracking accounts: they are more parsimonious, clearer and do a better job explaining why we prefer some judgments over others. See her (2006), 129-132.

would have to come up with an alternative explanation of why our moral

judgments track the truth. There are two ways of doing this: by rejecting the

dilemma or by embracing one of the horns (the replies I am discussing go for the

second option). Since the overall concern is to defend intuitionism, the viability of

moral realism is only of importance here insofar as it is required for the

intuitionist' project. Should it turn out that none of the realist' replies is

convincing, intuitionists could give it up in favor of anti-realist views.[401] This

would allow them to keep the normative part of intuitionism alive, and it might

also be compatible with accounts of non-inferential justification; after all,

debunking worries put only pressure on the claim that we are justified in believing

stance-independent moral truths.[402] However, there are good reasons for

intuitionists to prefer realism over anti-realism.[403] Switching metaethical views

should only be a last resort in case debunking worries prove to be irresistible.

5.6 Huemer's reply to evolutionary debunking arguments

In reply to evolutionary debunking arguments, Huemer's strategy is to

isolate intuitions that are immune from distortion and to use them in order to judge

moral judgments that are the outcome of a non truth-tracking process.[404] In order

to single out intuitions that "are most likely to be reliable,"[405] he identifies a

---

[401] For such a suggestion, see Bedke, (2009), 204.

[402] It would be a separate task to work out an account of epistemic justification for a non-realist intuitionism, but at least, such a theory wouldn't face debunking worries of the form discussed here.

[403] See, for example, Huemer (2008), 371f.

[404] See Huemer (2008), 370. He argues that his reply works against various kinds of debunking arguments (he explicitly states that his reply works against evolutionary debunking arguments and also refers to Street) but for present purposes, I will only consider whether his account works against the evolutionary blend of debunking worries.

[405] Huemer (2008), 383.

number of desiderata. Ideally, those judgments should "fit well together" and contribute individually to the plausibility of the set of beliefs. They should also be widely shared and accepted beyond one's own society. Judgments that favor behavior which promotes reproductive fitness are suspicious since they could also be the result of a Darwinian process. It counts against an intuition if it benefits oneself. The same goes for judgments that evoke strong emotions.[406] Apart from restrictions on content, Huemer argues that the judgments' level of generality matters. Abstract theoretical intuitions are to be preferred over concrete intuitions and mid-level judgments because those are most susceptible to bias.[407] They more readily evoke strong emotions, are more likely impacted by processes of natural selection ("biases with which evolution is most likely to have endowed us are biases favoring relatively specific forms of behavior that would have promoted our ancestors' inclusive fitness"),[408] and are more often culturally generated since cultures are expressed through specific rules of behavior. This leaves us with a set of abstract generalizations that Huemer calls "formal intuitions" which we have good reason to believe to be true (they count as "unusually trustworthy"):[409]

> (1) If x is better than y and y is better than z, then x is better than z.
> (2) If x and y are qualitatively identical in nonevaluative respects, then x and y are also morally indistinguishable.
> (3) 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.
> (4) If it is wrong to do x, and it is wrong to do y, then it is wrong to do both x and y.
> (5) If two states of affairs, x and y, are so related that y can be produced by adding something valuable to x, without creating

---

[406] See Huemer (2008), 381f.
[407] See Huemer (2008), 383-385.
[408] Huemer (2008), 384.
[409] Huemer (2008), 386.

anything bad, lowering the value of anything in *x,* or removing
anything of value from *x,* then *y* is better than *x.*
(6) The ethical status (whether permissible, wrong, obligatory, etc.)
of choosing (*x* and *y*) over (*x* and *z*) is the same as that of choosing *y*
over *z,* given the knowledge that *x* exists/occurs.[410]

Even though these intuitions don't contain normative content, they can be used to

evaluate ethical judgments. If a judgment violates any of those principles, we

should reject it. This way, we can accept that Darwinian pressures have left their

footprint in our moral inventory, but we possess independent criteria for judging

the veracity of our beliefs.[411]

5.7 A criticism of Huemer's reply

There are problems with Huemer's reply. First a clarificatory point: while a

case can be made that concrete intuitions fall prey to emotions and are often

influenced by culturally generated biases, those are not good arguments for

preferring abstract generalizations over mid-level intuitions. Are they especially

susceptible to emotions? Take some of Ross' *prima facie* duties. The claim that I

have a duty to keep my promises or that I should benefit others are not particularly

emotional statements. Further, it does not seem to be the case that *prima facie*

duties are tied to a particular culture. On those counts, abstract generalizations

don't enjoy an advantage over mid-level judgments. What carries the weight of his

argument is the claim that many of our concrete and mid-level moral intuitions

fulfill social functions and can therefore be considered as fitness enhancing. This

---

[410] Huemer (2008), 386.
[411] As the title of his paper indicates ("Revisionary Intuitionism"), Huemer believes that this method will diverge from classical intuitionism and question many common sense beliefs. Since this aspect of his argument is not relevant for the issue at stake, I will leave it aside.

makes them poor candidates for beliefs that are immune to evolutionary debunking worries.

The question is whether the formal intuitions that survive scrutiny are of much use. The problem can be set up as a dilemma. Assuming that only Huemer's list of abstract principles survives the process of eliminating beliefs that debunking arguments have rendered suspicious, there are two plausible options: either, no other than formal intuitions (and those beliefs that can directly be derived from them) are justifiable *or* beliefs that stand in no contradiction to formal intuitions are included in the set of justifiable beliefs.[412] The first option leads to moral skepticism (no moral beliefs can be derived from formal intuitions) and is therefore not an attractive option for an opponent of debunking arguments. The second option encounters the opposite problem. Too many beliefs are compatible with formal intuitions. As their name indicates, formal intuitions put almost no limitations on the content of moral beliefs; instead, they protect against inconsistencies in using those beliefs. For example, it does not matter which moral beliefs I hold, as long as I employ them in a way that respects various rules of reasoning such as transitivity ("If x is better than y and y is better than z, then x is better than z"), they don't violate any of the formal intuitions. The problem is that even beliefs that are clearly the outcome of a process of Darwinian selection and therefore not truth-tracking – e.g. the belief that I should prefer those who are genetically related to me over those who aren't – can pass this test. If Huemer

---

[412] The qualification "plausible options" is meant to exclude *ad hoc* possibilities that are logically possible, but hold no appeal, e.g. "beliefs that stand in no contradiction with formal intuitions and are compatible with consequentialism."

cannot rule those beliefs out, his argument is unable to distinguish between

intuitions that we can be justified in believing and those we don't. It is of no use

against evolutionary debunking arguments.[413]

5.8 Two problems for evolutionary debunking arguments

I want to advance two worries about evolutionary debunking arguments

which, even if they don't defeat the argument, might weaken its strength and show

the moral realist the direction for a way out. The first reply considers how much of

an accident it would if our judgments aligned with stance-independent facts even

granting that evolutionary forces have shaped those judgments. The second

argument considers whether evolutionary debunking arguments have shown that

moral reason is unable to override evolutionary pressures.

There are accidents and there are accidents, and some of them are worse

than others. To make their argument look as strong as possible, proponents of the

evolutionary debunking argument try to portray the accident they have uncovered

in the worst possible light. It would, they argue, be a *cosmic* accident, the least

likely event imaginable, if our evolutionarily shaped moral judgment were in line

with stance-independent moral truths as the moral realist claims. What justifies this

attribute is the incredible large number of evaluative beliefs that might have been

---

[413] Which horn of the dilemma does Huemer go? He gives two examples of beliefs that his theory is able to rule out. The first concerns sexual morality (especially our judgments about homosexuality, bestiality or polygamy) – see his (2008), 388. Those beliefs don't violate Huemer's abstract principles, but he argues that they violate his requirements for choosing benchmark beliefs that are used to rule out unreliable judgments (sexual moralities are often not the same among different cultures). Here, he seems to be adopting the first horn of the dilemma. His second example concerns "weak deontology" (see his (2008), 389f), a formulation of deontological ethics which, he argues, runs into problems with one of the formal intuitions. It is however not clear that this dispute has any bearing on concrete moral judgments.

chosen (any possibility that imagination can come up with) in comparison to the

one correct set of moral beliefs that corresponds to the realist' truth. However,

there is a complication in this picture. By Streets' own admission, "*full-fledged*

evaluative judgment with a given content – for example the acceptance of the

judgment that 'one ought to help those who help you'"[414] are not likely candidates

for natural selection because they are too specific to be genetically heritable.

Instead, what is genetically heritable are "*basic evaluative tendencies* [that] may be

understood very roughly as an unreflective, non-linguistic, motivational tendency

to experience something as "called for" or "demanded" in itself."[415] Those

'tendencies' have no propositional structure and can be shared with animals and

constitute e.g. "some kind of motivational "pull" in the direction of feeding […]

offspring."[416] It follows from this picture that evolutionary selective pressures have

had a "tremendous" direct influence on what Street calls "proto forms of

evaluative judgments,"[417] but only an indirect – and weaker – impact on our "full-

fledged evaluative judgments." This opens a gap between Darwinian pressures and

evaluative judgments. There are two respects in which the grip of evolutionary

forces on evaluative beliefs is weakened: there is not only room for more than one

factor to impact our moral judgments,[418] it is furthermore the case that the vehicle

---

[414] Street (2006), 118.
[415] Street (2006), 119.
[416] Street (2008), 119.
[417] Street (2008), 119.
[418] At places, Street accepts that evolutionary pressures might have a limited impact on our evaluative judgments, but at other places, she seems to take a much stronger view. Here is an example of the first view followed by a quote that instantiates her second position. "[…] I do not mean that we automatically or inevitably accept the full-fledged evaluative judgements that line up in content with our basic evaluative tendencies. Certainly not. For one thing, other causal influences can shape our evaluative judgements in

through which evolutionary pressures exert their influence on full-fledged evaluative beliefs, "evaluative tendencies," are much broader than beliefs with a propositionally structure and compatible with a wide range of possible belief sets.[419]

This allows for two criticisms against Street's argument. First, since evaluative tendencies are more abstract than our regular beliefs, these proto-beliefs can be translated in many different ways into full-fledged beliefs. This dramatically reduces their number; if proto-beliefs are (as Street suggests)[420] to be understood as those motivational tendencies that we share with higher developed animals, there might be barely more than a handful of them that we have adopted (e.g. motivational tendencies to take care of one's offspring, to be moved by suffering etc.) and not dramatically more counting also those that we have not adopted. Consequently, even if it was by chance that the proto-beliefs we hold were in line with a stance-independent truth, this outcome would be much less surprising than a cosmic coincidence. Street has however a reply. As long as our evaluative attitudes would have been different had evolutionary pressures caused us to adopt different proto-beliefs, our evaluative beliefs don't track the truth and evolutionary forces debunk the realist's claim to moral knowledge:

> This counterfactual claim is all I need for the purposes of my argument. While one might inquire into the exact causal process by which basic

---

ways that make them stray, perhaps quite far, from alignment with our more basic evaluative tendencies." (120); "[O]ur system of evaluative judgements is thoroughly saturated with evolutionary influence" (114).
[419] Street explicitly admits this: "Indeed, it is likely that we were selected above else to be extremely flexible when it comes to our evaluative judgements – not locked into any particular set of them but rather able to acquire and adjust them in response to the conditions in which we wind ourselves." (2006, 158 Fn. 20)
[420] See Street (2006), 117.

evaluative tendencies have influenced the content of human evaluative judgments, it is not necessary for me to enter into such questions here.[421]

This brings up a second criticism. Since by Street's own admission, Darwinian forces exert merely an indirect influence on our full-fledged evaluative judgments, it is not clear that her counterfactual holds. Despite her arguments, the details of the causal process matter, because *influencing the content of human evaluative judgments* might not be enough for the conclusion that this influence is *strong enough to change the outcome of the process of belief formation*. Given the distinction between proto-beliefs that are strongly influenced by evolutionary pressures and full-fledged beliefs that might be the result of various processes, all we know is that Darwinian selection has *some* impact on our moral beliefs, but it is unclear what role the other factors play. For the moral realist to score against Street, it is however not enough to merely point towards causal influences other than evolution. In order to challenge the debunking argument, the realist needs a truth-tracking process that is strong enough to defeat competing non truth-tracking influences on moral beliefs. Rational deliberation is such a candidate. Reason provides, so the argument goes, access to the right moral principles if properly employed. Furthermore, it is certainly not impossible that we accept the results of rational deliberation in the face of countervailing inclinations. This rationalistic view that intuitionists like W. D. Ross (and, more prominently, Immanuel Kant) would embrace does, even if it is far from uncontroversial, provide a serious challenge to evolutionary debunking worries. Street is aware of this argument. In

---

[421] See Street (2006), 158 Fn. 21. This is the counterfactual she refers to: "[H]ad the general content of our basic evaluative tendencies been very different, then the general content of our full-fledged evaluative judgements would also have been very different, and in loosely corresponding ways." (120)

reply, she offers a different view of the role rational deliberation plays: we "assess some moral evaluative judgments in terms of others."[422] Evaluative beliefs must be taken as a given and cannot be the result of the activity of reason. Therefore, if our evaluative beliefs are doubtful, rational deliberation cannot save them.

