THE PROBLEM(S) WITH PRINCIPLES:
TOWARDS A SKILL-BASED ACCOUNT OF MATURE MORAL AGENCY

by
Jennifer Lyn Cole Wright

A thesis submitted to the Department of Philosophy
and The Graduate school of The University of Wyoming
in partial fulfillment of the requirements
for the degree of

MASTER OF ARTS
in
PHILOSOPHY

Laramie, Wyoming
April, 2006
Let us start with a bit of terminology:

(AMJA) an *appropriate moral judgment or action*\(^1\) is one that is fully sensitive to the moral reasons present in a particular situation\(^2\)

(MMA) a *mature moral agent* is a moral agent that (reliably) forms appropriate moral judgments and actions

Much debate in normative ethics has centered on question of how moral agents ideally form AMJAs. Principle-based moral theories\(^3\) commonly agree that moral principles (MP) play a central role in AMJAs (though which and how many principles is debated) insofar as MPs identify the moral reasons MMAs are required to consider.\(^4\)

Ideally speaking, for principle-based theories AMJAs are those moral judgments and

---

\(^1\) Strictly speaking, judgments and actions are probably going to be “sensitive to moral reasons” in different ways – but for the purposes of this paper I am not going to be concerned with the difference. Though I could simply restrict my discussion to judgments, moral reasoning is a form of practical reasoning and, as such, is ultimately linked to action. As such, I think it is important to include both.

\(^2\) I think that what AMJAs are actually sensitive to are the moral *features* of particular situations which are *reasons for* (insofar as they obligate or recommend) particular AMJAs. As such, throughout I will talk about sensitivity to moral (reason-giving) features instead of to moral reasons.

\(^3\) The different types of principle-based theories will be outlined briefly in Chapter 1 and various sorts of principles will be discussed in detail in Chapter 2.

\(^4\) According to principle-based theories, MPs are (ideally) fully sensitive to moral reasons insofar as they identify the sorts of grounding-to-moral feature relations that, when present in a particular situation, are conclusive or presumptive reasons for MMAs to judge or do X.
actions that *conform* to MPs. This conformity may be understood in either one (or both) of two ways: as involving *moral guidance or normative authority.*

That is, AMJAs conform to MPs just in case either they are *guided* by such principles or they are *rendered appropriate* by such principles.

My overall aim is to show that there is reason (indeed, many reasons) to think that MPs are inadequate to this task: that is, because of the complexities and subtleties of actual moral life, MPs are unable to either adequately *generate* or *identify* AMJAs. Indeed, I will argue that conformity to MPs is neither *necessary* nor *sufficient* for the achievement of AMJAs. As such, AMJAs should not be spelled out in terms of conformity to moral principles.

To this end, I will begin in Chapter 1 by laying out the theoretical territory, summarizing what I take to be some of the primary criticisms that have been made in the literature against principle-based theories. This will allow us to distinguish the sorts of criticisms in which I am primarily interested from those that are orthogonal to my purposes. Chapter 2 will then weave together what I take to be the most serious problems faced by any attempt to cash out AMJAs in terms of conformity to MPs. I take it that the burden is currently on my shoulders to provide reasons to think that MPs are inadequate: the goal of Chapter 2 is to discharge this burden. Of course, there are many ways to discharge a burden. While one can, for instance, demonstrate that one’s opponent’s view is false, one can also simply shift in one way or another the burden onto the shoulders of

---

5 What I shall call ‘moral guidance’ has in the literature sometimes been referred to as MPs’ ‘investigative’ function (Shafer-Landau 2003).

6 Many of the reasons I discuss have already been given by other philosophers at one time or another in various forms. I nonetheless think it instructive to bring them all together and integrate them into one coherent, organized, and sustained critique. Of course, my aim is not merely to summarize existing criticisms of principles: throughout, I both introduce novel criticisms and elaborate upon more familiar, but undeveloped, ones.
one’s opponent. Accordingly, I take it that my objective in Chapter 2 will be to either demonstrate that MPs are inadequate and/or sufficiently shift the burden onto the defender of principle-based theories by providing reasons to be pessimistic about the ability of MPs to perform one or another of their functions. This should make it clear how I intend to proceed; namely, I will argue that either MPs are inadequate or defenders of principle-based theories must discharge a significant explanatory burden before earning the right to claim adequacy. Should I succeed in this endeavor, then we must look elsewhere for a satisfactory account of AMJAs and mature moral agency.

Accordingly, in Chapter 3, I will suggest that we must look instead to the development of something like a refined “moral sense”: one that, unlike the knowledge of principles, involves processes that reliably enable full sensitivity to the (relevant) moral features present in particular situations. I will outline, in broad strokes, such an account, relying on insights gained from empirical research on various forms of skill-mastery. I will consider a few objections and conclude with what I take to be the most important implications of both my negative and positive accounts.
Principle-based theories come in many shapes and sizes. The primary division is between consequentialist theories (e.g., Mill and Singer), which (roughly speaking) locate rightness/wrongness in consequences of actions and deontological theories (e.g., Kant and Korsgaard), which (roughly speaking) locate rightness/wrongness in the actions themselves. The former division specializes in MPs that identify appropriate moral ends (which our actions must then strive to achieve); the latter specializes in MPs that are more action-oriented, identifying those actions that are right (e.g., being compassionate) and those that are wrong (e.g., lying).

Each of these divisions can be further broken down into more specific sorts of theories. For example, there are a wide variety of consequentialist accounts: rule versus act consequentialism, those whose moral end is pleasure, happiness, or preference-satisfaction (to name only a few), and those who measure consequences in terms of either actual or expected outcomes. Deontological theories are likewise diverse, including such theories as universalization (e.g., Gewirth), inviolability of persons (e.g., Kant, Kamm),
natural law theories (e.g., Aquinas), and divine command theories (e.g., Adams). There is also intuitionism, which varies with respect to whether we have prima-facie duties (e.g., Ross) or prima-facie rights (e.g., Thomson) and contractualism (e.g., Rawls and Scanlon), which varies with respect to which guiding principle(s) are advocated (e.g., the principle of opportunity, equality, distributive justice, etc.), as well as how such principles are to be (ideally) chosen or constructed (and by whom).

These different approaches differ with respect to, among other things, whether (and how) MPs are discovered or constructed, as well as how they are justified. In addition, while some rely on one overarching MP to guide and/or track the appropriateness of AMJAs (such as the principle of utility or the Categorical Imperative), others embrace moral pluralism, formulating complicated sets of interacting MPs of different kinds. But, regardless of these many differences, one thing unites the general principle-based approach: for each of these approaches MPs lie at the heart of appropriate moral judgments and actions. In other words, the alleged job of MPs is to identify the right and wrong-making features of the moral domain, and AMJAs are those moral judgments and actions that conform to these principles (though, as I discussed above, how this conformity is understood varies).

Many different criticisms of principle-based theories have been offered over the years. The most prevalent criticisms take the following three forms: 1) those that

---

7 Not included as a principle-based deontological theory is “act-deontology”, often found in the existentialist writings, that focuses on the rightness/wrongness of particular acts without appeal to general principles, which operate as “underlings” to some supreme principle.

8 Of course, there are some readings of utilitarianism and Kantianism that recognize the need for a plurality of principles.

9 Some see the job of moral principles to be identifying that which is morally relevant – which may be distinct from identifying right/wrong-making features, because moral relevance does not by itself yet indicate rightness/wrongness (Montague 1986). This will come up (briefly) towards the end of Chapter 3.
challenge the requirements that MPs place on moral agents, 2) those that express skepticism about the guidance provided by MPs, and 3) those that question whether MPs can successfully capture general moral truths.

Probably the best known set of criticisms is the first. In particular, the most central requirement to be criticized is the requirement of impartiality, a criticism that principle-based theories are especially susceptible to because of their focus on the general, as opposed to the situation-specific, details of moral situations. In order to adequately evaluate moral situations, MMAs are called upon to abstract away from and ignore as irrelevant non-generalizable particular details: that is, they are called upon to be impartial, not letting such particular details influence their moral judgments and actions. This requirement for impartiality has given rise to several distinct (though related) criticisms.

Cottingham (1983), among others, has argued that the impartiality requirement leads to morally undesirable consequences. Because impartiality requires that we consider our own lives, goals, and interests as being on par with others’ lives, goals, and interests, the fact that these things are ours becomes morally irrelevant. Take, for example, the classic utility principle: strict adherence to this principle prohibits moral agents from considering their own needs and interests as carrying any more weight than the needs and interests of others in their utility calculations. Should I buy a book that I have been wanting to read? Only if my money wouldn’t be better spent helping meet the needs of others. Should I apply to a graduate program to pursue a higher degree in philosophy? Only if my time wouldn’t be better spent helping others to pursue their interests. In short, moral agents have no moral reason to place their own needs and
interests above the needs and interests of others. But, argues Cottingham, such a position is “grotesquely impracticable” and undesirable: the pursuit of one’s quality of life, which is something that should carry important moral weight, becomes morally indefensible.

Others (e.g., Nussbaum 1985, Railton 1984\textsuperscript{10}, Stocker & Hegeman 1996, Williams 1985) have argued, in a similar vein, that the impartiality requirement leads to alienation from one’s own projects/concerns, from one’s own and others unique value, and from the particularities and complexities (i.e., the richness) of moral life. Insofar as impartiality requires that we distance ourselves from our personal goals and interests, holding them as on par with the goals and interests of everyone else, we become alienated from them. We are likewise estranged from our affections for particular others, whomever they might be. This prohibits the attribution of “specialness” to particular relationships, rendering us unable to justify the recognition and appreciation of the uniqueness and irreducible worth of particular individuals (see also Sommers 1986).\textsuperscript{11}

Finally, some (e.g., Nussbaum 1985) have argued that impartiality alienates us from the particularities of moral life: particularities that are prior to and more significant than (and perhaps not entirely reducible to) general principles. For it is within the particularities (and not the generalities) that the moral richness and meaningfulness of

\textsuperscript{10} Railton (1984), while recognizing this worry, argues to save consequentialism from it by denying that moral agents must guide their moral judgments and actions according to the utility principle (that is, he denies conformity as moral guidance), stating instead that our judgments and actions must simply ultimately maximize utility (normative authority), however it is that we decide upon them.

\textsuperscript{11} The Utilitarian response that special relationships are morally called for because they promote general happiness (as people are generally happier when they have close relationships than when they do not) is considered by most critics to be a further source of alienation: if it is only morally appropriate for me to show a particular other special consideration just in case this love promotes happiness, then I must be continually vigilant for signs that the relationship is not successfully promoting happiness (or for other relations that might promote more). In other words, the worth of the relationship becomes located solely in the happiness it promotes and not in the actual individual worth of the persons involved.
individual lives is both discovered and created – all of which must be stripped away if we are to attempt to assume the “view from nowhere” (Nagel 1997).

For the most part, I will not be concerned with this class of criticisms (although the issue of the uniqueness of situations and agents will play a role in what follows). Instead, I will be more concerned with the other two classes: those aimed at the adequacy of MPs themselves. More specifically, I will concern myself with those criticisms that challenge the capacity of MPs to provide either moral guidance or normative authority.

With regards to the former challenge, many (e.g., McDowell 1998, Garfield 2000) have argued that MPs are not the right sorts of entities to provide answers about what to do in particular situations (at least not by themselves). To hold that ‘lying is wrong’, for example, does not itself give me the information I need to know when (and under what circumstances) such a principle applies. Nor does it tell me what specifically to do in particular situations (even when it applies). Something else is needed – something to bridge the gap between the general and the particular. This “something else” has been cashed out in a number of different ways: in terms of good moral judgment, developed moral sensitivity, acute moral perception, and so on.12 However this gets cashed out, it follows that moral agents must possess a faculty or ability (or set of faculties or abilities) that goes beyond the mere adoption of or commitment to particular MPs and whose function is to guide MMAs in recognizing 1) which MPs to apply in particular moral situations and/or 2) what AMJAs follow from the application of those principles (of course, my own inclination is to push further than this: In Chapter 2, I will argue that, given such a faculty or ability, MPs are no longer needed for AMJAs).

12 For variations on these themes, see (for example) Blum 1991, Garfield 2000, McDowell 1997, Mitchell 1963.
This concern is a different sort of worry for consequentialist and deontological theories. It is a worry for consequentialist theories because of their focus on achieving particular moral consequences: the problem of getting from the principle “maximize utility” to any particular judgment/action that in fact maximizes utility is a problem that no consequentialist theory has yet managed to sufficiently address. For the deontologists, the worry is action-oriented MPs admit of exceptions. And it is not clear how MPs, by themselves, can inform moral agents when they apply and when they do not: that is, when there are exceptions and when there are not. In addition, pluralistic accounts face the problem of how moral agents are to determine which of their many MPs apply in any particular situation.

With regards to the latter challenge, some have objected that “morality can get along perfectly well without principles, and that the imposition of principles on an area that doesn’t need them and is likely to lead to some sort of distortion” (Dancy 2004, 2). In a related vein, in Chapter 2 I will argue that moral situations are too complex for general MPs – it is simply not the case that principles are going to hold across situations in the way that principle-based theories must require them to. There simply are not the sorts of invariant relations (such as ‘lying’ to ‘wrongness’) necessary to make MPs work. While this challenge appears to be more of a worry for deontological theories (as consequentialist theories remain silent on the issue of which actions, in particular, are required), I will argue in Chapter 2 that it is a worry for consequentialist theories as well.

In summary, my critique of MPs (to follow in Chapter 2) will do two things: 1) it will provide the groundwork for a discussion of MPs, not as an exhaustive exploration of different theoretical approaches, but rather as a canvassing of conceptual space, and 2) it
will explore and expand upon existing criticisms, as well as introduce novel criticisms, that speak to the issue of whether AMJAs are those moral judgments and actions that conform to MPs, either by way of moral guidance and/or normative authority. Often I will leave implicit which of these two tasks I am performing at a given moment, since my central concern will be simply to present the strongest arguments against the adequacy of moral principles that I can find. The goal of this portion of my project is to offer a broad and inclusive view of the ongoing debate about MPs in order to demonstrate that we have reason to believe that the fate of the principle-based view that MPs are that to which AMJAs must conform is grim. As such, I think that it becomes crucial for philosophers and psychologists interested in understanding mature moral agency to explore more promising avenues. The positive portion of my project, gesture at in Chapter 3, will take us at least part of the way down one of those avenues.
CHAPTER TWO

~ The Problem(s) with Principles ~

“…Moral [principles]…have no place in the lives of saints or complete sinners. For saints are not still learning how to behave and complete sinners have not yet begun to learn.” (Ryle 1971)

The goal of this chapter is to develop my negative account (with Chapter 3 being dedicated to the development of my positive account). In other words, the goal for this chapter is to provide reasons to believe that conforming to moral principles is neither necessary nor sufficient for appropriate moral judgments and actions (AMJAs). As such, I will argue, we should look elsewhere for an account of AMJAs: in Chapter 3, I will at least gesture at what such an alternative account might look like.

I. LAYING THE GROUNDWORK

If we are to mount a sustained critique of moral principles, we must first have an account of both what moral principles are and in what forms they might come. My objective in laying this groundwork is not to merely recapitulate different philosopher’s accounts of moral principles, but rather to canvas (as thoroughly as possible) conceptual space.

I.1. ~ What is a moral principle?

In offering a definition of moral principles, it is important to attempt to capture what we take to be their essential components. Though rough, I take the following to do just this:
A proposition (or sentence-like structure) $p$ is a genuine moral principle (MP) iff $p$

(i) is (necessarily) true

(ii) contains both (what we might call) grounding and moral concepts

(iii) is normative

(iv) is (robustly) generalizable

(v) is theory-grounded (i.e., non-arbitrary)

(vi) is specified (or specifiable) in advance (i.e., not ad hoc)

Perhaps it is worth saying a bit more about a few of the above conditions: With respect to (ii), an example of this would be the MP ‘lying is wrong’, which identifies both a grounding feature (lying) and a moral feature (wrongness). Of course, the grounding feature could itself be a moral feature (e.g., ‘promoting justice is good’), though most will not be. With respect to (iii), the point is that MPs cannot be merely descriptive, but must entail certain prescriptive imperatives. Of course, strictly speaking, MPs can only obligate moral agents to perform certain actions through the derivation of an imperative, which requires an additional (if unspoken) premise:

(P1) Lying is wrong.