> [I]f the fund of evaluative judgements […] was thoroughly contaminated with illegitimate influence […] then the tools of rational reflection were equally contaminated, for the latter are always just a subset of the former. It follows that all our reflection over the ages has really just been a process of assessing evaluative judgements that are mostly off the mark in terms of others that are mostly off the mark.[423]

According to this picture, rational deliberation can influence our moral beliefs,[424] but only by choosing to adopt one already preexisting belief over another. If those beliefs are the outcome of a process that is not truth tracking, moral deliberation cannot fix that problem. This view of moral reasoning resembles a trolley driver who can choose which way the trolley takes in case the rails split, but who is unable to change the way the trails are set. Another view of moral reasoning also believes that rational deliberation can influence which moral judgments we accept, but has a stronger understanding of what that influence implies. This understanding compares to a car driver who, even though the steering wheel is out of line and steers towards the left, is able to correct for that tendency and to choose into which direction to drive. Shafer-Landau, for example, defends such a view. There are many cases, he argues, where we are able to transcend by reflection

---

[422] Street (2006), 124.
[423] Street (2006), 124.
[424] Throughout her essay, Street speaks of "evaluative judgments" whereas I am interested in moral judgments, a subgroup of evaluative judgments. For present purposes, I will ignore the question whether anything of importance hangs on that distinction.

beliefs that we have been pressured to adopt by various non-truth tracking cultural or historical circumstances.[425] He quotes the struggle to abandon slavery and the victory of female suffrage, and more examples come easily to mind (racism, sexism, speciecism etc.). Why, Shafer-Landau argues, should reflection fail when it comes to correcting for evolutionary influences? Much more would have to be said to turn this view into a complete theory, but for present purposes, it is plausible that this understanding of how reason works is – at least as it stands – not any less plausible than Streets' view. If rational deliberation is able to override the influences that debunkers appeal to, it is not clear that our well-considered judgments are unable to fulfill the role the moral realist believes they play.

From a larger perspective, this conflict over the role of reason in moral deliberation is part of the familiar debate between Humean versus Kantian conceptions of rationality with Humeans taking Street's side and Kantians the opposite position. This makes evolutionary debunking arguments both stronger and weaker. On the one hand, the argument has the resources of a well worked-out conception of how reason work, but on the other hand, those targeted by the argument can rely on an equally well developed conception of reasons that potentially offers an escape route.[426]

---

[425] See Shafer-Landau (2012), 19.

[426] It should be stressed that Kantian constructivists would reject the idea that there are mind-independent moral truths. However, even they would be friendly to the claim that the role of reason extends beyond the adjudication of various pre-existing moral beliefs. It would take a chapter – perhaps a dissertation – of its own to develop how different strands of Kantians would respond to debunking arguments, but it seems clear to me that they would reject Street's interpretation of moral deliberation, 12andand this is all that has to be established for present purposes.

5.9 Conclusion

Ideally, intuitionists should possess a convincing positive account that explains why we are justified in our deeply held moral beliefs. Such a theory would (a) demonstrate that those beliefs are truth-tracking and (b) show that our process of belief formation can withstand non truth-tracking pulls caused by e.g. psychological defeaters or evolutionary pressures. A solution that satisfies both (a) and (b) is however at short supply. As a first step towards a complete intuitionist epistemology, intuitionists should address the negative arguments brought forward by debunkers. After all, if there is no reply to those challenges, there is also no way to answer (a). This would undermine the attractiveness of classical intuitionism.

As I have tried to show in this chapter, intuitionists have a way to reply to two prominent kinds of debunking arguments: the intuitionist is able to isolate moral beliefs that are free from psychological defeaters, and he is also able to withstand the claim that Darwininan pressures render moral beliefs unjustifiable.

Chapter 6: Conclusion

Ross' ethical theory is flawed in many regards. Here are some of them. (1) His strong confidence in what seems self-evident lacks a sufficiently worked out philosophical theory. (2) The list of moral principles he proposes is of limited appeal to the modern reader: it omits notions of autonomy, respect or equality and contains items such as a duty of self-improvement that many would not consider foundational. (3) Many of his arguments against utilitarianism are outdated and have lost force against more recent versions of the view. (4) His claim that we can never have moral knowledge about duties all things considered seems exaggerated and is based on questionable assumptions. (5) His view that motives are irrelevant for the moral evaluation of the rightness of an act has few adherents.

Despite these problems, Rossian intuitionism is far from dead. In fact, it is seeing a revival, attracting an increasing group of defenders. The reason is that Ross and the philosophers he builds upon have deep insights to share that are drawn from close observations of how everyday moral deliberation proceeds. This sets intuitionists apart from their competitors who might have more unified and sophisticated theories, but whose views are also farther removed from common sense morality. Biomedical ethics, where intuitionism has become one of the main approaches for both moral philosophers and practitioners, demonstrates this point (appendix 2).

Three core elements of intuitionism help explain its attractiveness. First, Ross' claim that there is a plurality of irreducible first principles which consist of consequentialist *and* deontological considerations reflects a fact about morality that other theories, which derive all obligations from *either* consequentialist *or* deontological

174

principles have difficulty capturing. Second, Ross' argument that moral principles are *prima facie* in the sense he defines captures the complex and unpredictable interplay between conflicting moral considerations while it also takes into account that those principles are stable insofar as they always count in favor or against an action. Third, Ross' foundationalist moral epistemology assigns common sense beliefs a more central role than other ethical theories.

In this dissertation, I have traced the recent history of one central aspect of Ross' view, his moral epistemology. I started by presenting Ross' theory and its critics (chapter 2). After a period in which intuitionism was neglected, two dominant contemporary interpretations have emerged. On the one hand, Robert Audi's minimalist theory attempts to increase the appeal of intuitionism by minimizing its philosophical commitments and by connecting it to other normative theories such as Kantian ethics (chapter 3). On the other side, Jonathan Dancy pursues the opposite strategy of radicalizing Ross' view (appendix 1). He takes to the extreme Ross' rejection of rules that determine the weight of a morally relevant consideration by extending his argument to the conclusion that there are no moral principles at all.

Instead of settling for either one of these options, I suggest that intuitionists should separate two questions that are epistemically relevant – the first descriptive and the second normative. This separation is already implicit in the writings of Ross, but to the detriment of intuitionism, the descriptive issue has so far been neglected. The descriptive question asks whether people rely in their normative judgments on an underlying set of moral principles. If this is so, which are those rules (chapter 4)? I answer by drawing on recent research in moral psychology and in linguistics. Second,

there is the normative issue whether those empirically verifiable principles are also justified (chapter 5). Rather than responding to this problem head-on, I discuss a preliminary challenge against the possibility of such a justification that has been receiving much attention in the present debate. According to this argument, the causal origin of our moral beliefs undermines their credibility. On behalf of intuitionism, I argue that the problem is not as devastating as the debunkers make it seem.

I want to end by briefly mentioning three issues that lie beyond the scope of my dissertation, but which a comprehensive Rossian intuitionism would have to deal with.

1. On the normative level, opposition to consequentialism is a defining feature of Rossian intuitionism. It characterizes his view that not all *prima facie* duties are concerned with the consequences that an act has. Recent versions of rule utilitarianism, however, have found a way to incorporate deontic principles into their theory.[427] If those attempts were successful and able to account for common sense morality to the same extent as Rossian intuitionism, Ross would loose his advantage over certain forms of consequentialism. Further, the arguments against his strongest competitor would have lost its force. The same problem comes from Kantian ethics. If, as Audi argues, the second formulation of the categorical imperative is able to explain Ross' list of *prima facie* duties without remainder, some might be tempted to ask whether we still need Ross theory at all. Rather than offering additional support for the theory, the question becomes whether defending Ross' view offers any additional benefit over Kantianism.

---

[427] See e.g. Hooker (2003). The idea of grounding *prima facie* duties in utilitarianism goes back to Sidgwick. See Irwin (2009), 369 Fn. 38.

2. Is Kantianism the best theory to supplement intuitionism? One problem is that the second formulation of the categorical imperative that Audi relies on in his Kantian intuitionism may be too indeterminate to derive any concrete moral principles.[428] Another issue concerns divergences between Kant and Ross in regards to moral motivation.[429] Instead, it might be more promising to derive Ross' list of *prima facie* duties (or a more plausible successor of that list) from Aristotle's ethics. This would have various advantages: Aristotle's ethics fits better than Kant's theory with Ross' motivational externalism and his notion of *prima facie* duties. Further, Ross refers to Aristotle in his account of moral judgment. An explicit connection between the two views would allow deepening Ross' theory in this respect.

3. It would be interesting to pursue further the line of empirical research started by Mikhail about the considered moral judgments of mature agents. As seen, one of his main problems was to explain the variance of response patterns in response to trolley cases. This issue is important for determining whether there is a common sense morality that extends across all mature moral agents, an assumption that Rossian intuitionists are firmly committed to. Further research should explore whether there are moral rules that do not exhibit substantial variations in intuitions. How likely are agents who disagree with the majority on one issue to also disagree in response to other scenarios? It might also help comparing how other disciplines such as linguistics explain those aberrations.

---

[428] See Hurka (2007).
[429] See Darwall (2002).

Appendix 1: In Defense of *Prima facie* Duties

7.1 Introduction

In this essay, I provide a novel response to particularist attacks on the

Rossian conception of *prima facie* duties. This attack consists in challenging the

intuitionist idea that the valence of *prima facie* duties – whether they count in

favor or against an action, or not at all – is invariable. Particularists like Jonathan

Dancy argue that given the context-sensitivity of reasons, both the strength and the

valence of a *prima facie* duty depend on circumstances. In defense of a Rossian

intuitionism, I first consider Robert Audi's reply against particularism. I argue that

his notion of invariant valence is coherent, but too weak for intuitionism. Then, I

turn to a second line of defense: Sean McKeever and Michael Ridges' claim that

context-sensitivity is compatible with invariant principles. In response to

McKeever and Ridge, I try to show that their argument is flawed. Finally, I outline

a solution for the intuitionist that offers a robust notion of invariance of *prima*

facie duties by arguing that they contain thick ethical concepts.

7.2 The particularist's challenge

Moral particularists agree with ethical intuitionists that the strength a *prima*

facie duty carries in a certain situation depends on the context.[430] However,

particularists go further and argue that also its valence might vary from situation to

---

[430] A number of different theories are called "moral particularism." For the purpose of this paper, I take the label to refer to Jonathan Dancy's claim that the deontic valence of a consideration is determined by circumstances and is therefore not invariant (see especially his (2004)). For a helpful overview of the various forms of particularism, see Audi (2004), 69-74.

situation. What motivates their view is holism, a certain view of how reasons work. According to that view, the ethical impact of a consideration cannot be understood in isolation from the situation in which it occurs. Hence, reasons are "sensitive to context."[431] There is no good reason, the particularist claims, that allows us to draw a line between the variant strength and the allegedly invariant valence of *prima facie* duties.

Both Ross and recent ethical intuitionists like Robert Audi explicitly reject the particularist claim and hold an "invariant valence view," according to which e.g. the fact that an act injures another person gives always rise to a *prima facie* duty against performing the act.[432] This places a burden on the intuitionist to explain why the valence of *prima facie* duties is exempt from variability.

7.3 Audi's intuitionist defense

In order to vindicate the invariance of *prima facie* duties against particularism, Audi needs to explain cases in which it seems that a *prima facie* duty has been silenced or reversed. He offers various strategies of how this can be achieved. If I have promised to pick up a friend, but I discover that my car contains a bomb that will be detonated as soon as we meet, then the fact that I have made a promise plays no role in my deliberation since it falls "below the threshold of deliberative relevance."[433] This does however not mean that the normative function

---

[431] Dancy (1993), 89.
[432] Audi (2004), 71.
[433] Audi (2004), 71.

of the promise has been reversed: Later on, I owe an explanation to my friend.[434]

The particularist can grant these arguments, but insist that the intuitionist still owes us an explanation of cases like Schadenfreude (taking pleasure in another's misfortune) where it seems that the valence of a *prima facie* duty has been reversed.

In order to explain these cases, Audi distinguishes between two separate claims, one about intrinsic value and another about organic value.[435] According to his claim about organic value, the contribution an aspect makes to the moral status of an overall situation is determined by its relation to other elements; furthermore, a consideration that normally counts in favor of an experience can due to the context count against it:

> Pleasure in something intrinsically good may contribute non-additively and even negatively – both reducing the overall value of an experience and doing so by more than its own 'quantity'.[436]

Like Schadenfreude, sadism is an example of organic value: The pleasure somebody feels in inflicting pain on somebody else makes his experience worse from a moral standpoint.[437] According to Audi's claim about intrinsic value, bearers of intrinsic values – which are following his experientialism "qualitative mental states or processes"[438] – always count in favor of an action. Taken together, the claims about intrinsic and organic value allow Audi to respond to the

---

[434] In a similar vein, considerations can fall below a "threshold of ordinary visibility" without losing their valence: I am so overwhelmed by other factors that I don't take a certain consideration into account. Still, the valence of the consideration has not changed. See Audi (2004), 72.
[435] See Audi (2004), 122-138.
[436] Audi (2004), 129, but see as well 138.
[437] See Audi (2004), 137.
[438] Audi (2004), 124f.

particularist that *prima facie* duties always count in a certain direction, while at the

same time explaining cases like Schadenfreude or sadism. The question to ask is

whether this is a stable position. On the face of it, it seems paradoxical to say that

pleasure always makes an experience better if its contribution to the overall value

in an act of sadism is negative. In defense to that challenge, Audi's intrinsic value

claim can be understood in two ways. In a stronger interpretation, pleasure always

counts in favor of an action. Some of what Audi says suggests such a reading:

> The most important single point is that whatever one may consider
> intrinsically good or intrinsically bad, one is committed to taking it
> to provide a reason for action, specifically, some positive
> consideration that is normative at least in the wide sense that it
> counts toward the rationality of the action in question. For instance,
> if we believe that pain is intrinsically bad, we are committed to
> taking it to provide (negative) reasons for action, thus to regarding
> the fact that doing something causes pain as a reason to avoid doing
> it.[439]

This strong interpretation of the intrinsic valence claim is incompatible with the

organic valence claim. Examples like sadism or Schadenfreude demonstrate that

pleasure sometimes provides negative reasons; therefore, it is hard to see how Audi

can say that pleasure always counts in favor of an experience. At other places

however, Audi defends a weaker claim:

> To be sure, in order to appreciate the positive intrinsic value of
> pleasure in the bad, one must abstract from the (intentional) content
> of the pleasure. [. . .] [W]e have the possibility of negative
> contextual value combined with positive intrinsic value, much as we
> may have this overall result in the case of pleasure in the bad.[440]

Here, we have a much less ambitious statement. If we isolate a feature with an

---

[439] Audi (2004), 130.
[440] Audi (2004), 129.

invariant valence, the feature always counts in favor or against the moral value of an experience. This does not conflict with the organic value claim since the weaker statement does not make any assertion about the moral valence of a feature in context. The problem with the weak understanding of the intrinsic value view is that it threatens to diminish the utility of the concept of *prima facie* duty for the analysis of moral reasoning. After all, the appeal of intuitionism is that it captures how we morally deliberate in everyday situations.[441] If all the concept of *prima facie* duty means is that taken in isolation, benevolence, gratitude, justice etc. count in favor of an action, we haven't learnt anything about the way we actually deliberate. In real life, these considerations occur in contexts, and what we are interested in is their impact in these situations.