(P2) (Roughly) Moral agents should not do that which is wrong.

(C) Therefore, (moral agents:) do not lie (or tell the truth).

For my purposes, I will basically ignore the need for (P2), treating imperatives as alternative forms of the MPs from which they are derived. In other words, I will use imperatives such as ‘tell the truth’ and ‘do not lie’ and the principles they were derived from (‘lying is wrong’) interchangeably.
With respect to (iv), this basically means that MPs must apply to more than one situation: whenever the grounding feature(s) contained within the MP is present, the moral feature(s) must be instantiated\textsuperscript{13}. And, finally, (v) and (vi) mean that MPs cannot simply be generated “willy nilly” or “on the fly”. Not unlike other principle-generating domains of enquiry, the adequacy of any principle-based ethical theory must be determined by whether or not it can articulate theory-grounded (non-arbitrary) general principles that specify in advance (non ad hoc) a systematic means for accurately determining the appropriateness of various moral judgments and actions.

\textbf{1.2. ~ In what forms do moral principles come?}

Many different forms (or types) of MPs have been proposed over the years. They can largely be captured along the following two dimensions: (1) The invariant or variant dimension and (2) The action-specific or abstract dimension.

With respect to the first dimension, invariant MPs are typically the sort of MPs in which normative ethical theories are interested. Invariant MPs are \textit{universally quantified conditionals}: propositions expressing relations claimed to invariantly obtain between certain grounding features and certain moral features (for every $x$, if $GFx$, then $MFx$).\textsuperscript{14} For example, ‘killing is (always) wrong’ or ‘for every death, if it is a killing, then it is wrong’ is an invariant MP because it states that the fact that a particular act is an act of killing invariably makes that action wrong. Invariant MPs are thus \textit{fully general}: whenever the grounding feature(s) is present, the moral feature(s) is always instantiated.

We can thus summarize as follows:

\textsuperscript{13} With exceptions (if there are any) being noted and an adequate explanation for why they are legitimate exceptions given. More will be said on this issue as we go.

\textsuperscript{14} This definition of (invariant) MPs comes largely from Russ Shafer-Landau (1997).
A MP is an *invariant* MP iff it identifies a grounding feature(s) that always instantiate(s) a given moral feature(s).

Variant MPs, on the other hand, are *heuristics* (often referred to as ‘rules-of-thumb’). Though certainly not as desirable to normative ethicists as invariant MPs, normative theories have nonetheless had occasion to appeal to variant MPs for a number of reasons, the most common of which is to get around the challenge of human limitations. For example, Mill (1979) advised that even though the one “true” MP was the principle of utility, an invariant MP, since utility calculus is extremely difficult (if not impossible) for people to manage, they must rely instead on rough approximations, or variant MPs.

Of course, by “rough approximation” we could mean one of two things: first, that variant MPs identify grounding feature(s) that *sometimes* (though not always) instantiate(s) moral feature(s); second, that variant MPs identify feature(s) that inexactely *approximate* grounding feature(s) that instantiate(s) moral feature(s). In other words, variant MPs either get it exactly right only sometimes or they always get it approximately, but never exactly, right.\(^{15}\) Thus, we can summarize as follows:

(2) A MP is a *variant* MP iff it identifies (a) a grounding feature(s) that *sometimes* (though not always) instantiate(s) a given moral feature(s), or (b) feature(s) that inexactely *approximate* a grounding feature(s) that instantiate(s) a given moral feature(s).

Normative theories are typically not interested in variant MPs, because they lack both the invariance and generalizability to which they aspire. Hence, although I will consider the

---

\(^{15}\) The latter option could be thought of as something like rounding numbers: when adding 305.1 and 263.6, rounding to 305 and 264 gives the answer 569, which is close to the actual answer – 568.7 – but not exact.
usefulness of variant MPs for moral discourse at the end of Chapter 3, our discussion of MPs will otherwise be restricted solely to the invariant sort.

With respect to the second dimension noted above, it is important to keep in mind the difference between those invariant MPs that are action-specific and those that are abstract. Action-specific principles identify grounding-to-moral feature(s) relations that obligate moral agents to perform (or refrain from performing) certain actions. For instance, ‘lying is wrong’ identifies an invariant grounding-to-moral feature relation that directly obligates moral agents to act in accordance with the imperative ‘do not lie’ (or ‘tell the truth’). Because it is an invariant MP that obligates moral agents to perform (or refrain from performing) specific actions, ‘lying is wrong’ is what we might call a concrete MP. In general,

\[(3) \quad \text{A MP is a concrete MP iff it (a) is an invariant MP and (b) obligates moral agents to perform (or refrain from performing) specific actions.}\]

As we will see, concrete MPs admit of exceptions. One way to handle such exceptions is to deny the invariance and/or the generalizability of concrete MPs, claiming that they are variant MPs instead. However, ethical theorists commonly prefer to maintain invariance and generalizability by modifying concrete MPs into qualified MPs.

Qualified MPs come in two forms: prima facie (or pro tanto) principles and specified principles. Classically, Ross (1930) introduced prima facie MPs as a way of handling the sorts of exceptions that concrete MPs run into. One way to conceive of exceptions is that they involve a conflict of principles. Prima facie MPs provide a way for resolving such conflicts because, though they identify invariant grounding-to-moral
feature(s) relations, they regard these relations as ‘presumptive’. For example, if lying is wrong, then it always gives rise to a *presumptive* obligation to refrain from doing it: but it may nonetheless be the case that other principle(s) are more important in a particular situation, thereby overriding the obligation to so refrain. Importantly, this does not mean that prima facie MPs are variant principles: the wrongness of lying remains, even though the obligation to refrain from lying can be overridden. Thus, prima facie MPs *recommend* rather than obligate moral agents to perform or refrain from performing certain actions. In general, then,

\[(4) \quad \text{A MP is a prima facie MP iff it (a) is an invariant MP, (b) recommends specific actions to be either performed or refrained from, (c) this recommendation can be overridden by other prima facie MPs, and (d) if not so overridden, it becomes an obligation.}\]

Specified MPs come in two flavors: ‘narrow’ and ‘wide’. Narrowly specified MPs restrict the grounding features. That is, the grounding feature(s) identified by a particular MP is/are narrowed so as to specify fairly determinate forms of the more general grounding feature(s) identified by concrete MPs. For example, ‘lying is wrong’ becomes ‘lying* is wrong’, where ‘lying*’ denotes a particular kind of lying. Widely specified MPs, on the other hand, include an enumeration of all overriding and mitigating exceptions to the more general grounding feature(s) identified by concrete MPs. For example, ‘lying is wrong’ becomes ‘lying is wrong, except in circumstances C₁, C₂, ..., Cₙ’. In short:

---

16 In the literature, ‘presumptive’ reasons are sometimes referred to as ‘contributory’ reasons. These are contrasted with ‘absolute’, or ‘subsumptive’, reasons, which are identified by concrete MPs. See, e.g., Dancy 2004.
(5) A MP is a *narrowly specified* MP iff (a) it is an invariant MP and (b) if X is a grounding feature(s) identified by a concrete MP, then it specifies the *kinds of X* that instantiate(s) a given moral feature(s).

(6) A MP is a *widely specified* MP iff (a) it is an invariant MP and (b) if X is a grounding feature(s) identified by a concrete MP, then it specifies the *conditions under which* X instantiates a given moral feature(s).

Let us now briefly consider abstract MPs. Abstract MPs do not obligate moral agents to perform or refrain from performing specific actions: rather, they orient moral agents towards particular moral ends (commonly construed as the ‘ends’ of morality), towards which a wide range of actions may be appropriate. An example of an abstract MP is ‘maximize utility’. In the case of ‘maximize utility’ there are no specific actions that moral agents are called upon to perform or refrain from performing; rather, they are called upon to perform *whatever* actions promote or achieve this end (as well as refrain from performing *whatever* actions might thwart it, as it were). In general, then,

(7) A MP is an *abstract* MP iff it (a) is an invariant MP that (b) recommends or obligates an unspecified set of actions which promote or achieve a specified moral end.

II. THE CRITIQUE OF PRINCIPLES

My critique of MPs (specifically, invariant MPs) shall be organized in the following manner:

First, I will consider whether moral judgments and actions must be *guided* by MPs in order to be AMJAs. This will involve considering first whether guidance by MPs

---

17 Strictly speaking, this abstract MP ought to read something like ‘maximizing utility is the highest moral good’. However, for simplicity I will employ its most common form: ‘maximize utility’.
is *sufficient* for AMJAs and then whether it is *necessary*. In addition, it requires that we consider both the different types of principles and the direct and indirect mediation models of moral reasoning. The basic structure of this part of my argument will be as follows:

1) Being a moral judgment or action that is guided by a MP(s) is not *sufficient* for being an AMJA
   
a. Because MPs are too rigid/insensitive, complex, or abstract to adequately guide MMAs.
   
b. This is true regardless of whether we adopt the direct or indirect mediation model of moral reasoning.
   
c. Therefore, AMJAs cannot be formed on the basis of MPs alone – something in addition to (or other than) MPs will be required for moral guidance.

2) Being a moral judgment or action that is guided by a MP(s) is not *necessary* for being an AMJA
   
a. Though many philosophers introduce an additional faculty/ability (good moral judgment, moral sensitivity, and the like) to enable MPs to guide MMAs, an adequate account of good moral judgment/sensitivity makes it clear that it renders MPs obsolete: MPs need play no significant role in the achievement of AMJAs.

Second, I consider whether moral judgments and actions must be rendered appropriate by MPs. I will argue that:
3) Being a moral judgment or action that is rendered appropriate by a MP(s) is not *sufficient* for being an AMJA
   a. Because (most) MPs do not successfully identify *invariant* grounding-to-moral feature(s) relations (therefore, something in addition to, or other than, MPs is needed to track appropriateness).
   b. Even if MPs *did* successfully identify invariant grounding-to-moral features(s) relations they would be either
      i. Too situation-specific to be considered MPs, or
      ii. Too abstract to track appropriateness in the right way.

4) Being a moral judgment or action that is rendered appropriate by a MP(s) is not *necessary* for being an AMJA
   a. There are other ways in which the appropriateness of moral judgments and actions can be tracked. One such route would be a reliabilist account of moral judgments and actions, along the lines of the one proposed by Shafer-Landau (2003). According to such an account, a moral judgment or action is appropriate iff it is the result of a reliable process(es).

   Given the above arguments, I will conclude that either MPs fail to hold up their end of the bargain or principle-based theorists owe us a better explanation of how they do so. Either way, it behooves us to look elsewhere for an alternative account of both the generation and identification of AMJAs.

   **II.1 ~ Conformity as moral guidance**

   (II.1.a) What is *moral reasoning*?
Before moving on, it is important to be clear about how MPs allegedly figure into moral guidance, according to principle-based theories. I will assume that they do so in virtue of playing an important role in moral reasoning.\textsuperscript{18} Thus, we can say that

\begin{quote}
(MG) A given moral judgment or action is guided by a given MP(s) iff the moral judgment or action is the conclusion of (implicit or explicit) reasoning that relies on the MP(s).\textsuperscript{19}
\end{quote}

In this regard, Cullity (1998, 1) writes, “a widespread view is this: good moral thinking identifies correct moral principles under which the moral judge then subsumes particular instances to produce a moral verdict about them.” This is the \textit{direct mediation} model of MPs. On this model, MPs are employed (presumably, as premises) in explicit reasoning whose conclusion is a moral judgment or action. This could involve employing deductive reasoning (in the form of either means-end or rule-case syllogisms), the conclusion of which is a moral judgment or action in response to a particular situation.

An example would be:

\begin{quote}
(P1) ‘Lying is wrong’ (concrete MP)
(P2) Telling X that p (in this particular instance A) would be lying
(C) \textit{Therefore}, I should not tell X that p (in A).
\end{quote}

The direct mediation model also allows for an inductive step to move from a particular case to a MP, as in:

\begin{quote}
(P1) Lying in this \textit{particular instance A} was wrong.
(P2) (Generalization): Lying is wrong.\textsuperscript{20}
\end{quote}

\textsuperscript{18} Perhaps there are other ways besides reasoning in which a MP could guide moral judgments and actions, e.g., as a source of motivation (see, for example, Korsgaard 1997).

\textsuperscript{19} This places a constraint on principle-based theories: if MPs are to guide moral judgments and actions, then MPs must be \textit{manageable}, which is to say that they must be such that a normal agent is able to employ them in his/her moral reasoning.
(P3) Telling X that p (in a particular instance B, where B ≠ A) would be lying.

(C) Therefore, I should not tell X that p (in B).

Some have argued that the direct mediation model allows for a broader conception of moral reasoning. Richardson (2003, 3) defines explicit moral reasoning as “responsibly conducted thinking in which the reasoner attempts to reach a well-supported answer to a well-defined question.” Such a conception may be acceptable, as long as it is able to show that reaching a “well-supported” answer somehow relies (in a substantial way) on MPs.21 In effect, this account would be consistent with MG.

Of course, direct mediation models are not the only available options. Some principle-based theories appeal to ‘indirect mediation’ models instead.22 Let us consider two such models: the periodic guidance model and the internalization model.

According to the periodic guidance model, moral agents need only make reasoned appeals to MPs (a) in times of moral crisis or (b) periodically, when deliberating about (or reflecting upon) one’s character traits, goals, and/or motives. More formally, this model supports the normative claim that

(8) Moral agents must satisfy MG only in certain circumstances (namely, when faced with a crisis and/or when engaged in periodic reflection).

Clearly, this model leaves the majority of our moral judgments and actions — namely, those that are not inspired by crises or periodic reflection — unprincipled. Those that are

---

20 Ross (1930) called this process “inductive intuition”. According to Ross, when we recognize a particular act of lying as wrong, what we recognize is not just this particular truth, but an underlying universal truth (namely, that lying is wrong). In this way, MPs are supposed to be on par with other sorts of a priori knowledge.

21 Because Richardson and other defenders of this broad conception do not make it clear exactly what ‘responsibly conducted thinking’ is supposed to be, it remains to be seen whether this proposal is in fact a real alternative to the conception described above.

22 See, for example, Railton (1984), Herman (1981), and Baron (1984).
principled, of course, will be subject to the same criticisms I will raise against principled judgments that occur in any situation. Therefore, I will set the period guidance model aside.

Unlike the period guidance model, the internalization model is a descriptive claim about how moral reasoning in fact operates. According to the internalization model, MPs mediate moral judgments and actions unconsciously: that is, MPs are internalized and thus guide moral judgments and actions implicitly. Of course, the details of the internalization process are unclear: neither the philosophical nor psychological literatures have fully articulated just what exactly ‘internalization’ amounts to. That aside, the meaning of ‘implicit’ also remains unclear, implying that there are (at least) two ways of understanding the internalization model:

(9A) MPs function as implicit guides in moral judgments and actions at the personal level – though moral agents do not typically use such principles consciously they could (at least in principle)

(9B) MPs function as implicit guides in moral judgments and actions at the subpersonal level – thus moral agents could not employ them consciously (even in principle).

While these (and other) accounts of indirect moral reasoning may certainly be plausible, they do not change the fact that whatever account of moral reasoning is given, it must maintain the reliance of moral judgments and actions on MPs. After all, no matter what the details, MPs must play a guiding role in the formation of moral judgments and actions in order for that moral judgments and action to satisfy MG.