Hence, Audi's solution to the problem of valence-switching faces a dilemma. His strong version of the intrinsic valence claim is incompatible with the organic value claim, while his weak interpretation of the intrinsic valence claim does not offer an account that helps to understand our everyday moral reasoning. What the intuitionist needs is a theory that accomplishes two things. It has first to explain why *prima facie* duties have an invariant valence – even when they occur in contexts; secondly, the theory needs to allow for cases in which valence-switching occurs. In the last part, I am going to argue for such an account. In the meanwhile, I consider a rival explanation that defends the invariance of *prima facie* duties against particularism.

---

[441] See Audi (2004), 1f: "[I]n its richest forms [intuitionism] [. . .] has a normative core that is, at least in its best- known version, developed by W.D. Ross, close to the kinds of generally uncontroversial everyday judgments that any ethical theory seeks to account for." See as well p. ix.

7.4 McKeever and Ridge's argument for codified holism

McKeever and Ridge have developed an argument that can be read as an attempt to address the problem that Audi's account of establishing the univalence of *prima facie* duties faces. Like Audi, they recognize that there are genuine cases of valence-switching, but they go beyond him in defending a more robust notion of univalence which holds even if the *prima facie* duty is not taken in isolation. They do this by codifying the contexts which affect the valence of a *prima facie* duty. This allows them to isolate cases like Schadenfreude or sadism and to maintain that all other instances of e.g. pleasure count always – and not only in isolation, but also in relation to other considerations – in favor of doing an action.[442] McKeever and Ridge's notion of invariance does not imply that it is never the case that e.g. bringing pleasure counts against an action. However, it comes close to this by claiming that the valence of a *prima facie* duty is stable apart from a limited and in advance definable set of circumstances (I will call this form of invariance quasi-invariance). Ross, textual evidence suggests, would be happy with quasi-invariance.[443]

The way McKeever and Ridge's argument works is that they question how particularists interpret holism. In order to imply particularism, holism must contain two theses that are better kept separated: the first concerns the context-sensitivity

---

[442] McKeever and Ridge, who make this claim, didn't specifically intend to defend *prima facie* duties, but ethical principles in general. However, their argument can be used on the intuitionist's behalf. If principles in general can be quasi-invariant, so can *prima facie* duties.

[443] In response to W. A. Pickard-Cambridge (1932) who argues that utilitarianism fits better than intuitionism with common-sense because it can better explain cases in which special circumstances invalidate promises, Ross (1939) replies that for him, the binding-force of promises depends on certain circumstances (e.g. that the promisee has not lied), and this explains as well as the utilitarian these cases. Hence, Ross codifies the circumstances in which the fact that I have promised carries no positive valence.

of reasons, while the second holds that reasons are not codifiable. These claims can come apart since it is possible that a reason depends upon context, while all circumstances in which the context is able to affect this reason can be indicated. The denial of codifiability is therefore an extra ingredient of holism which is nothing more than a denial of quasi-invariance itself.

Hence, to claim that holism consists in the conjunction of the two claims is to beg the question about the particularist's claim.[444]

For their argument to be useful for the intuitionist' project, it has to be interpreted as an epistemological claim: We can rely on the quasi-invariant valence of ethical principles such as *prima facie* duties because it is possible for us finite beings to indicate all conditions which affect the direction in which the a consideration counts. It might be true that a complete list of those valence-changing conditions exists, but we as finite human beings need to have epistemic access to them in order to give *prima facie* duties the status of quasi-invariance.[445] In order to attack McKeever and Ridge's argument, it is therefore enough to show that the weaker epistemic claim is wrong.

McKeever and Ridge's argument can be illustrated with an example. Take the following formulation of the *prima facie* duty of benevolence:

(U) The fact that an action would promote pleasure gives a *prima facie* duty to perform the action if and only if the object of the pleasure is not another's pain.

(U) is a context-sensitive formulation of the *prima facie* duty of benevolence

---

[444] McKeever and Ridge develop this argument in their (2006), 25-32 and 41-43.
[445] See Dancy (2004), 116f.

insofar as the promotion of pleasure is a reason in favor of the action only in certain circumstances. The theory is codified as these circumstances are limited and fully indicated in the formulation of (U). The example is compatible with quasi-invariance. Importantly, holism as understood by McKeever and Ridge cannot be used to criticize (U) because the example is not only compatible with the context-sensitivity of reasons, it even requires it. Hence, holism does not rule out quasi-invariance.

In order to evaluate McKeever and Ridge's argument, it is crucial to see that it implies a certain definition of holism. Their strategy is to drive a wedge between (A) context-sensitivity and (B) non-codifiability and to show that holism consists in (A), but not in (B). (A) alone allows for the quasi-invariance of *prima facie* duties, and does not support particularism. The particularist on the other hand believes that the concept of holism entails both (A) and (B); this entitles her to deny quasi-invariance.

Hence, the question becomes whether non-codifiability is part of the definition of holism or not. Dancy himself proposes a formulation of holism which, unfortunately, does not settle this question:

> (H) "A feature that is a reason in one case may be no reason at all, or
>
> an opposite reason, in another."[446]

This minimal definition of holism talks only about the fact that the same reason can count differently in different situations, but remains silent on whether the conditions that are responsible for the changes are codifiable or not.

---

[446] Dancy (2004), 7 and 73

To make progress, it is helpful to step back and to turn to the phenomenon from which holism arises. This should clarify how the idea is to be formulated most plausibly. Particularists, especially Dancy, claim that not only moral but all reasons behave holistically.[447] It is easy to demonstrate that theoretical reasons are holistic: That something in front of my eyes seems to be red is normally a reason to believe that it is in fact red. If I have however taken a drug that changes my color perception so that all red objects appear blue and vice versa, the fact that something seems red to me is rather a reason to believe that it is blue. The same can be shown for moral reasons. My having borrowed a book from you is normally a reason for me to return it to you; but if I discover that you have stolen it from the library, this reason might vanish.[448] In another example, that an act that is fun is usually a reason for doing it, but the fun experienced in fox hunting seems to be a reason against it.[449] Further examples can be produced at will, but I think the point is clear. The structure of the examples presented is identical:

(1) We have a reason normally counting in one direction.

(2) In certain unusual contexts, the reason changes its behavior.

(3) Therefore, the reason is not invariant.

The variance of *prima facie* duties is only codifiable if all valence-changing contexts in 2. can be specified in advance. But why should we be able to exclude that unforeseen circumstances occur which affect the valence of a reason? This sounds implausible for both theoretical and practical reasons. Take theoretical

---

[447] See Dancy (2004), 74-76.
[448] See Dancy (1993), 60.
[449] Dancy (1993), 61. See as well McNaughton (1988), 193.

reasons first: Until very recently, the fact that a planet was lacking either carbon, hydrogen, nitrogen, oxygen, phosphorus or sulphur was taken to be decisive evidence that it didn't contain organic life. New experiments have however demonstrated that a qualification has to be added to that rule. If phosphorus is replaced by arsenic, microorganisms can grow. A further point that this finding shows is that we shouldn't be confident that in the future, no further elements will be found that are able to substitute other elements from the original list. Or consider aesthetic reasons: It was once thought as always counting in favor of a piece of music if it respected the laws of tonality, but in the 20th century, some musicians have come to the belief that in the right context, atonality can also make good music. The same goes for moral reasons. Other than in the past, many people nowadays believe that the pursuit of profit is only morally legitimate if it does not harm the environment. As a second example, it has over time become common sense that a punishment for a crime is justified only if the criminal does not suffer from a mental condition. Through reflection and new empirical information, people have concluded in these cases that their moral rules have to be qualified.

As these examples demonstrate, we shouldn't put confidence in definite lists of valence-changing contexts. Reasons are context-sensitive, but not codifiable.[450] This insight makes it possible to formulate a more exhaustive definition of holism:

> (H') A feature that is a reason in one case may be no reason at all, or an opposite reason, in another. The circumstances that affect the ethical significance of a feature cannot be codified in advance.

---

[450] As discussed above, McKeever and Ridge's project should be interpreted as the epistemic claim that finite beings have access to the conditions that change the valence of a *prima facie* duty. Therefore, my argument against their claim questions whether we have such an epistemic access. This leaves however open the metaphysical question whether a list of valence-affecting conditions exists or not.

If we accept (H'), this allows us to reject McKeever and Ridge's argument. In contrast to (H'), McKeever and Ridge go for the interpretation that claims that we are able to indicate the conditions under which the changes occur. If we add their position about codifiability to the minimal definition of holism, we get:

> (H'') A feature that is a reason in one case may be no reason at all, or an opposite reason, in another. The circumstances that affect the ethical significance of a feature can be codified in advance.

As seen above, (H'') is wrong. This allows the particularist to accept with McKeever and Ridge that "there is still going to be no argument from context-dependence to uncodifiability,"[451] but to hold against them that holism, if spelled out more thoroughly, contains an additional clause, the denial of codifiability. This clause allows for a rejection of quasi-invariance which is based on (H''): if valence-changing contexts cannot be specified in advance, then McKeever and Ridge cannot claim that a codified morality is compatible with holism.

7.5 Prima facie duties and thick ethical concepts

Both Audi and McKeever and Ridge fail to adequately respond to the particularist' attack against the univalence of *prima facie* duties. However, their arguments bring to light the requirements a satisfactory reply has to fulfill. As the discussion of Audi shows, intuitionists need to explain how a robust invariant valence-claim can be squared with the existence of cases such as sadism or Schadenfreude where a genuine valence-reversal occurs. Can a criterion be found that allows separating those cases from the ones where there is no change in valence? The problem with McKeever and Ridge's argument demonstrates how

---

[451] McKeever in a private correspondence about an earlier draft of this paper.

not to answer this question: We cannot simply create a list of contexts which cause

valence-switching. Given our epistemic limitations, we can never be sure that we

have such a list. An adequate response has therefore to take seriously holism in the

theory of reasons.

A more promising solution to the problem of univalence starts by analyzing

the meaning of *prima facie* duties. As it turns out, they contain an element that sets

them apart from most other considerations: thick ethical concepts.

The notion of thick ethical concepts has been introduced by Bernard

Williams.[452] What sets thick ethical concepts apart from other concepts is that they

can be analyzed into a descriptive and a prescriptive element. Take reparation as an

example. There is certainly a descriptive element in reparation. In order to count as

an act of reparation, its beneficiary must have suffered a harm for which the person

who pays the reparation is responsible. But in addition to that, there is also a

prescriptive element involved since only those harms that result from a wrongful

act require reparation. If the state has punished somebody who was rightfully

convicted, it wouldn't make sense to ask for reparations.

Both the descriptive and the prescriptive element are constitutive for the

understanding of the thick term.[453] Knowing that somebody has been wronged is

taken by itself to be insufficient for identifying acts that amount to reparation

(perhaps no harm has occurred), whereas knowledge of the fact that somebody has

---

[452] See Williams (1985), 140-143. Although Williams' discussion has become the point of reference for debates about thick ethical concepts, similar distinctions can be found in Anscombe (1958) and Foot (1977).

[453] Thick ethical concepts stand in the middle between thin ethical concepts and purely non-moral concepts. In contrast to thick concepts with both evaluative and descriptive content, thin ethical concepts like "morally good" or "morally bad" lack descriptive aspects and consist merely in an evaluative element, while non-moral concepts lack evaluative parts and contain only descriptive elements.

been harmed does not provide enough information either to pick out those cases

for which reparation can be paid (possibly the harm has not been wrongfully

inflicted).

Many of Ross' *prima facie* duties contain thick ethical concepts. The

concept of reparation as it occurs in the *prima facie* duty of reparation, as

discussed above, illustrates this idea. As another example, take the *prima facie*

duty of fidelity. Whether or not I have violated my duty of fidelity by breaking a

promise depends not only on whether there was an (implicit) agreement; it must

also be the case that the promise was not deeply immoral itself. Similarly, I have a

prima facie duty of gratitude only in case the act that was done to my benefit

didn't violate moral standards. For example, if I discover that unbeknownst to me,

the organ transplant that I received was taken from somebody else against his or

her will, I have no duty of gratitude towards those who organized it for me.

Reasons of space do not permit to discuss all *prima facie* duties here, but the idea

should be clear:[454] the elements on Ross' list contain thick ethical concepts.[455] It is

in virtue of these thick concepts that *prima facie* duties have their invariant

valence.[456]

---

[454] Especially, I cannot deal here with complications that arise from the *prima facie* duty of justice – some philosophers argue that it is a thick ethical concept (see McNaughton and Rawling (2000), 268f.), while others believe that it is a thin concept (see Dancy (2004), 8

[455] Another question in this context concerns the variance of those thick ethical concepts that don't appear in the list of *prima facie* duties. Why are, according to the thick intuitionist, the thick ethical concepts that are part of *prima facie* duties invariant, but not those that don't? Here, I don't have the space to take up that issue, but I am inclined to agree with McNaughton who discusses the problem in his (2002). There, he argues that *prima facie* duties are basic and that all other thick ethical concepts can be defined in terms of them.