(II.1.b) Is satisfying MG sufficient for being an AMJA?
Concrete Principles

Concrete MPs identify invariant grounding-to-moral feature(s) relations that obligate moral agents to perform (or refrain from performing) specific actions. Consider two examples: ‘lying is wrong’ and ‘knowingly ending human life is wrong’ (better known in imperative form as ‘do not lie’ and ‘thou shalt not kill’). One of the values of such principles is that they are simple and exceptionless. They decompose the moral domain into basic, easily identified grounding-to-moral feature(s) relations that are recognizable even by those who lack much experience in the moral domain. As useful as this might be for early moral instruction, however, there are (at least) two serious (and well noted) problems with concrete MPs: the problem of rigidity and the problem of insensitivity. The problem of rigidity can be expressed as follows:

(10) Because of their inability to adapt to context, concrete MPs greatly restrict moral agents’ capacity to respond to the situation-specific characteristics of particular circumstances.

The problem of insensitivity is the following problem:

(11) Because concrete MPs are structurally insensitive to the complexities of particular moral situations, they are unable to track the details which, in the end, play a role in determining the appropriateness of moral agents’ moral judgments and actions.

These problems are two sides of the same coin (though the former has more to do with MG and the latter with NA). These problems, while not particularly troublesome for moral agents who are merely “learning the ropes” of moral interaction (for it is expected
that their moral judgments and actions will fall short of being fully appropriate),
nonetheless become unacceptable at the level of mature moral agency.

To illustrate these problems, consider the concrete MP ‘lying is wrong’. While it
certainly seems uncontroversial to hold that lying is often morally reprehensible – and
thus, is often an action moral agents ought to refrain from – it is clearly not only possible,
but likely, that situations will arise in which lying is morally acceptable, if not
praiseworthy. To cite a by-now-classic example, consider a MMA (Hilda) who hides her
Jewish neighbors in her basement during the Nazi occupation of Poland. What should
Hilda do when a group of Nazi soldiers come to her door and inquire as to the
whereabouts of her neighbors? According to the concrete MP in question, the answer is
simple: she must not lie. Yet, it strikes us\textsuperscript{23} that the AMJA would be for Hilda to lie
(indeed, through her teeth, should she need to). In effect, it seems that the rigidity of the
crude MP in this case would lead Hilda astray, guiding her to what we take to be an
inappropriate moral judgment and action.

Similar exceptions arise for all concrete MPs: we can, for example, think of
situations in which it would be not be appropriate to keep a promise, in which it would
not be appropriate to be considerate, in which it would not be appropriate to save a
human life. This is a serious problem when such MPs are supposed to be invariant. So,
how might a concrete MP principle-based theory respond? There are two possibilities.
The first would be to simply hold the principled line: the moral judgments and actions

\textsuperscript{23} Or, at least, most of us. And I take the fact that it strikes us as such as evidence against the concrete MP
in question. While some remain skeptical of the reliability of intuition, I find Bealer’s (2004) point about
the value of concrete case intuition compelling: “The...solution [to skepticism about a priori judgments] is
to defer to concrete-case intuitions. ...We witness success at this in an impressive list of cases that we now
all take for granted. For example, the perceptual-relativity refutation of phenomenalism; the Spartan-
pretender refutation of logical behaviorism; the defective-instrument refutation of instrumentalism; the
unrepeatable-events refutation of the inductivist theory of justification; and on and on.” (14-15)
that our concrete MPs lead to are always appropriate, appearances notwithstanding. For those willing to take this hard line, there is little more to be said. For those of us, however, who consider this an undesirable and untenable position, the obvious next move would be to introduce qualified MPs: a strategy I will consider momentarily. Before doing so, however, I think it is instructive to reflect upon the relationship between concrete MPs and AMJAs a bit more. For, not only do concrete MPs fail to tell MMAs when to conform to their requirements, they also fail to provide any guidance as to how they ought to so conform (a problem that has been largely overlooked).

The principle ‘lying is wrong’, for example, is silent on the question of how one ought not lie – on how one ought to tell the truth. And yet, there are many ways to tell the truth, only a few of which, given any particular situation at hand, would qualify as an AMJA. Take, for example, a MMA (Joe). Joe’s best friend, Bob, has been dating their mutual friend, Katie, very seriously for several years now and is deeply in love with her. One evening, Joe sees Katie out at a restaurant, behaving fairly intimately with another man (whom Joe recognizes as a local married man). Suppose that soon after this event, Bob expresses his desire to marry Katie and his plans to propose to her this coming weekend – a plan prompted, in part, by what Bob has sensed to be a growing distance in their relationship. Bob asks Joe for his thoughts on the matter. Clearly, Joe knows something about Katie that bears importantly upon this issue. But, what (if anything) should he tell Bob? According to our concrete MP, the answer is (again) simple: Joe must tell Bob the truth.

But, that is as far as the concrete MP takes Joe. And thus, it fails him with regards to the most morally important issue: namely, how he should tell Bob. For, not only does
the concrete MP not help him decide how much information he ought to divulge in order to be appropriately “truthful” (for example, is it appropriate for him to divulge the identity of the man with whom Katie was having dinner? Is it appropriate for him to share the specific details of how Katie and the man interacted?); more importantly, it does not tell him how that information must be conveyed in order for his response to qualify as an AMJA. Indeed, as far as this particular concrete MP is concerned, any method or manner of truth telling is as good as the next – it cannot distinguish between them. Yet, consider the possible scenarios:

Bob: Hey, Joe, as you know, I’ve been thinking about marrying Katie for some time now and I’ve decided that I’m going to propose to her this weekend. We are supposed to go to the lake together for a picnic and I plan to propose to her then. I’ve been feeling like Katie and I are growing apart a little of late, she’s been spending more and more time with her friends, and I think that making a commitment like this to one another will help to bring us back together. I’m feeling a little bit nervous about it, though – what do you think?

Joe1: Gee, Bob, I wish I knew what to say. I understand how much you love Katie and if you are feeling a bit of distance developing between the two of you, I suppose it is natural to want to do something to pull things together. Unfortunately, I think there is something you should know before you decide to go forward with the proposal. I recently saw Katie having dinner at a restaurant with someone else, a man. I don’t know the details, I don’t know why she was there with him, but it didn’t strike me as a purely platonic arrangement. Of course I could be just reading things into what I saw, but they did leave together and acted like they
were planning on spending more time together. I think you should talk with Katie and see if you can get clear on what is really going on before you go any further with plans for marriage.

Now, since Joe is a MMA, there is a good chance that his discussion with Bob would go something like the above. That is, he would reveal what he knows about Katie in a morally appropriate fashion. Importantly, though, he could just as easily respond in the following manner and still meet the requirements of the concrete MP:

Joe2: I think you’d be crazy to marry Katie, Bob. She’s been seeing more than just her friends. Just the other night I saw her having dinner with another man and they were acting pretty tight. I mean they were way digging one another – and not too worried about anyone noticing, either. They were holding each others hands, kissing, and mooning over each other pretty heavy. And I recognized the guy she was with. He’s some local married fellow, the guy who owns that little bookstore over on 7th street – he has kids too! Seems like he might be up to some extra-marital hanky-panky. I overheard Katie telling him what room number she was in – seems like she’d rented a motel room or something. I think she’s got you looking like the laughing stock, buddy – you better wise up!

This response, however crude and insensitive, cannot be ruled out as a legitimate response by our concrete MP: if anything, it tells more of the truth – revealing both specific details and Joe’s beliefs about the situation – than Joe1 does. But, it seems that we would reproach Joe for such a response to Bob’s question. As such, it seems questionable to hold that concrete MPs can adequately guide Joe to an AMJA.

Conformity to concrete MPs isn’t enough for Joe to form an AMJA because just as it
seems clear that lying isn’t always (and equally) wrong, it seems clear that the many different ways of telling the truth aren’t always (and equally) right: the appropriateness of his moral judgments and action are determined in large part by both context and the manner in which they are formed.

One could argue, of course, that this problem can be handled by introducing other concrete MPs. We can tell Joe, for instance, that as a MMA he not only has to ‘tell the truth’, but he has to ‘be compassionate towards others’ as well. This pluralist move would not only allow us to understand Joe1 as an AMJA because it complied with both principles, but it would also allow us to account for why Joe1 response was superior to other potential truthful responses (such as Joe2) that would not have qualified as AMJAs: while they might have complied with the ‘lying is wrong’ principle, they would have failed to comply with the ‘being compassionate towards others is right’ principle.