[456] McNaughton and Rawling (2000) argue as well for the claim that *prima facie* duties contain thick ethical concepts, but my version of "thick intuitionism" differs from theirs: McNaughton and Rawling don't have a satisfactory answer to the argument from holism: Why should we suppose that the valence of *prima facie* duties remains stable in all contexts? They claim that holism demonstrates only that the reasons used in

The prescriptive element that is part of the definition of the thick ethical concept establishes a necessary link between the *prima facie* duty and a thin ethical concept.[457] For this reason, Schadenfreude does not fall under the duty of benevolence, although it brings about pleasure:[458] Only those kinds of pleasure that are worthy of support can fall under this duty. The prescriptive element that is part of benevolence blocks all kinds of counterexamples that particularists might bring up. Even if we are unable to codify valence-changing contexts, the fact that *prima facie* duties contain a thick ethical concept guarantees that they do not apply to these cases.

This response solves Audi's problem of explaining why in contexts like Schadenfreude, pleasure has a negative valence. As Audi himself notices, "this is the wrong kind of pleasure."[459] It was difficult to fit this statement with his claim that all instances of pleasure have a positive intrinsic value. Thick intuitionism however can argue that given the prescriptive part in the meaning of benevolence, cases where pleasure has a negative valence are excluded. This solution avoids as well McKeever and Ridge's problem of presupposing that we are able to codify all

---

the particularist's examples vary in valence; nothing in the argument from holism shows that context-sensitivity extends to other reasons as well (267). In addition, even if some primary reason turns out to depend on context, we can spell out these circumstances: "the list of conditions is not open-ended, and it is knowable in advance" (269). As argued above, these replies underestimate the argument from holism. Our previous discussion has shown that we cannot trust our ability to recognize in advance which contexts change the valence of a reason. Furthermore, to reduce the argument from holism to a list of examples shows that McNaughton and Rawling misunderstand the argument; its purpose is to unveil a general feature of how reasons interact with each other.

[457] A problem this approach raises is that it already presupposes knowledge of ethical terms. As Audi argues, "if any set of principles is to be an independently good guide in moral conduct, we need a way of applying them that does not require an appeal to an independent basis of moral judgment" (Audi (2004), 115). I cannot deal with this problem here, but I believe that it does not amount to a rejection of thick intuitionism.

[458] Another example of inappropriate pleasure particularists have brought up is the pleasure in foxhunting. See McNaughton (1988), 11.

[459] Audi (2004), 72.

valence- switching circumstances. Instead, thick intuitionism blocks counter-examples even without knowing them in advance. Once a situation occurs in which e.g. pleasure has a negative valence, it does not fall under the *prima facie* duty of benevolence.

Yet, this solution seems to come at a high cost. It apparently reduces thick intuitionism to a mere conceptual claim. If and only if pleasure counts in favor of an action, we call its bringing about a *prima facie* duty of benevolence; otherwise, we don't. Or, as Margaret Little puts it: "In the end, the most one can say is that, in contexts in which a feature is good-making, it is good making – not exactly late-breaking news."[460]

Responding to this charge, the intuitionist can argue two things. First, she can claim that a correct conceptual analysis of a central class of ethical concepts is already a noteworthy achievement. If the present analysis clarifies our understanding of how we use concepts such as justice, benevolence or fidelity, we should not criticize it for such a result.

At the same time, it is true that the intuitionist' analysis would be more attractive if it gave us a more substantive reason why *prima facie* duties are univalent. Such a reason would have to indicate what makes these duties important to us; this would not only explain that only those considerations that count in favor of an action fall under a *prima facie* duty, but as well why this was the case.

Such a reason can be found in Audi's own account. He provides a systematization of *prima facie* duties by reference to a version of Kantian ethics.

---

[460] Little (2000), 288f.

Contrary to what Ross might have thought, such a project is compatible with the plurality of morally relevant considerations: *prima facie* duties can be non-inferentially known, have a normative force that is not reducible to an underlying factor, and can even be better grounded as more basic reasons, and still be illuminated and systematized by reference to a fundamental rationale.[461] Kantian ethics offers such an overarching moral theory in the Categorical Imperative.[462] If we accept the universality formulation and the formula of humanity as our basic ethical framework and ask what concrete moral principles we would derive from them for our everyday moral deliberation, the result would be something close to Ross' list of *prima facie* duties.[463] The perfect duties in Ross' list – those of fidelity, reparation and non-malevolence – stem from the universality formulation, whereas the positive duties – those of gratitude, justice, benevolence and self-improvement – can be derived from the formula of humanity.

As Audi argues, the Categorical Imperative in its formulations expresses the idea of respect for the dignity of persons.[464] If this idea reflects what stands behind our ethical reasoning, and if *prima facie* duties translate this abstract thought into moral principles for everyday situations, we have a further explanation of why *prima facie* duties have an invariant valence. Insofar as we care for the dignity of persons, the fact that a certain situation helps to achieve this aim always counts in

---

[461] See Audi (2004), 90 and 104. These claims need more elaboration as I can offer here. Audi argues for them in his (2004), especially 48-54.
[462] Audi stresses that he is not interested in a faithful interpretation of Kant's texts; he rather relies on an ethical framework that is inspired by Kant's ideas. See Audi (2004), 90.
[463] See Audi (2004), 102f.
[464] See Audi (2004), 110 and 112.

favor of doing it. *Prima facie* duties play according to this Kantian reading the role of indicating when this is the case. Therefore, whenever they are instantiated, they have an invariant valence.

Appendix 2: Can ethical intuitionism serve as a framework for biomedical ethics?

8.1 Introduction

If one of the strengths of Rossian intuitionism lies in its close relation to moral phenomenology,[465] this raises the expectation that it be able to explain how we solve concrete moral problems. To make good on this promise, intuitionists have to offer a convincing model of moral deliberation for applied areas of normative enquiry such as biomedical ethics.

The authors that come closest to defending an intuitionist picture of biomedical ethics are Tom L. Beauchamp and James F. Childress. In their textbook *Principles of Biomedical Ethics* (henceforth *Principles*) which is widely adopted in medical and nursing schools and which is recognized for presenting the "most widely used account in biomedical ethics"[466], the authors explicitly endorse a Rossian framework.[467] They accept that ethical decision-making is based on a plurality of non-reducible principles that are valid for all moral agents and derived from common sense;[468] there is no cardinal ordering among the items on the list of principles and no algorithm exists indicating how principles have to be weighed against each other: the principles are defeasible and contain *prima facie* obligations whose strength depends on the circumstances. Importantly, both Ross and Beauchamp and Childress' so-called "Four principles approach" (henceforth *4PA*) combine consequentialist (nonmaleficence and beneficence) with deontological

---

[465] See Ross (1930), 19.
[466] Gert, Culver and Clouser (2006), 99.
[467] Beauchamp (2012), 14-16.
[468] Beauchamp (2012), 4.

principles (respect for autonomy, justice), a feature that sets pluralists in ethics

apart from theories that are built on one kind of principle.

Despite the obvious similarities to Ross's ethical theory, 4PA is in important

aspects different from classical intuitionism. The most apparent difference concerns the

choice of considerations that count in favor or against an action. Ross believes that there

are seven *prima facie* duties: fidelity, reparation, gratitude, non-maleficence, justice,

beneficence and self-improvement.[469] Beauchamp and Childress limit their list to four

ethically relevant considerations: respect for autonomy, nonmaleficence, beneficence and

justice.[470]

For present purposes however, this difference is superficial. Three of Ross' *prima*

facie duties – beneficence, justice and non-maleficence – have equivalents in 4PA.

Considering that strictly speaking, Beauchamp and Childress don't speak of *prima facie*

duties, but define basic normative features in terms of "clusters of principles" – a notion

that is not further defined and rather vague, but intended to be broader than Ross's

concept – the remaining duties on Ross's list can also be encompassed in 4PA: the duty

to fidelity in Ross's framework can be derived from a combination of non-maleficence

and beneficence in 4PA, reparation can be understood in terms of justice and gratitude as

a form of beneficence or justice.[471] Even if the overlap is not perfect (e.g. one could argue

that Ross's *prima facie* duty of fidelity holds even if no considerations of non-

maleficence or beneficence are at stake or that Beauchamp and Childress's notion of

---

[469] See Ross (1930), 21f.

[470] Beauchamp (2012), 12f.

[471] Admittedly, the derivation of gratitude from beneficence or justice is a stretch, but even if this derivation fails, the overlap between Ross and 4PA is still very close.

autonomy has no direct correspondence in Ross or that Ross's duty of self-improvement

– a rather odd one anyway – is left out in 4PA), the difference between Ross and

Childress and Beauchamp is rather one of emphasis, especially given that Ross stresses

that his list of *prima facie* duties is tentative and can be improved upon.[472]

Another difference between 4PA and Ross is harder to reconcile. Beauchamp and

Childress don't claim that the common morality reflects an independent moral reality; if

the conditions of our lives change, the principles have to be adapted.[473] Ultimately, what

matters is that the basic moral principles 'serve society',[474] and an ethical theory which

achieves this aim in current circumstances might differ from the most appropriate set of

moral rules in another setting. That is not Ross's position. As a moral realist, he does not

believe that what is *prima facie* right or wrong can change over time;[475] even if he

considers his list of *prima facie* duties only as tentative, we can in the long run and

through reflection arrive at the true set of *prima facie* duties which is as timeless as

mathematical truth are.[476] Other than Beauchamp and Childress, Ross believes in the

notion of moral progress and in the idea that as a fruit of moral deliberation carried out

through generations, our current moral views are superior and not just different from

those of the past.[477] Although the present chapter deals with methodological issues in

applied ethics and not with metaethics, this difference will play a role in the discussion

about relativism in 4PA below.

---

[472] See Ross (1930), 23.
[473] See *Beauchamp (2012)*, 389-392.
[474] Beauchamp (2012), 389.
[475] See Ross (1930), 15.
[476] See Ross (1930), 29f.
[477] See Ross (1930), 12.

Is Beauchamp and Childress' theory plausible? Can it serve as an illustration that ethical intuitionism in the spirit of Ross is applicable to applied ethics or is Rossian intuitionism in important aspects different from 4PA? If yes, which theory is more plausible? My focus will be on the methodology of biomedical ethics, not on the substantive claims that Beauchamp and Childress make. In the first sections, I will discuss how they apply general norms to concrete situations and consider issues that are related to their solution. Later sections will discuss 4PA's method of justification and their attack on other views. Finally, I will present and evaluate an application of Rossian intuitionism to biomedical ethics.

8.2 4PA's official view on specification

From the very beginning of Principles, the gap between general norms and concrete moral judgments is the focus of Childress and Beauchamp:

> Practical ethics [...] is the attempt to interpret general norms for the purpose of addressing particular problems and contexts. [...] Often no straightforward movement from norms – in the form of theories, principles, or precedents – to particular judgments is available. General norms are usually only starting points for the development of concrete norms of conduct.[478]

The general norms are comprised of a core of universally shared ideas that bind all moral agents; they form the "common morality"[479] which contains the four basic principles. However, "content thin" general norms need to be supplemented by "content rich particular moralities" in order to be applicable to concrete moral problems.[480] Particular moralities express the moral convictions of different cultures, religious traditions or

---

478 Beauchamp (2012), 1f.
479 See Beauchamp (2012), 3f.
480 Beauchamp (2012), 5.

professional communities. They differ from general norms not only in their level of specificity, but also lack the objectivity and universal applicability of common morality.[481] There is no single correct way of spelling out moral rules in the biomedical field.

The roles and relationships that occur within the medical settings demonstrate the need for particular moralities; rules of informed consent or medical confidentiality for example are, although related to the basic principles of autonomy and protection from harm, specific to the professional community.[482]

Although there is more than one way in which specifications can occur, the connection between general norms and particular moralities is not random. Specifications have to be coherent with the general norms, and if it turns out that the specification leads to consequences that are incompatible with common morality, the specification needs to be revised.[483] While this limits the number of possible specifications, it still allows for a variety of ways to spell out what general norms mean in specific contexts.[484] Since specified rules cannot cover all possible circumstances (in order to provide exact directions for all cases, an impractically large – perhaps indefinitely large – number of rules would have to be required), cases of conflict between several specified rules or between specified rules and general principles will occur. When rules are in conflict, the overall obligation has to be established by weighing all relevant considerations against

---

[481] "[W]e accept moral pluralism (some would say relativism) in particular moralities" Beauchamp (2012), 4.
[482] See Beauchamp (2012), 6f.
[483] For a discussion on whether this is – as Beauchamp and Childress believe – compatible with a reflective equilibrium in Rawls' sense, see below.
[484] See Beauchamp (2012), 12f.

each other. Their relative strength is context-dependent. In other words, all morally

relevant considerations have the status of *prima facie* duties.[485]

Can more be said about the relation between general principles and particular

rules? In response to this question, Beauchamp and Childress develop an account of

specification, a mechanism that narrows down the scope of the basic principles by

"spelling out where, when, why, how, but what means, to whom, or by whom the action

is to be done or avoided."[486] It is important that this process goes beyond the mere

analysis of the basic principles; rather, it "adds content."[487]

Beauchamp and Childress discuss two scenarios to illustrate this point. In the first

case, psychiatrists sometimes receive no informed consent from a patient for conducting

a forensic evaluation in a case where the procedure has been ordered by a court.