Unfortunately, there are two problems with this response. To see the first, consider Hilda and the Nazis. Can we likewise understand her situation by introducing another concrete MP such as ‘never knowingly endanger human life”? This would allow us to say that Hilda’s lying to the Nazi soldiers was appropriate because it successfully complied with the new concrete MP (knowing full well the Nazi’s intentions for her Jewish neighbors), though it did not comply with the imperative not to lie. But, this runs contrary to the situation with Joe. While Joe’s moral judgment and action was fully appropriate because it complied with two concrete MPs, Hilda’s was fully appropriate even though it only complied with one. So, clearly on the pluralist picture there would be some situations in which MPs worked together in the formation of AMJAs and other situations in which they worked against each other. Yet, how are we to know, in any
given situation, which is the case? We would need to introduce some sort of system of
higher-order regulatory principles that determined both which MPs to apply in particular
situations and how to apply them: i.e., whether to treat them as ‘additive’ (i.e., working
together), ‘subtractive’ (i.e., working against each other), etc.24

More importantly, given that we are working with exceptionless concrete MPs,
the introduction of additional MPs cannot solve the problem. If what Joe says in Joe2 is
the truth then nothing he says can be ruled out as morally appropriate on the basis of
some other principle – it cannot be both morally appropriate on the basis of one principle
and not morally appropriate on the basis of another. Concrete MPs are not supposed to
interact with one another in this way: only prima facie MPs do. Thus, how additional
congrete MPs would address cases like Joes’ and Hilda’s is unclear: it seems we might
have to remain neutral on the moral status of Joe’s unsavory (yet truthful) response to
Bob’s troubles and we’d have to view Hilda’s situation as an irresolvable moral dilemma
in which one concrete MP is obligating her to do $X$ and the other concrete MP is
obligating her to not do $X$.25

Qualified Principles

One way to mitigate the problems with concrete MPs is to turn to qualified MPs.
As mentioned, qualified MPs come in (at least) two forms: 1) prima facie principles or 2)
specified principles.

24 Of course, there is reason to believe that interaction of concrete MPs would not be nearly this
straightforward: for, in addition to having ‘additive’ and ‘subtractive’ functions, they could have
‘multiplicative’ or ‘exponential’ functions, and so forth, as well. For a discussion of this, see Kagan (1988).
25 I do not mean to suggest that principle-based moral theories currently advocate conformity solely to
congrete MPs (presumably because of difficulties I’ve articulated). Nonetheless, I do think concrete MPs
are relied on quite frequently in various folk-applications of moral theory (take, for example, the huge stone
pillar of the Ten Commandments that sits out in front of one of the local churches in my hometown).
Prima facie principles. Recall that the essential difference between concrete MPs and prima facie MPs is that prima facie MPs are contributory principles: though, like concrete MPs, they identify invariable grounding-to-moral feature(s) relations, they do not provide invariant reasons for action. Instead, they provide presumptive reasons: e.g., if lying is wrong, then we always have a presumptive reason to refrain from doing it, but it may nonetheless be the case that other prima facie MPs weigh in as more important in a particular situation, thereby effectively overriding our reason to so refrain. In this way, prima facie MPs are invariant MPs: the wrongness of lying remains even when our reason for refraining from doing it has been overridden. So, in Hilda’s case, the moral accounting process might go something like this: lying is wrong, but not as wrong as knowingly endangering the lives of others; overall, then, Hilda’s lying was morally appropriate.

There are, as I see it, two general problems with this approach. First, it seems mistaken to say that Hilda’s lying in this situation was wrong – in any way. But this is an issue that will have to wait for our discussion of normative authority. More relevant to the issue of moral guidance is the second problem, which is a specifically epistemological problem: because they are merely presumptive, prima facie MPs are by themselves inconclusive. As such, it is unclear how MMAs are supposed to use prima facie MPs to form AMJAs, which are conclusive. The problem, in short, is that it is not clear how MMAs are supposed to employ prima facie MPs without some independent source of information that will enable them to determine when a prima facie MP is overridden by the presence of other prima facie MPs and when it remains in force. Prima facie MPs on their own cannot tell us that, so they appear to be insufficient guides to AMJAs. This
looks like the same problem we ran into with concrete MPs: there needs to be some way to make sense of prima facie MPs presumptive quality, and presumably this would require either organizing prima facie MPs into some sort of invariant hierarchical casuistry and/or introducing higher-order principles that inform us under what circumstances (and by which other prima facie MPs) prima facie MPs can be overridden.\(^\text{26}\)

The successful generation of such a system of principles is unlikely (at best). To see this, let us return to Bob and Joe for just a moment. Presumably, the prima facie story of why the Joe1 response is morally superior to the Joe2 response would go something like this: Joe’s obligation to reveal certain truths (e.g., that Katie had been with a married man who owned a book store of 7\(^{\text{th}}\) street) has been overridden by their inconsistency with his obligation to be compassionate towards his friend. And this is a fine story: but, it leaves us with questions. For example, are we to take it to mean that the obligation to be compassionate always overrides the obligation to be honest? This certainly doesn’t seem right – it could easily be the case that there are times when the obligation to be honest will override the obligation to be compassionate (as well as times when they are consistent with one another). Consider the Bob and Joe case tweaked just a bit – let us imagine that the relationship that Bob has with Katie is deeply problematic. Katie sleeps around on Bob frequently and Bob remains in total denial about it. Perhaps the appropriate thing for Joe to do in this case is precisely that which we did not think was

\(^{26}\)We would also need a set of higher-order principles to tell us under what circumstances there are exceptions to this first set of higher-order principles. For example, the higher-order principle that tells us that in cases like Hilda’s the moral obligation against lying is overridden by the moral obligation to not endanger human life might itself be overridden by the higher-order principle that tells us the moral obligation to not endanger human life is overridden by the moral obligation to not harbor the murderers of children (as Hilda’s neighbors blew up a school house earlier in the week that was educating the children of local Nazi leaders in an attempt to force the Nazis to leave the area) – and so on, potentially \textit{ad infinitum}.\)
appropriate a moment before: that is, he should go into all the “gory details” of what he had witnessed in order to try to shock Bob into acknowledging that his relationship with Katie needed to end. And, of course, quite the opposite might be the case if Joe knew Bob to be a man of erratic temperament, prone to explosive fits of violence: in this situation, revealing too much information could lead to very serious trouble. The worry, then, for the principle-based theory is to figure out how such rich context-sensitivity is to be captured in any principled fashion.

To date, there has been no adequate response to this problem. Indeed, the seeming impossibility of such a task led Ross (1930), as well as others, to forgo principles altogether past a certain point and rely on something like “good judgment” – by which can only be meant judgment that does not employ, involve, or appeal to principles – to make decisions about one’s actual (as opposed to merely presumptive) obligations. Such a response is ultimately an admission of defeat: not only does it mean that AMJAs fail to meet the requirements of MG, but more importantly, it leaves principle-based theories in the conundrum of having to explain why, past a certain point, AMJAs are unprincipled. If principles are capable of doing what principle-based theories think they can do then there should be universal, general, and non-ad-hoc explanations for why and when one prima facie MP trumps another one. If there are, then good judgment is not (ultimately) required and if there are not, then principle-based theories are not (ultimately) principled.

*Specified principles*. An alternative route to achieving qualified principles is to incorporate situational variables up front, as a component of the concrete MP itself, in the form of specified MPs. As mentioned, such specified MPs come in two forms: ‘narrow’ (NSP) and ‘wide’ (WSP). For any $X$ identified as invariably right/wrong by a concrete
MP, a narrowly specified MP specifies the *determinate forms of* \( X \) that are invariably right/wrong. So, ‘lying is wrong’ might become something like ‘lying \( \text{to-cover-up-another-wrong-you-have-committed} \) is wrong’, ‘lying \( \text{with-the-intent-of-harming-another-person} \) is wrong’, and so on. On the other hand, for any \( X \) identified as invariably right/wrong by a concrete MP, a widely specified MP specifies the *conditions under which* \( X \) is invariably right/wrong. So, ‘lying is wrong’ becomes something like ‘lying is wrong, unless it will save someone else’s life or it generates a net-increase in happiness or…’.

Certainly, these MPs will eliminate some of the most obvious exceptions. Yet, in order to be genuinely exceptionless, such principles must capture all of the morally relevant situational details up front – and in a non-arbitrary, non-ad-hoc manner. Imagine considering whether or not, in a particular situation, it would be appropriate to withhold or alter the truth. Imagine all of the potentially relevant considerations that may be present – and how changes in these considerations could easily change what might be appropriate. Imagine all of the possible exceptions (or determinate forms of lying) you would have to wade through in order to determine whether a particular judgment or action would be appropriate. Either this, or you would need (once again) some higher-order system for determining whether certain considerations count as exceptions in this particular instance or not. In other words, the immediate worry of this approach is that genuinely invariant specified principles would be entirely too cumbersome and complex to be functional with respect to MG, especially given our limited cognitive architecture. For every concrete MP we would get either a (potentially infinite) multitude of narrowly specified MPs or one widely specified MP with a (potentially infinite) multitude of
exception clauses. This creates a serious problem as far as MG is concerned: unless we envision MMAs as being god-like rational agents with brains the size of supercomputers, it is unclear how qualified MPs of either sort will be consistent with the requirements of MG.

Perhaps this response seems quick. Perhaps it is the case that the human cognitive architecture can employ complex systems of principles quite effortlessly. We might turn to language comprehension and production, for example – is it not the case that humans somehow implicitly employ complex linguistic principles in order to comprehend and produce language? The problems with using language as an example, however, are several. For one, whether or not language comprehension/production requires or employs principles of any sort is controversial.²⁷

Secondly, it is the comprehension/production of particular sentences that is infinitely complex (because there are infinite possibilities) – this fact does not require that the principles (if there are any) used for such comprehension/production are likewise complex. There may in fact be very few actual “rules” involved in determining whether particular sentences are grammatically or syntactically correct. In other words, though there are infinitely many particular sentences, the underlying structure that determines appropriateness (i.e., grammaticality) may be fairly simple – and it never changes. As such, the appropriate comparison between language and ethics, in this instance, is between the linguistic “principles” that go into the production/comprehension of specific grammatically correct sentences and the MPs that go into the production or comprehension of specific AMJAs. And the whole point of specified MPs, unlike

²⁷ Indeed, there are some who argue that language production/comprehension does not even require knowledge (much less principles) – see, e.g., Dean Petit (2002).
linguistic principles, is that the “rules” one should employ to determine the appropriateness of a moral judgment or action will shift from situation to situation. In other words, they are not simple and they do change.

There is also the worry that at least some accounts of specified principles (e.g., Scanlon, 1998) eliminate the possibility of moral conflict. According to Scanlon, there is one (and only one) specified MP to be employed in any given situation. Yet, if there is a specified MP form-fit to handle every situation that arises, then there should be a definitive answer in each case as to what is morally required. Yet, as Dancy (2004) points out, any principle-based account that does not allow for at least some forms of moral conflict is problematic, since in moral life conflicts must certainly be possible. Finally, there is a worry, raised by Montague (1986) and others, about whether specified moral principles are legitimate adaptations of concrete MPs. But this will be addressed later on, under the heading of normative authority (see footnote 35).

Abstract Principles

Perhaps principle-based theories ought to abandon the quest for action-specific principles entirely and focus on providing abstract MPs instead. Abstract MPs do not identify the same sorts of invariant grounding-to-moral feature(s) relations that concrete and qualified MPs do: rather, they identify an abstract grounding-to-moral feature(s) relation (e.g., ‘happiness is good’) that must be instantiated, remaining silent on the question of which specific sorts of actions will instantiate it. In other words, rather than specifically telling moral agents what to do, abstract MPs identify desired moral ends (or ultimate moral goods) and point towards them, so to speak. Examples of abstract MPs (in imperative form) are: ‘Uphold justice’, ‘Maximize utility’, ‘Do only that which is
consistent with everyone’s self-love’, and so on. These MPs identify certain desired moral ends (justice, utility, and everyone’s self-love), and moral judgments and actions are AMJAs iff they promote or achieve these ends.

The problem with abstract MPs is that they are only useful for identifying moral ends. While such an endeavor is highly laudable and an important element of moral inquiry (especially at the philosophical level), nonetheless when it comes to AMJAs its value is limited. It may be helpful early on in moral agents’ development, when general orientation is what is needed in order to aim in the right direction (and thus hit the “moral dartboard”, so to speak). However, MMAs do not need general orientation; they are in the business of AMJAs (their darts must reliably hit the bulls-eye and not merely the dartboard). Abstract MPs are of little to no help in this regard, for although they tell moral agents what end to promote or achieve, they do not tell them what exactly is required in any given situation to do so. Knowing that the desired moral end is to ‘uphold justice’, for example, tells us next to nothing about how to go about doing so. Indeed, this is what makes moral life so incredibly difficult: it is hard to know, in any particular situation, what is truly called for if justice (or everyone’s self-love or what have you) is to be served.

One might be tempted to respond that the whole point of abstract MPs is that they are not supposed to provide this kind of guidance: rather, once oriented, moral agents are supposed to “fill in the blanks”, as it were, on their own. The trouble is that in order for MPs to be sufficient for moral guidance (that is, to meet the requirements of MG), these blanks must be “filled” with action-specific MPs. But it is questionable whether MPs designed to provide action-specific guidance can perform this function, as we saw in our
discussions of concrete and qualified MPs. Of course, the other frequent response is to say that the blanks are filled in (once again) with “good judgment”. But, for reasons already given, I think this is a bit of a “slight-of-hand” response: in part because no one is particularly clear on just what “good judgment” amounts to (or why principle-based theorists shouldn’t be concerned by the fact that is must be, by definition, unprincipled), but also because it is not clear why, if we had the good judgment to know precisely what was morally required of us in a given situation (which is the hard part), we would need abstract MPs to orient us (the easy part) in the first place. But, this is an issue that we will take up in more detail shortly, so I will leave it at that for now.

The final problem with receiving one’s guidance from abstract MPs is alienation – one of the main objections to principle-based theories discussed in Chapter 1. As has been pointed out by numerous philosophers (see Chapter 1), it is questionable whether guiding oneself according to an abstract MP, such as “maximize utility”, brings about the sort of moral life we envision for a MMA. It seems wrong, for example, for us to have to question whether saving one’s child (who is drowning nearby another drowning child and you can only save one) or dedicating one’s time to those activities that bring personal satisfaction are right actions – that is, whether they are actions that maximize utility. Thus, while all AMJA’s may in fact promote or achieve the moral ends specified by our abstract MPs (though this will be challenged later), it seems nonetheless questionable whether we should be guiding our moral judgments and actions with them.

28 E.g., Blum 1991, McDowell 1997, among others. Indeed, considering the amount of work that “good judgment” does in most principle-based theories, it is amazing how little press it actually gets.
Indirect Mediation

Since my discussion of MPs so far has assumed the direct mediation model of MG, it is possible that adherents of principle-based theories might object that I’m being unfair. Many theorists have proposed indirect mediation models instead: as such, unless my arguments can be shown to work against these models of MG as well, they may fail to hold.

Now, as I mentioned, there are (at least) two kinds of indirect mediation models: the periodic-guidance and the internalization models. Since the criticisms of MPs given so far will apply to the periodic-guidance model (which basically holds that we only need appeal to MPs in times of crises or periodic-reflection), I will not consider it here. The internalization model holds that MPs mediate moral judgments and actions unconsciously: that is, MPs are internalized and thus guide MMAs’ judgments and actions implicitly either (1) at the personal level – thus, though moral agents are not typically using such principles consciously, they could (at least in principle) or (2) at the subpersonal level – thus moral agents could not be conscious of them, even in principle.