Nevertheless moving ahead with the procedure can however be justified by introducing a

specification: "Respect the autonomy of persons who are the subjects of forensic

evaluations, where consent is not legally required, by disclosing to the evaluee the nature

and purpose of the evaluation."[488] The second example concerns situations where

physicians decide to lie on insurance forms in order to obtain reimbursement for an

effective treatment for their patient that might normally not be covered under their

insurance plan. Although physicians have no obligations to break the law, in a survey a

majority of them decided to put the interests of patients over their legal duties; this

---

[485] See Beauchamp (2012), 14-16.
[486] Beauchamp (2012), 17.
[487] Beauchamp (2012), 17.
[488] Beauchamp (2012), 17. I will come back to this case below.

suggests that in this case, physicians rely on an implicit specification that details how in the context the duty of beneficence stands to the duty of justice (obeying the law).[489]

Beauchamp and Childress stress that these examples don't stand for isolated cases; rather, moral rules are always subject to specification. Furthermore, the process of specification is never ending as the circumstances that require ethical judgments are constantly changing.[490]

8.3 Descriptive and normative specification

Is this a plausible picture? There are at least three problems with Beauchamp and Childress' understanding of the relationship between general norms and specification which will be discussed in the next sections: (1) There is more than one way of adding content to general norms in order to get to specified rules. Which way do Beauchamp and Childress prefer and for what reason? Is the one they prefer the most plausible one? (2) Beauchamp and Childress are ambiguous in their use of the terms "pluralism"/"relativism." Why are their reasons for supporting relativism? (3) It is unclear what form of specification is required in order to make sense of how general norms are unfolded in different contexts. (4) What status do specified principles have? Do they express *prima facie* duties or duties all things considered? Is 4PA's solution compatible with the role Beauchamp and Childress assign to general norms?

Let's start with the question of whether Beauchamp and Childress provide a satisfactory account of the relation between general norms and particular moralities. As

---

[489] See Beauchamp (2012), 18 and 295f.

[490] However, there are some specifications that are not subject to change, such as rules forbidding unnecessary infliction of pain. See Beauchamp (2012), 19.

explained above, specification in 4PA is more than the mere unfolding of what is already contained in the general norms; rather, it "adds content." Unfortunately, Beauchamp and Childress don't specify what this "extra ingredient" consists in.

At least two answers are possible. First, "specification" can mean the breaking down of abstract general norms into applicable specific rules by taking into account the rich *factual context* that is salient to certain kinds of situations. For example, an investigation of what the principle of autonomy means for the professional duties of physicians needs to reflect e.g. how the expectations of privacy in medical settings have developed, what circumstances would be considered as a breach of autonomy and in which form medical records are being stored and accessed by third parties such as insurance companies. These are contingent facts that change over time and vary from society to society. What specification *does* is to apply these empirical facts to the "thin" general norms; the general norm of respect for autonomy, for example, becomes more specific insofar as it takes into account how far the sphere of privacy extends in certain contexts. What specification *does not* do is to introduce rules that couldn't be directly derived by taking general norms and qualifying them by adding content that can be described in empirical terms. This form of specification can be called "descriptive."

A second way of understanding specifications adds normative content to the general norms (normative specification). The complexities found in new kinds of situations require new rules which – while they don't contradict general norms – go beyond the four basic principles. The example of physicians who think it is okay to lie on insurance forms can serve as an illustration. As described above, when asked whether they would exaggerate the condition of a patient on an insurance form in order to make

the insurance pay for an annual screening mammography, an overwhelming majority of physicians decides to bend the truth in favor of the patient's interest. Beauchamp and Childress suggest that implicitly, the physicians engage in a specification of the general norm "Doctors should put their patients' interests first" into the more specific rule "Doctors should put their patients' interests first by withholding information from or misleading someone who has no right to that information, including an insurance company with unjust policies of coverage, who thereby forfeits his or her right to accurate information."[491] This rule is new and could not have been deduced by simply considering general norms plus the facts relevant in the situation (after all, it justifies conditions under which it is ok to break the rule and claims that certain kinds of insurance coverage are unjust).

Both descriptive and normative specification address the same issue: the list of general norms taken in isolation cannot do justice to the variety of contexts in which ethical problems arise. The two methods diverge however about how to remedy this shortcoming. Descriptive specification suggests that empirical contextual information should be used to clarify how to best apply a rule in a given context, whereas normative specification allows for additional normative context that is added in specific circumstances to the stock of general norms.

The two interpretations of specification are not mutually exclusive. It is possible to believe that descriptive facts inform the formulation of our concrete rules while also holding that concrete duties are not entailed by general norms, but are added to the normative content. However, if one wants to avoid the implausible consequence of

---

[491] Beauchamp (2012), 18.

having to invent new normative content whenever new circumstances arise, those who

defend the normative interpretation of specification are well advised to allow as well for

descriptive specification. After all, this offers an unproblematic way of adapting norms to

ever changing empirical circumstances; normative specification can then be reserved for

cases that cannot be dealt with in this way. Hence, normative specification goes hand in

hand with descriptive specification, but the opposite does not seems to be the case. If one

believes that general norms have to be responsive to empirical facts, there is no pressure

to allow for new normative input as an additional source of specification (unless it can be

shown that there are cases that cannot be resolved with empirical specification alone). It

is certainly possible to subscribe to empirical and normative specification at the same

time, but such a hybrid view is not the only option.

8.4 Which interpretation makes more sense?

While both kinds of specification attempt to address the same underlying

problem, what are the implications for each of them? In regard to normative

specification, it is helpful to have a look at two of Beauchamp and Childress' own

examples, one of which is already familiar. Is normative specification needed in order to

make sense of these cases?[492]

In the example of the physicians asked to fill out insurance forms, the conflict is

between the duty of veracity (which is derived from the basic principles of respect for

---

[492] Even though discussing selected examples provides no conclusive argument in favor of empirical over normative specification, it shows how complex circumstances can in principle be accommodated by means of empirical specification alone.

autonomy among others)[493] and the duty of beneficence towards patients ("Doctors should put their patients' interests first"). In solving the problem, Beauchamp and Childress infer that physicians engage in a process of specification that results in a new principle. This principle specifies the obligation doctors have toward their patient's interests.[494]

How is this process best understood? Beauchamp and Childress interpret it as implicitly relying on a new rule which, although it is based on a basic principle, goes beyond it by introducing a further normative element that specifies conditions under which the basic principle trumps other basic norms (in this case, the principle of justice, according to which the physician owes an honest report to the insurance company). However, there is an alternative and simpler explanation: taking into account relevant empirical information about the benefit the procedure provides for the patient, the costs involved and the patient's economic situation, the physicians weigh their duty of truthfulness against their obligation towards their patients and conclude that the second duty outweighs the first. All that is needed is an awareness of the general duties involved, a mature sense of moral judgment plus a sufficient degree of knowledge about the relevant circumstances. This simpler explanation of how are physicians are deciding these cases is also superior as it allows for flexibility. If, for example, the health care system was at the brink of financial ruin, the costs for the procedure exceedingly high and the benefit marginal, the physician might come to another conclusion and put the insurance interests above the patient's. For Beauchamp and Childress to allow for

---

[493] For a discussion of how obligations of veracity are related to basic principles, see Beauchamp (2012), 289.
[494] Beauchamp (2012), 18.

conditions which justify a different behavior, they would have to introduce further normative specifications establishing under what conditions insurance policies are to be considered just or unjust.

The second example concerns the right of privacy applied to problems of compulsory screening for HIV and tuberculosis.[495] The general norm of autonomy implies a "right to privacy" according to which an agent controls access to his or her information, personal spaces, choices or interests.[496] Consequently, a patient should not be forced to undergo screening for certain diseases. On the other hand, duties of beneficence sometimes require that in order to protect others, doctors and to those who are in contact with a patient need to know whether he or she is infected by a dangerous and contagious disease. Given this tension, how should screening for HIV-infections be handled? Beauchamp and Childress argue that given recent medical advances in the treatment of HIV/AIDS, compulsory screening is only justifiable in narrow circumstances such as in prisons where the HIV-rate is high and where transmissions occur frequently. However, they take a stronger position in regard to tuberculosis, a disease which spreads by airborne transmission and has a high fatality rate. Here, they defend compulsory testing in a wider variety of cases and in some circumstances even "directly observed treatment," that is, monitoring whether patients are taking their medication.[497] As in the previous example, conflicting principles are weighed by taking into account the empirical circumstances; no new rule that couldn't be derived from the basic principles is being introduced.

---

[495] See Beauchamp (2012), 298-302.
[496] See Beauchamp (2012), 296f.
[497] See Beauchamp (2012), 300f.

What these examples illustrate – even if they don't amount to a proof – is that empirical specification can be sufficient to explain how general norms are applied in complex circumstances. As long as defenders of normative specification don't challenge their opponents with an example that can only be met by means of normative specification, the burden of proof lies on those who want to introduce more contentious methods that go beyond descriptive specification which – as explained above – all parties rely on. No such argument can be found in Beauchamp and Childress. What is their motivation for relying on normative specification despite the availability of the less problematic method of empirical specification?

8.5 What motivates normative specification?

As mentioned above, two key assumptions that Beauchamp and Childress make is that morality consists (1) of a common core (expressed in general norms) that binds every moral agent and (2) of an application of this core. At the level of application, there is a plurality of particular moralities that are relative to communities. It is the function of specification to bridge this gap. Does descriptive specification suffice to provide this function, or is it necessary to add normative specification as well? If normative specification were required to account for a plurality of particular moralities, this would explain what motivated Beauchamp and Childress to include it in 4PA. But do they also provide independent and convincing reasons for assuming pluralism/relativism on the applied level? If not, there would be no need for normative specification.

Before addressing these questions, the meaning of "particular moralities," "pluralism" or "relativism" as used by Childress and Beauchamp needs to be clarified. Two interpretations of these terms are possible, and both receive textual support in

Principles. First, particular moralities can be relative to different contexts. Although

separate contexts have different moral rules, these different sets of rules don't conflict

because the areas to which they apply normally do not overlap. For example, while

physicians are subject to rules of informed consent when dealing with patients, these

rules have no equivalent among the professional duties of firefighters since their activities

usually don't give rise to situations in which informed consent is an issue. However,

pluralism can also be understood in a stronger sense. Under this interpretation (which

might as well be called "relativism"), a plurality of *possibly conflicting* sets of moral

rules can be justified.[498] As an example, a Catholic hospital might claim that it violates

fundamental moral rights of conscience if institutions or individual health care

professionals are being forced to participate in abortions, but the government or

differently-minded interest groups see no such moral constraint and demand that the

hospital offer the procedure. According to Beauchamp and Childress, particular

moralities must not be tied to entire communities or groups; even individuals can have a

personal morality that they "have very good moral reasons for."[499]

In other words, while for the weak form of pluralism, the reason for the plurality

lies in the fact that there are different *contexts*, strong pluralism is grounded in *different

views of subjects*. Although Beauchamp and Childress do not differentiate between weak

---

[498] Note that Beauchamp and Childress' position is not merely that different moral viewpoints exist; they also hold the normative view that a plurality of moral codes can be justified; this is implied in Beauchamp (2012), 4f.
[499] Beauchamp (2012), 6.

and strong forms of pluralism, the overall thrust of their argument makes it clear that they support both forms.[500]

Now that it is clear what Beauchamp and Childress mean by "pluralism," we are in a position to ask what form of specification is necessary in order to derive a (strong) plurality of particular moralities from a single set of general norms that are universally valid.

Descriptive specification leads to a plurality of particular moralities, but the pluralism is weak. The resulting principles might differ from context to context, but for each separate context, there is only one correct way of specifying norms. It is true that disagreement can arise over the assessment of the facts, but this is no objection to the claim that such disagreement can under ideal circumstances be solved. For example, in determining how the protection of patient autonomy has to be codified in certain medical settings, we need to figure out first how the notions of privacy and autonomy are interpreted by patients, how records are stored and used etc. Reasonable agents should be able to agree on these descriptive questions, as difficult as they might be; but once this is achieved, we have the material at hand to determine the significance of the medical context for the application of the general norm of autonomy – we can now say how

---

[500] Support for the weaker form of pluralism can be found at Beauchamp (2012), 6, where Beauchamp and Childress justify plurality be reference to the needs of different professions: "In medicine, professional morality specifies general moral norms for the institutions and practices of medicine. Special roles and relationships in medicine require rules that other professions may not need." Other passages are ambiguous. Beauchamp (2012), 16: "The reason why directives in particular moralities often differ is that abstract starting points in the common morality can be coherently specified in more than one way to create practical guidelines and procedures;" both weak and strong forms of pluralism are compatible with this statement. However, overall there is a clear preponderance of cases that support strong pluralism. To quote one, in Beauchamp (2012), 5 Beauchamp and Childress state that "[m]any moral norms are not shared by all cultures, groups, and individuals. [...] These specific moralities include the many responsibilities, aspirations, ideals, sympathies, attitudes, and sensitivities found in diverse cultural traditions, religious traditions, professional practice standards, institutional expectations, and the like." Further passages supporting strong pluralism can be found at Beauchamp (2012), 6, 18/19 and 384.

important autonomy is for doctor-patient relationships and what factors trigger autonomy-related duties. Since other norms might play a role as well, this does not always settle what our duty all things considered is in a given situation; how to weigh different duties against each other is still a matter of judgment. What is important here is that descriptive specification provides no support for relativism or strong pluralism. Quite to the contrary if one assumes descriptive specification, then the view that there is one and only one correct way of specifying what context means for the application of general norms to specific areas of normative inquiry seems to follow naturally.

Things are different with normative specification: if it is possible for any individual or group to add new rules to the universally valid stock of principles whenever they think the situation calls for it, it is easy to see how a plurality of alternative and possibly conflicting moral views (i.e. strong pluralism/relativism) emerges. The only limitation on what normative content can be added is that the new set of rules must be coherent with the general norms and with other rules the group/individual accepts (internal consistency). As long as this requirement is fulfilled, the view has no resources to distinguish between better and worse forms of specification.[501] Consequently, the theory can offer no resolution if different particular moralities come into conflict over some case.

In order to argue for the stronger version of pluralism/relativism, Beauchamp and Childress hence need to assume normative specification. But can the adaptation of normative instead of descriptive specification be justified as necessary for solving the

---

[501] A possible reply to this argument comes however from Dworkin's distinction between "best" and "justification" in his (1977), Ch. 4.

original problem, the adaptation of general rules to the specific context of biomedical ethics? If, as I have argued above, empirical specification alone is sufficient in order to account for the ways in which circumstances affect abstract general norms, the adaptation of normative specification turns out to be an unnecessary and costly step for Beauchamp and Childress.

This reveals that Beauchamp and Childress are engaged in two different projects at the same time and that the solutions for them fall apart. The first project seeks to bridge the gap between *abstract* general norms and *diverse and complex situations*; descriptive and normative specification are the two options, but the descriptive form of specification does a better job. The other project introduces the additional requirement that the move from general norms to specific rules also needs to be a move from a *single* set of moral norms to a *plurality* of sets of principles that can co-exist and possibly even compete against each other. To fulfill this second project, descriptive specification is not sufficient and the normative form of specification is required. Preferring the more problematic normative over empirical specification can however only be made plausible (a) if there are independent reasons to embrace relativism or (b) if there are cases that I have ignored that require normative specification. Beauchamp and Childress demonstrate neither of these.

8.6 The status of specified rules

What status should concrete, specified rules should have? Several things can be meant by this question, and it pays to be careful. The distinction at stake could be between (1) rules that cast *all-things-considered judgments* versus rules which state that there is merely a *prima facie duty* or between (2) rules that, when instantiated, *always*

indicate the presence of a (prima facie) duty versus rules which state that if certain conditions are realized and *if no defeaters are present, we have a certain (prima facie) duty*. Here is an example of the second distinction: We have reason to support laws which increase the availability of donor organs, but if the law achieves this by allowing prison inmates to be killed for that purpose, this isn't merely a reason that has to be weighed against the benefits, but a defeater which cancels out the duty of beneficence in favor of the legislation.[502]

In regards to the first distinction, Beauchamp and Childress are clear that they interpret rules as expressing *prima facie* duties: "All general norms are justifiably overridden in some circumstances."[503] It is not clear whether they believe that specified rules are defeasible in the second sense because they don't address these cases.[504] However, this should in principle not be a problem, as nothing in their theory excludes the existence of defeaters. What raises more issues is Beauchamp and Childress' interpretation of rules as *prima facie* duties. This implies that even if specified rules are available, general norms can still play a central role in our practical deliberations. In other words, when trying to figure out what our overall duty is, we have both general *prima facie* rules and specified principles that are not conceived as *prima facie*. It is not clear from their account how to adjudicate between those two kinds of rules. It is no solution to continue specifying rules as long as needed in order to account for all exceptions: "A scheme of comprehensive specification would constitute a package of potentially

---

[502] Assuming that it is not an option to vote for a version of the law that removes the problematic clause.
[503] Beauchamp (2012), 15.
[504] Some of what Beauchamp and Childress say could be interpreted as endorsing such an option: "In light of the range of possibilities for contingent conflicts among rules, even the firmest rules are likely to encounter exceptive cases" (Beauchamp (2012), 19).

hundreds, thousands, or millions of rules, each suited to a narrow range of conduct."[505]

And since there is often no alternative specified rule that captures in a correct way what

our duties are, the additional considerations must sometimes be expressed as a *prima*

facie duty.[506] If *prima facie* duties can directly be weighed against specified rules, this

gives them a stronger role than Beauchamp and Childress are willing to grant. For them,

general norms are only "starting points for the development of concrete norms of

conduct"[507] or "general guidelines"[508] that "do not function as precise guides to action"[509]

and that by themselves are not designed to solve practical problems: "Our framework is

spare, because *prima facie* principles do not contain sufficient content to address the

nuances of moral problems. However the principles can be specified to provide more

specific guidance."[510] According to 4PA, their function is rather to set the boundaries that

specified rules should not transgress (i.e. specified rules must not violate the general

principles).[511]

      This raises a dilemma for Beauchamp and Childress: either, they allow general

rules to play a central role in practical deliberation as is required by their understanding

of specified rules as *prima facie* duties, but this weakens their argument that without

specification principles and particular moralities, moral problems cannot be solved. Or

they hold on to the belief that general norms are too abstract to be of use in concrete

---

[505] Beauchamp (2012), 22.
[506] At other times, the additional considerations can be expressed in non-normative terms.
[507] Beauchamp (2012), 1f.
[508] Beauchamp (2012), 12.
[509] Beauchamp (2012), 13.
[510] Beauchamp (2012), 16f.
[511] See Beauchamp (2012), 12.

moral deliberations, but are consequently unable to explain how normative considerations that are not captured by a specified rule can affect what our duty all things considered is.

8.7 Justification

By what method can ethical theory be justified? In this section, I will first discuss Beauchamp and Childress' approach to justification and then turn to their discussion of rival theories, especially what they call the "deductive method." I will argue that Rossian intuitionism falls under that label and consider whether Ross can respond to the criticisms.

Beauchamp and Childress base their own method of justification on Rawls' theory of a "reflective equilibrium."[512] According to Rawls,[513] justification takes the form of a reflective process where considered judgments (i.e. judgments formed under ideal circumstances that allow for thorough reflection and rule out bias) about concrete cases are tested against our considered judgments at more abstract levels of moral thinking. The aim is to find a coherent set of beliefs. If, as is to be expected, we encounter a conflict between our beliefs, we have to adjust some of our convictions. These changes can occur at any level, affecting either our judgments about concrete cases or their underlying abstract principles. No belief is immune to modification. The revised set of beliefs is then again subject to the test of coherence. Assuming that we hold countless beliefs that can potentially conflict with each other and given that we constantly encounter new situations to which our moral views are being applied (which might yield to results that don't match

---

[512] See Beauchamp (2012), 381-388.
[513] See Rawls (1971), 42-45.

other well considered judgments), there seems little hope of reaching an endpoint in the process of rendering the whole of our beliefs coherent.

There are two forms of reflective equilibrium, narrow and wide. In contrast to narrow reflective equilibrium, the wide form includes not only all of our moral beliefs, but also empirical background theories, extending the test of coherence to the complete set of our beliefs. Beauchamp and Childress adopt a wide form of reflective equilibrium.[514]

In one important respect, 4PA departs from Rawls: Beauchamp and Childress believe that "bare coherence never provides a sufficient basis for justification."[515] First, a moral code could be unacceptable but still coherent with all other beliefs (Beauchamp and Childress quote a "Pirate's Creed of Ethics").[516] Second, justifying every moral judgment by reference to another moral judgment leads to a "vicious circle."[517] Therefore, Beauchamp and Childress add a foundationalist element to Rawls' coherentist methodology: common morality is justified without dependence on other moral beliefs. Rather, general norms stem from "considered judgments that are the most well-established moral beliefs."[518] This limits the scope of the reflective equilibrium to the specifications and exempts general norms from revision.[519]

---

[514] See Beauchamp (2012), 383.

[515] Beauchamp (2012), 384.

[516] Does this example show that coherence cannot rule out a moral code that allows for theft, slavery etc. (see Beauchamp (2012), 407)? One could argue that it is incoherent to respect your fellow pirates, but to enslave other fellow human beings without being able to point to an ethically relevant difference, as seems to be the case in Beauchamp and Childress' example. Furthermore, even Rawls' reflective equilibrium requires more than mere coherence, but also reliance on well-considered moral judgments. Rawls explicitly believes that those judgments rule out slavery (Rawls (1971), 17f.).

[517] See Beauchamp (2012), 408.

[518] Beauchamp (2012), 385.

[519] Is common morality really immune from revision? Beauchamp and Childress are not always consistent. At places, they seem to defend such a view: "The thesis is that reflective equilibrium needs the common morality to supply initial norms, and then appropriate development of the common morality requires

A number of problems cast doubt on this marriage between coherentism and foundationalism. Why should we put more trust in our considered judgments about general norms as compared to our considered judgments about concrete cases? After all, we often have the firmest convictions when it comes to concrete cases that display clear wrongdoings or praiseworthy behavior. To give general norms a privileged status seems *at hoc* and reveals a bias on Beauchamp and Childress' part in favor of generality. Here is a possible explanation: general norms are shared by all moral agents, whereas particular moralities – if one adopts normative specification – can differ from group to group or even from agent to agent. If Beauchamp and Childress want to ensure that the same general norms apply to all moral agents, they have to find a way to make those norms immune against changes that result from the process of reflective equilibrium. Since agents belonging to different communities adhere to different particular moralities, they face different challenges to render their beliefs system coherent; hence, if general norms are not immune from change, some particular moral communities might end up solving inconsistencies between the common morality and concrete judgments by changing the general norms. A foundationalist stance towards general norms might avoid those

specification, balancing, and reflective equilibrium, a method of coherence. A warranted approach using reflective equilibrium does not involve the relentless reduction to coherence of any set of preferred beliefs. We start in ethics with a particular set of beliefs – namely, the set of considered judgments that are acceptable initially without argumentative support. We cannot justify every moral judgment in terms of another moral judgment without generating an infinite regress […]. The way to escape this regress is to accept some judgments as justified without dependence on other judgments." (Beauchamp (2012), 407f); "[W]e do not speak of our account as coherentism or as a coherence theory, but present our work as reaching for coherence after starting with considered judgments as basic building blocks" (Beauchamp (2012), 408). However, other passages are less clear and would allow that basic judgments are open to revision: "A moral belief that is used initially and without argumentative support can only serve as an anchor of moral reflection if it survives subsequent testing for coherence" (Beauchamp (2012), 408) Overall, I interpret Beauchamp and Childress to hold the first view; they repeatedly insist that in regard to the justification of basic beliefs, their approach differs from Rawls, and a Rawls starts with considered judgments that are then made coherent with other beliefs (add reference in Rawls). Hence, if Beauchamp and Childress want to hold a view that gives considered judgments about general norms more weight than Rawls, the way to go is to hold that they are immune from revision.

problems. Even if this is a correct explanation of their *motivation*, it does not yet *justify*

the privileged position 4PA gives to general norms. In Rawls own elaboration of the

method of reflective equilibrium, considered judgments about rather *concrete* beliefs (e.g.

that we should be tolerant towards other religions or that racism is bad) play the role of

"provisional fixed points which we presume any conception of justice must fit."[520]

Second, what exactly is the difference between the foundationalist justification for

general norms and the coherentist justification for all other ethical judgments? General

norms, as Beauchamp and Childress stress, are completely justified (i.e. without the

further need for reflective equilibrium with other beliefs) because they are the result of

considered judgments. But what counts as a considered judgment? Beauchamp and

Childress provide a list of epistemic qualities that qualify an agent to make considered

judgments:

> Moral judges are entitled to claims to have reached considered
> judgments only if those judgments have been framed from an impartial
> perspective that reins in conflicts of interest and other temptations of
> self-interest; the evaluator exhibits attitudes of sympathy and
> compassion for the welfare of others; the evaluator possesses pertinent
> information about the relevant matters; the evaluator is able to display
> these attitudes in a consistent and sustained way; and the like.[521]

The problem is that these four criteria are at least in part already contained in the

definition of the reflective equilibrium, especially in regards to the notion of coherence. If

that is the case, the method is a mere variation on Rawls' method, not a contrast to a

coherentist theory. Let's consider those four criteria one by one. The "impartial

perspective that reins in conflicts of interest and other temptations of self-interest"

describes an attitude where no undue preference is given. In other words, benefits and

---

[520] Rawls (1971), 17f.
[521] Beauchamp (2012), 409.

burdens are coherently applied. Its opposite, an ethical belief system that allows for

conflicts of interests, would clearly not be in a state of reflective equilibrium. The second

criterion, "the evaluator exhibits attitudes of sympathy and compassion for the welfare of

others," is not as easily explained in terms of coherence alone. However, if somebody

exhibited sympathy for some beings but ignored others in similar circumstances, that

person would display a deficiency in his or her abilities to express sympathy. Therefore,

coherency in applying sympathy is required for a fully developed sense of sympathy. The

next condition, "the evaluator possesses pertinent information about the relevant matters"

comes close to a wide reflective equilibrium which requires – by Beauchamp and

Childress' own account – an "assessment of the strengths and weaknesses of the full body

of all relevant and impartially formulated judgments, principles, theories, and facts."[522]

Finally, the requirement that "the evaluator is able to display these attitudes in a

consistent and sustained way" relies openly on the notion of coherence. In sum,

Beauchamp and Childress' attempt to find a foundationalist ground that solves the

problem of circularity they see in Rawls' coherentist method fails to deliver.[523] If the

method of reflective equilibrium cannot justify our moral beliefs (as Beauchamp and

Childress believe), it is difficult to see how their own account can fix the problem.

Next, there is the question of whether the notion of coherence can be applied in

the context of particular moralities. As discussed above, general norms are valid for all

moral agents, but particular moralities specify general norms in different ways for

different communities. Hence, there is no tight relation that allows the deduction of

specified principles from general norms. At the opposite, specification is a process of

---

[522] Beauchamp (2012), 406.
[523] See Beauchamp (2012), 408.

"inventive content expansion."[524] From the way general norms are applied for physicians, no prediction is possible how norms will be worked out for policemen, politicians or teachers. But if that is the case, what are particular moralities supposed to be coherent with? If "many particular moralities are reasonably coherent ways to specify the common morality,"[525] then the fact that different particular moralities diverge from each other does not in itself count as a "discrepancy"[526] that needs to be fixed. What about the relation between general norms and specified principles? As discussed above, general norms and specified principles are of different kinds: in contrast to specified principles, the "indeterminate" general norms don't have "sufficient content" to be action-guiding.[527] How can it then be established whether a specified principle is in tension with a general norm? If no direct conflict between the two is even possible, it is hard to see what a reflective equilibrium between them amounts to. In the face of those limitations, what work is reflective equilibrium supposed to do? What is left is coherence within a local, specified framework and a vague notion of coherence between general and particular morality. This, however, is far less than Beauchamp and Childress intended with their combination of wide reflective equilibrium with foundationalism.

Before moving on to Beauchamp and Childress' criticism of competing methods, it is worth briefly noting that the issue of how different particular moralities are related to one another is problematic not only in regard to justification, but also on the normative level. If 4PA were applied across various areas of applied ethics, this would result in a rag rug of particular moralities which differed as much as they resembled each other; as

---

[524] Beauchamp (2012), 159.
[525] Beauchamp (2012), 384.
[526] Rawls (1971), 18.
[527] Beauchamp (2012), 17.

mentioned above, physicians would adhere to different rules than policemen, politicians or teachers. The differences would go beyond the obvious fact that these communities encounter circumstances that are empirically different from each other and therefore need to specify general norms for their respective environments; normative specification allows for different ways of interpreting and developing the normative content that is present in the general norms and shared by all moral agents. How should we deal with situations in which there is an overlap between various professions and communities as it frequently happens between e.g. investigative police work and the psychiatric treatment of suspects (in one of Beauchamp and Childress' own examples, the question arises whether psychiatrists are morally allowed to evaluate suspects and share the results with the police)? Which moral code should take precedence? Is there any way to integrate the various perspectives that bear on a situation? Given 4PA's commitment to moral pluralism, it would be difficult to come up with a ranking of competing ethical codes.

Besides presenting their own method, Beauchamp and Childress also attack alternative ways of justifying biomedical frameworks. Of interest here are their arguments against the "deductivist model."[528] The basic idea of deductivism is simple: "[J]ustification occurs if and only if general principles and rules, together with the relevant facts of a situation, support an inference to the correct or justified judgment(s)."[529] The application of norms to situation works as follows:

    1. Every act of description *A* is obligatory.
    2. Act *b* is of description *A*.
    Therefore,

---

[528] Beauchamp (2012), 369-371.
[529] Beauchamp (2012), 391.

3. Act *b* is obligatory.[530]

Ethical judgments are justified because they can be subsumed under a moral rule. Those rules in turn may themselves be justified in virtue of being covered by higher-level principles (but they don't have to).

Beauchamp and Childress raise several criticisms against the deductivist model. Although it works for simple and clear cases that can be solved by reference to a single rule, they claim that the deductivist model does not have the resources to deal with more complicated situations. First, there might be several rules that bear on a case and that have to be balanced against each other, but the deductivist model has no room for balancing. Second, "abstract rules and principles in moral theories are extensively indeterminate;" without specification, we are not able to take into account "factual beliefs about the world, cultural expectations, judgments of likely outcome, and precedents"[531]; the deductivist model does not allow for that. Third, different people might believe that different rules bear on the issue.[532] Last, deductivist top-down models 'create a potentially infinite regress of justification'.[533] Since moral principles cannot be self-justifying, they need to be justified by reference to a higher-level principle, and the problem repeats itself *ad infinitum*. Hence, the deductivist model lacks proper epistemic justification.

---

[530] Beauchamp (2012), 391.
[531] Beauchamp (2012), 392.
[532] "Even if we have all of the facts straight, our selection of pertinent facts and pertinent rules will generate a judgment that may be incompatible with another person's selection of facts and rules" Beauchamp (2012), 371.
[533] Beauchamp (2012), 393.

Beauchamp and Childress' criticisms are difficult to assess in a void. In order to evaluate

their arguments, we have to test them against a concrete deductivist theory. Ross' ethics

will serve as an example.

8.8 Rossian bioethics

While Ross' work predates the division of ethics into normative and applied

branches, he would certainly have thought that an agent facing a biomedical issue would

find recourse in intuitionism. His basic idea is that the various ways in which we relate to

others give raise to a number of different duties;[534] as in 4PA, these duties are irreducible

to a unifying monistic theory and include consequentialist and deontologist

considerations. Some duties, such as the duty of beneficence, are general and are owed to

all people, but others, such as the duty of reparation or gratitude, are special and apply in

certain circumstances only.[535]

What makes Ross a deductivist in Beauchamp and Childress' sense is that

together with the relevant facts about a situation awareness of those duties suffices to

come to a justified conclusion about what to do. What sets Rossian intuitionism apart

from other deductivist theories is the notion of *prima facie* duties, i.e. the idea that the

strength of the duties that bear on a situation depends on context.[536] We need to exercise

judgment in order to weigh conflicting *prima facie* duties against each other in order to

come to a duty all things considered ("duty sans phrase"). Consequently, our duty all

things considered isn't the result of a situation "falling under some general

---

[534] See Ross, (1930), 19.
[535] See Ross, (1930), 26f.
[536] See Ross (1930), 19f.

description."[537] Ross explicitly rejects theories that assign absolute strength to a single rule as being in contradiction "with what we really think."[538]

As mentioned above, he proposes a list of seven *prima facie* duties, for example the duty of fidelity that rests "on a promise or what may fairly be called an implicit promise, such as the implicit undertaking not to tell lies" or the duty of non-maleficence demanding 'not to injure others' and which Ross takes pain to distinguish from the general duty of beneficence.[539] Depending on the circumstances, a duty all things considered does not need to expressible in the language of any single *prima facie* duty, but can be "compounded together in highly complex ways."[540] For example, the duty to obey the law (to use Ross' example) results from the *prima facie* duty of gratitude towards one's country combined with the *prima facie* duty of fidelity (based on an implicit promise).

What epistemic justification does Ross offer for the list of *prima facie* duties? He believes that convictions regarding our most basic duties have the status of self-evident truths.[541] Although Ross offers no extended defense for his foundationalist approach to moral epistemology, recent scholars such as Robert Audi and others have developed sophisticated accounts of self-evident beliefs.[542] Even though there is no space here to enter into those debates, it should be clear that Beauchamp and Childress' unargued claim

---

[537] Ross, (1930), 33.
[538] Ross, (1930), 18f.
[539] Ross, (1930), 21f.
[540] Ross, (1930), 27.
[541] See, for example, Ross (1930), 20fn1.
[542] See Audi (2006), Ch. 2.

that the project of grounding ethical theory in self-evident beliefs is philosophically

unfeasible[543] is premature and ignores the state of the discussion.

It might be objected that up to this point, Ross has delivered the ingredients for a

general normative theory but that applied ethics, especially biomedical ethics, has

different requirements which intuitionism has not met yet. Specifically, an ethical

framework must be able to issue professional guidelines for reoccurring types of

situations. Although the context-sensitivity that comes with the introduction of *prima*

facie duties is important and in line with common sense, too much of that flexibility can

stand in the way of formulating general rules such as those that Beauchamp and Childress

have developed throughout *Principles.* If there is no way to predict which considerations

are in certain types of situations weightier than other ones, generalizations become

impossible. In response, defenders of Rossian intuitionism can allow for generalizations

understood as heuristics (see the discussion in section 6): As a *general rule*, killing

human beings is wrong, but not in certain cases where the patient is suffering from an

incurable and debilitating disease (and agrees to the procedure etc.). Most would agree

that *overall* lying is morally blameworthy, but not when much is at stake and little

sacrificed or when we are forced to answer against our will (again, it all depends on the

details). The purpose these rules serve is to remind us how certain situations are normally

evaluated and to guide our attention to what features are generally relevant. To serve

those functions it is not necessary that these rules be viewed as immune to defeaters or

the weight of other *prima facie* duties that are also relevant. As long as we are clear that

---

[543] See Beauchamp (2012), 393.

this is the status we assign to specifications, Beauchamp and Childress' substantive work in *Principles* can remain unaffected by the kind of methodology that is being adopted.

Having outlined how a Rossian bioethics could be developed, it becomes clear that the view can respond to the criticisms that Beauchamp and Childress have leveled against the deductivist model of justification. The ability to balance the various duties that bear on a situation against each other is a feature of Ross's theory, not a challenge against him. The charge that the abstract rules which are then applied to concrete cases are too indeterminate without further specification does not seem very problematic against Ross' list of *prima facie* duties. His duties (see the examples above) are specific enough to apply directly to concrete cases. It is true that different people might come to different conclusions about what duties are most important [?] or even about what *prima facie* duties bear on a situation. As Ross points out, however, intuitionism is in the same boat as other theories in this respect. Beauchamp and Childress themselves don't seem to have a unique solution to the problem of how to solve ethical disagreement.[544] In response to their criticism of deductivist theories of justification, Ross would disagree with Beauchamp and Childress that our fundamental ethical beliefs cannot be self-evident. As stressed above, different methods of justification in moral epistemology are being defended in contemporary metaethical debates and while it's not evident that Ross' foundationalism will have the last word, it is also clear that his method cannot be as easily dismissed as Beauchamp and Childress suggest.

In conclusion, 4PA faces a number of challenges in regard to its epistemic justification: Beauchamp and Childress give no adequate reason for their preference of

---

[544] See Ross (1930), 23f.

general norms over considered judgments in concrete cases. They fail in their goal to clearly differentiate their own method from Rawls' reflective equilibrium, a notion they deem deficient as a complete method of justification. It is questionable whether the idea of a wide reflective equilibrium is even applicable to a plurality of specified moralities.

Rossian intuitionism does not face those problems, and it avoids the criticisms that Beauchamp and Childress raise against deductivist methods of justification. This makes Ross' theory of justification more plausible than 4PA's. Now that the outline for a Rossian framework in biomedical ethics is on the table, we can extend the question and compare intuitionism and 4PA in other regards as well. In fairness to Beauchamp and Childress, without a fully developed Rossian theory in biomedical ethics, it is difficult to be confident that Ross' theory wouldn't run into other problems that make it overall weaker than its competitors, but if it can be established that Ross at least avoids Beauchamp and Childress' weaknesses and has *prima facie* plausibility,[545] this would amount to a presumptive argument in favor of Ross and against 4PA.

8.9 Is 4PA an intuitionist' theory in Ross' sense?

In the last section, I have already presumed that 4PA is not a version of Rossian intuitionism. While this is a fair assumption in the context of discussions about epistemic justification where 4PA and Ross explicitly differ, it is less clear in respect to normative theory. The very reason to investigate the plausibility of Beauchamp and Childress' theory was to see whether it can be understood as an application of Rossian intuitionism to biomedical ethics. Since we now have an idea of how Ross could extend his theory

---

[545] "*prima facie*" is here used not in Ross', but in the legal sense, meaning "at first sight"

into applied fields of ethics, we can consider how closely 4PA is related to intuitionism. The result is that there are significant differences. If it is important for the overall plausibility of Ross' theory that its relevance to areas such as biomedical ethics can be shown, supporters of intuitionism are well advised to pursue such a project without relying on 4PA.

As has emerged in our previous discussion, there are three major areas of disagreement between the two views. Most fundamentally, Beauchamp and Childress believe that biomedical ethics is in the business of spelling out particular moralities which are relativized to certain (professional) communities; they also believe that there are several valid ways of developing moral codes for the same group of people. Ross, who defends moral realism, rejects such relativism in favor of one set of basic moral rules that remain the same in all circumstances.

The second and related difference concerns the way that specification, the application of rules to concrete cases, proceeds. According to 4PA, there a small number of general norms that lie at the core of morality, but without the addition of empirical and normative content, they are too indeterminate in order to play a direct role in moral deliberation. Ross, on the other hand, believes that the basic ethically relevant considerations directly apply to concrete cases, even though the details of the situation matter in determining the strengths of each relevant duty. For him, *prima facie* duties are more than abstract general norms that play only a remote role in everyday moral deliberations. Third, Beauchamp and Childress disagree with Ross, as discussed in the last section, about the right method of moral justification. Whereas Ross builds his theory on self-evident beliefs, 4PA prefers a combination of considered judgments for basic

norms with the method of reflective equilibrium for specified principles. Whether or not

the commonalities between Rossian intuitionism and Beauchamp and Childress' theory –

both rely on a plurality of first principles, some of which are deontological and some of

which are consequentialist, and both believe that principles need to be weighed against

each other – warrant the label "intuitionism" for 4PA, it is clear that they are in many

respects direct competitors.

8.10 4PA versus Rossian intuitionism

Which competitor is more convincing? This depends at least partly on whether

Rossian intuitionism is able to avoid the traps that Beauchamp and Childress fall into. As

the previous discussion of 4PA and the sketch of Rossian theory have shown, it is not

difficult to see that Ross comes out ahead of Beauchamp and Childress.

It has already been argued that 4PA runs into a number of problems with their

approach to justification; Ross, defending a different theory, avoids those issues.[546]

Beauchamp and Childress' reliance on normative specification is problematic for a

number of reasons: we are mostly left in the dark about how the process of normative

specification works; 4PA contains a host of concrete examples of specifications, but lacks

a systematic description of how the process works. Consequently, it is unclear what the

criteria for good versus bad specifications are (apart from the appeal to coherence with

general norms, which, as argued above, is only if limited help). Ross' method of

specification is descriptive, and even though it is a matter of judgment which empirical

facts have ethical relevance, this will in most cases be uncontroversial. This allows for an

---

[546] However, Ross' account of self-evidence faces challenges of its own that would require a defense of
their own.

objective assessment of specifications based on whether they adequately account for the empirical context, making Ross' method of specification more transparent and objective than 4PA. Next, Beauchamp and Childress are unclear about the role that general norms play for practical deliberation. In order to make sense of the claim that specified rules are open to modification from additional normative factors that are relevant in a certain situation, general norms – e.g. that we should avoid causing harm – have to have the ability to make a direct difference. At the same time, Beauchamp and Childress deny in their justification for the need of specification that general norms can directly apply to concrete situations. Ross' more transparent theory avoids these problems. Since knowledge of *prima facie* duties plus awareness of relevant circumstances is generally enough to make a practical judgment, there is no ambiguity about the role *prima facie* duties play. 4PA, as argued above, defends a form of limited relativism (limited because there are constraints on what moral code is being adopted insofar as particular moralities may not contradict general norms), whereas Ross believes that there is only one correct application of *prima facie* duties in any given situation.[547] The weight of the *prima facie* duties depends on context, and this allows for flexibility in regards to our duty all things considered without giving up the belief in an objective moral truth. Since relativism in moral theory is a view that is of limited appeal to most, any viable alternative that eschews such a commitment – such as Rossian intuitionism – will by many be preferred over 4PA.

---

[547] Even though he admits that there are epistemic difficulties in finding out what the right thing to do all things considered is. See Ross (1930), 134.

8.11 Conclusion

Ross offers a theory that is not subject to the same criticisms as Beauchamp and Childress. As stressed above, this alone does not establish that intuitionism is the most plausible theory for bioethics, but it at least creates a presumptive case in its favor. A complete assessment, a demanding task that would require a dissertation of its own, would have to show that its resources (its understanding of *prima facie* duties, its conception of specification, its theory of justification and of what status rules possess etc.) are sufficient to deal with all the concrete issues that Beauchamp and Childress address in *Principles*. The present conclusion however allows making good on the promise mentioned at the beginning of the chapter: in principle, there is no obstacle for intuitionism to give guidance in areas of applied ethics. Other than it might seem at first sight, it is not sufficient for intuitionists to refer to theories such as 4PA as an already worked out answer to this challenge. Beauchamp and Childress' view suffers from internal problems and turns out to differ from Rossian intuitionism in important respects.

This conclusion is not only of relevance for ethical intuitionism, it also contributes to a larger debate about the nature of applied ethics. The underlying issue is this: is applied ethics an autonomous discipline, or can it be reduced to normative ethics plus awareness for specific circumstances? Beauchamp and Childress would argue for the first claim and Ross for the second. Even if Ross is right and intuitionism can without fundamental modification be used to solve problems in applied ethics, this does not render those applied issues trivial. Taking into account context can be a demanding task, especially in areas such as biomedical ethics where the stakes are high, the facts complicated and the

outcomes often uncertain. Furthermore, biomedical ethics is faced with questions

where Rossian intuitionism is of no help: What is the ethical status of a fetus?

What is the most plausible definition of death? Who should shoulder the burdens

of ever increasing costs for health care? Intuitionism presupposes that these

problems are already solved and addresses only a subset of issues: given that all

relevant information is available, what is our duty all things considered?[548] Given

these limitations, if Ross is right and intuitionism is the most convincing view in

normative ethics, no different theoretical approach above or besides normative

theories is needed in order to answer questions concerning what we should do. In

this picture, doing biomedical ethics does not involve developing a framework of

its own, choosing the best from various normative theories.[549] It simply is the

application of the most convincing normative theory to a confined set of problems.

---

[548] Ross also offers an extensive discussion on what counts as intrinsic goodness that could be fruitful for discussions in biomedical ethics (see Ross (1930), 65-173), but our present discussion is confined to his views on the right.
[549] See Beauchamp (2012), Ch. 9.

Works cited

Anscombe, G.E.M. "Modern Moral Philosophy". *Philosophy* 27 (1958): 1-19.

Aquinas, T. *Summa Theologiae* (Vol. 17). Seattle: The Aquinas Institute, 2012.

Audi, R. "Intuition, Inference, and Rational Disagreement in Ethics". *Ethical Theory and Moral Practice*, *5* no. 11 (2008): 475-492.

Audi, R. *Moral Perception.* Princeton: Princeton University Press, 2008.

Audi, R. "Self-Evidence". *Philosophical Perspectives* 13 (1999): 205-228.

Audi, R. "Skepticism About a Priori Justification: Self-Evidence, Defeasibility". In J. Greco, *The Oxford Handbook of Skepticism* (pp. 149-175). Oxford: Oxford University Press, 2008.

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

Ayer, A. *Language, Truth and Logic.* London: V. Gollancz, 1936.

Beauchamp, T. L., & Childress, J. F. *Principles of Biomedical Ethics* (7 ed.). Oxford: Oxford University Press, 2012.

Bedke, M. "Ethical Intuitions: What They Are, What They Are Not, and How They Justify". *American Philosophical Quaterly*, *3* no. 43 (2012): 253-270.

Bedke, M. "Intuitive Non-Naturalism Meets Cosmic Coincidence". *Pacific Philosophical Quaterly*, *2* no. 90 (2009): 188-209.

Bedke, M. "Intuitional Epistemology in Ethics". *Philosophy Compass*, 5 no. 12 (2009): 1069-1083.

Binkley, L. *Contemporary Ethical Theories.* New York: Philosophical Library, 2009.

Brandt, R. *Ethical Theory - The Problems of Normative and Critical Ethics.* Englewood Cliffs: Prentice-Hall, 1959.

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

Chomsky, N. *Syntactic Structures.* New York: Mouton, 1979.

Cuneo, T. and Shafer-Landau, R. "The Moral Fixed Points: New Directions for Moral Nonnaturalism". *Philosopical Studies* 34 (2014): 1-45.

Dancy, J. *Ethics without Principle.* Oxford: Oxford University Press, 2014.

Dancy, J. *Moral Reasons.* Oxford: Blackwell, 1993.

Darwall, S. "Ethical Intuitionism and the Motivation Problem". In P. Stratton-Lake, *Ethical Intuitionism - Re-evaluations* (pp. 248-270). Oxford: Oxford University Press, 2002.

Dreyfus, H. L. *What Computers Can't Do.* New York: Harper and Row, 1972.

Dworkin, R. *Taking Rights Seriously.* Cambridge: Harvard University Press, 1977.

Dwyer, S. "How Good is the Linguistic Analogy?" In P. Carruthers, S. Laurence, & S. Stich, *The Innate Mind: Culture and Cognition* (Vol. 2, pp. 145-167). Oxford: Oxford University Press, 2006.

Edmonds, D. *Would You Kill the Fat Man?* Princeton: Princeton University Press, 2013.

Fitch, G. *Saul Kripke.* New York: Acumen, 2004.

Foot, P. "Euthanasia". *Philosophy and Public Affairs* 32 (1977): 85-112.

Foot, P. "Moral Beliefs". *Proceedings of the Aristotelian Society* 59 (1958): 83-104.

Foot, P. "The problem of abortion and the doctrine of double effect". *Oxford Review* 5 (1967): 5-15.

Frankena, W. *Ethics* (2 ed.). Englewood Cliffs: Prentice-Hall, 1963.

Gert, B., Culver, C. M., & Clouser, K. D. *Bioethics: A Systematic Approach* (2nd Edition ed.). Oxford: Oxford University Press, 2006.

Hare, R. *Sorting out Ethics.* Oxford: Oxford University Press, 1998.

Harman, G. "Moral Grammar". *unpublished manuscript* .

Hauser, M. *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong.* New York: Harper Collins, 2006.

Hooker, B. *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality.* Oxfod: Oxford University Press, 2003.

Huemer, M. *Ethical Intuitionism.* Hampshire: Palgrave Macmillan, 2005.

Huemer, M. "Revisionary Intuitionism". *Social Philosophy and Policy, 1* no. 25 (2008): 368-392.

Hurka, T. "Audi's Marriage of Ross and Kant". In M. Timmons, J. Greco, E. Eslick, & A. Mele, *Rationality and the Good: Critical Essays on the Ethics and Epistemology of Robert Audi* (pp. 64-72). Oxford University Press, 2007.

Hurka, T. "Common Themes from Sidgwick to Ewing". In T. Hurka, *Underivative Duty - British Moral Philosophy from Sidgwick to Ewing* (pp. 6-25). Oxford: Oxford University Press, 2011.

Hurka, T. "Introduction". In T. H. (ed), *Underivative Duty - British Moral Philosophers from Sidgwick to Ewing* (pp. 1-5). Oxford: Oxford University Press, 2011.

Hurka, T. "Normative Ethics: Back to the Future". In B. Leiter, *The Future for Philosophy* (pp. 246-265). Oxford: Oxford University Press, 2004.

Irwin, T. *The Development of Ethics, Volume 3: From Kant to Rawls.* Oxford University Press, 2009.

James, S. M. *An Introduction to Evolutionary Ethics.* Wiley-Blackwell, 2011.

Johnson, O. A. "Ethical Intuitionism - A Restatement". *Philosophical Quaterly* 28 (1957): 193-203.

Joyce, R. "The Many Moral Nativisms". In K. Sterelny, R. Joyce, B. Calcott, & B. Fraser, *Cooperation and its Evolution* (pp. 549-572). Cambridge: MIT Press, 2013.

Kahane, G. "Evolutionary Debunking Arguments". *Nous*, *1* no. 45 (2011): 103-125.

Knobe, J. "Intentional Action in Folk Psychology". *Philosophical Psychology*, 2 no. 16 (2003): 309-325.

Kohlberg, L. *Essays on Moral Development. Volume 1: The Philosophy of Moral Development.* New York: Harper and Row, 1981.

Kohlberg, L. *Essays on Moral Development. Volume 2: The Psychology of Moral Development.* New York: Harper and Row, 1984.

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

Krikpe, S. A. *Naming and Necessity.* Cambridge: Harvard University Press, 1980.

Le Guin, U. *The Left Hand of Darkness.* Seattle: Ace Trade, 1957.

Mackie, J. L. *Ethics - Inventing Right and Wrong.* London: Penguin Books, 1977.

McKeever, S., & Ridge, M. *Principled Ethics - Generalism as a Regulative Idea.* Oxford: Oxford University Press, 2006.

McNaughton, D. "Intuitionism". In H. Lafollette, *The Blackwell Guide to Ethical Theory* (2nd Edition ed., pp. 287-310). London: Wiley-Blackwell, 2013.

McNaughton, D. "Unprincipled Ethics". In B. Hooker (ed), *Moral Particularism* (pp. 267-289). London: Clarendon Press, 2000.

McNaughton, D. *Moral Vision: An Introduction to Ethics.* London: Wiley-Backwell, 1988.

Mikhail, J. *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment.* Cambridge: Cambridge University Press, 2011.

Mikhail, J. "New Perspectives on Moral Cognition: Reply to Zimmerman, Enoch, and Chemla, Egre, and Schlenker". *Jerusalem Review of Legal Studies* 13 (2013): 1-49.

Mikhail, J. "The poverty of the moral stimulus". In W. Sinnott-Armstrong, *Moral Psychology: The Evolution of Morality: Adaptation and Innateness* (pp. 81-91). Cambridge: MIT Press, 2008.

Mikhail, J. "Universal moral grammar: Theory, evidence, and the future". *Trends in Cognitive Sciences* 11 (2007): 143-152.

Moore, G. *Ethics.* London: Routledge, 1912.

Nagel, T. "Mortal Questions". In T. Nagel, *Mortal Questions* (pp. 128-141). Cambridge: Cambridge University Press, 1979.

Nelson, M. T. "Intuitionism and Conservatism". *Metaphilosophy*, 3 no. 21 (1990): 282-293.

Nowell-Smith, P. H. *Ethics.* Melbourne: Penguin Books, 1954.

Piaget, J. *The Moral Judgment of the Child.* New York: Free Press, 1965.

Pickard-Cambridge, W. "Two Problems About Duty (I.)". *Mind* 41 (1932): 72-96.

Prinz, J. "Against Moral Nativism". In D. Murphy, & M. Bishop, *Stich and His Critics* (pp. 167-189). London: Wiley-Blackwell, 2009.

Railton, P. "Moral Realism". *Philosophcial Review*, *2* no. 95 (1986): 163-207.

Rawls, J. *A Theory of Justice - Revised Edition.* Cambridge: Harvard University Press, 1999.

Roeser, S. *Moral Emotions and Intuitions.* New York: Plagrave Macmillan, 2011.

Ross, D. *Aristotle.* London: Routledge, 1923.

Ross, D. *Foundations of Ethics.* Oxford: Oxford University Press, 1939.

Ross, D. "The Basis of Objective Judgements in Ethics". *International Journal of Ethics* 37 (1927): 113-127.

Ross, D. *The Right and the Good.* Oxford: Oxford University Press, 1930.

Sandel, M. *Justice: What's the Right Thing to Do?* London: Farrar, Straus and Giroux, 2011.

Shafer-Landau, R. *Moral Realism: A Defence.* Oxford: Oxford University Press, 2003.

Shafer-Landau, R. "Evolutionary Debunking, Moral Realism, and Moral Knowledge". *Journal of Ethics and Social Philosophy* 1 no. 7 (2012): 1-37.

Sinnott-Armstrong, W. *Moral Skepticisms.* Oxford: Oxford University Press, 2006.

Sinnott-Armstrong, W. "An Empirical Challenge to Moral Intuitionism. In J. G. Hernandez (ed), *The New Intuitionism* (pp. 11-28). London: Bloomsbury Academic, 2011.

Smith, A. *Theory of Moral Sentiments.* Cambridge: Cambridge University Press, 2002/ 1759.

Sripada, C. S. "Adaptationism, Culture, and the Malleability of Human Nature". In P. Caruthers, S. Laurence, & S. Stich, *The Innate Mind - Foundations and the Future* (pp. 311-330). Oxford: Oxford University Press, 2008.

Stevenson, C. L. *Ethics and Language.* Oxford: Oxford University Press, 1944.

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

Stratton-Lake, P. "Introduction". In D. Ross, *The Right and the Good* (pp. ix-l). Oxford: Oxford University Press, 1930/2002.

Street, S. "A Darwinian Dilemma for Realist Theories of Value". *Philosophical Studies* 1 no. 127 (2006): 109-166.

Warnock, G. J. *Contemporary Moral Philosophy.* New York: St Martin's Press, 1967.

Warnock, M. *Ethics Since 1900.* Oxford: Oxford University Press, 1960.

Williams, B. "What does intuitionism imply?" In B. Williams, *Making Sense of Humanity* (pp. 182-191). Cambridge: Cambridge University Press, 1995.

Zimmerman, A. "Mikhail's Naturalized Moral Rationalism". *Jerusalem Review of Legal Studies* 23 (2013): 1-22.

Zimmerman, A. *Moral Epistemology.* London: Routledge, 2010.