In regards to the second option, I see two worries. First, we can only really say that MPs are providing adequate guidance for moral judgments and actions at the subpersonal level if what is actually providing the guidance at the subpersonal level is, in fact, MPs. That is, an internalized version of ‘lying is wrong’ guides one’s achievement of AMJAs only if there exists at the subpersonal level some computational (or representational) equivalent of the principle ‘lying is wrong’. Otherwise, what we have is something else (e.g., emotions, automated pattern recognition, etc.) that is guiding moral judgments and actions – not principles. Second, even if this version of the internalization
model is somehow correct, the subpersonal chunking of principles hardly seems to count as the sort of principle-guided behavior that principle-based theories are looking for: MMAs are supposed to be upstanding, law-abiding rational agents who live their lives according to rationally endorsed principles that they employ in a process of responsibly conducted reasoning. In other words, it doesn’t seem appropriate to call this sort of subpersonal processing “principled”: though moral agents may be following MPs in some way, they are not (and cannot be) aware of it.

On the first option, I fail to see how MPs subconsciously employed would be any more successful than MPs consciously employed: the principles themselves haven’t changed, so they would suffer from the same inadequacies. The only MPs that might benefit from this set-up would be the specified MPs because processing speed may be faster at the subconscious level, which in turn might help with the practicality of employing complex qualified MPs. Importantly, though, unless we are willing to posit some sort of innate “moral module” in which these principles are already present (in much the same way some posit an innate language parser), these complex qualified MPs will have to be learned and integrated into our cognitive architecture in such a fashion as to run subconsciously. And this seems like a dubious prospect: the dominant take-home message we get from cognitive science and cognitive psychology is that humans are cognitive misers and that most implicit cognitive reasoning is the product of “fast and frugal” heuristics (Gigerenzer 2000, Gigerenzer & Goldstein 1996, Kahneman, Slovic, & Tversky 1982, Kahneman & Tversky 1979, Stanovich 1999) and cognitive biases. Such heuristics and biases are notorious for their simplicity and efficiency – not their complexity and accuracy.
The empirical literature on expertise has direct bearing on this point. For here the view that novices become experts through the internalization of domain-specific rules (so that experts make decisions on the basis of the same rules as novices, only more quickly) has come under serious criticism. Such criticism points out that if more rapid rule-following behavior is what constitutes expertise, then it is unclear how experts are able to respond appropriately to such a diversity of unfamiliar and novel sets of circumstances – an ability that is one of the hallmarks of expertise. After all, rules are general decision procedures designed to capture general regularities (the more general the better) and novel circumstances are typically exceptions, in one way or another, to these rules (that is why they are novel). As such, the experts’ skill at adapting to novel circumstances cannot be explained merely through a competence with rules. Of course, it is possible that experts have formulated more sophisticated, more complex rules and that this is why they are able to respond so much more adeptly than novices – but this seems unlikely. And, indeed, research in both expertise development and artificial intelligences suggests otherwise (see Dreyfus & Dreyfus 1986). Perhaps this is why the notion that expertise cannot be reduced down to sophisticated rule-following behaviors is widely accepted across different domains.

Consider, as a representative example, the following quote:

The greatness of critics like Bazin in France and Agee in America may have something to do with their using their full range of intelligence and intuition, rather than relying on formulas. Criticism is an art, not a science, and a critic who
follows rules will fail in one of his most important functions: perceiving what is original and important in new work and helping others to see.29

As this quote suggests, expertise might better be understood as the ability to rapidly perceive, identify, understand, and respond to the various patterns of environmental stimuli that are encountered in a given domain: that is, as the development of a trained, intuitive (and largely non-inferential) responsiveness that replaces (in fact, transcends) rule-guided behavior. As I will argue in Chapter 3, this view of expertise has important implications for mature moral agency and the formation of AMJAs – but to say any more would be getting ahead of myself.

The Gap

The take home message of our inquiry thus far is that MPs, whatever their stripe, cannot be all there is to achieving AMJAs. The problem with MPs concerns their nature, not their use. There is a gap between how far MPs can take moral agents and where AMJAs require them to be. Thus, in order to achieve AMJAs (as well as avoid inappropriate – if not repugnant – moral judgments and actions), something more (or other) than MPs is required.

(II.1.b) Is satisfying MG necessary for being an AMJA?

As I mentioned in Chapter 1, the gap between how far MPs can take moral agents and where AMJAs require them to be has not gone unnoticed in the literature. Several sorts of faculties or abilities have been proposed to bridge it: good moral judgment, refined moral sensitivity, and the like (Blum 1991, McDowell 1998; hereafter simply sensitivity). It would appear, then, that the fact that MPs by themselves are not sufficient for AMJAs is no big deal, since, at first glance, principle-based theories can

29 Pauline Kael (1994).
accommodate the need for an additional faculty/ability without a hitch. Indeed, it seems straightforward to say that of course sensitivity plays a role in principle-based theories – who would have thought otherwise?

I think, however, that the introduction of such a faculty (at least of the sort that would allow MMAs to form AMJAs) is actually problematic for principle-based theories. For it makes the role that MPs are supposed to play in AMJAs obsolete. In other words, positing a sensitivity to bridge the gap between MPs and AMJAs misses the point: it does not succeed in bridging the gap so much as in removing it (by discarding the need for MPs in the formation of AMJAs). Indeed, as I will now argue, there seems little reason to believe that such a sensitivity merely supplements MPs: it seems far more likely that it replaces them. In other words, there may be reasons to think that MPs are not necessary for AMJAs, and that such a sensitivity just is that by which MMAs achieve AMJAs.

So, let us look at the issue more closely. When it comes to an account of what the sensitivity required for AMJAs amounts to, it seems that principle-based theories have three options: they can say that it is a sensitivity to non-moral (grounding) features, to moral features, or to both. Now, these last two options clearly make MPs obsolete: if we have direct sensitivity to the moral features of situations themselves, then there is no need for us to employ MPs in order to determine what they are. As such, it seems that principle-based theories must go with the first.

There are several problems with this however – the most important of which is that the sensitivity they posit is no longer really a moral sensitivity. In addition, such an account does not seem to jive with our moral experiences and the way we talk about those experiences. We talk about perceiving and being sensitive to moral features: we
cringe at the callousness of someone’s behavior, we smile and feel gratitude at someone’s kindness. If we were to explain why a little boy cries whenever his father hits his mother by stating “the boy is quite sensitive to his father’s cruelty towards his mother”, no one would bat an eye or wonder what we meant by such a statement. On the other hand, it seems quite strange to say that the boy infers, by reference to MPs, the cruelty of his father’s actions. Indeed, it seems strange to say that generally speaking, we must employ MPs to infer the rightness or wrongness instantiated by particular grounding features – it seems that we often are able to simply perceive that rightness/wrongness immediately and directly.

What is more, a sensitivity to (non-moral) grounding features cannot provide the sort of guidance required by principle-based theories. To see this, let us consider how the proposed (non-moral) sensitivity is supposed to function. There are two gaps that this sensitivity must bridge. The first gap is between the situation and the MPs: the proposed sensitivity is supposed to determine which MPs apply (and which do not). The second gap is between the MPs and the AMJA: the proposed sensitivity is supposed to determine what moral judgments and actions appropriately follow from the application of the MPs, once chosen. Let us look at both of these functions in turn.

With regards to the first gap, consider Hilda’s case once again and how a (non-moral) sensitivity might help to determine which MPs to apply. The proposed sensitivity picks up on grounding features – in this case, lying – and then determines which MP to apply in this case. Yet, such a sensitivity does Hilda little good in her case. If the MPs she is working with are concrete, then her sensitivity to the fact that lying is (potentially) involved will indicate to her that the she must apply the concrete MP “lying is wrong”.
As such, she must clearly judge in this case, as in all cases, that lying to the Nazis would be wrong. But, this seems mistaken.

Suppose that “lying is wrong” is a prima facie MP instead. How will her sensitivity to the grounding feature (lying) help to determine whether or not the principle is overridden? Perhaps she is also aware of some other grounding feature: say the (potential) protecting of human lives. But why should she think that protecting the lives of her neighbors overrides her obligation not to lie? She must have a sensitivity to the (potential) wrongness of each action in order to determine which carries more weight. Of course, maybe she has been given a systematic hierarchy of principles (e.g., lying is wrong, unless protecting human lives is involved, in which case it overrides the lying, etc.): yet even here, in the very creation of such a complex hierarchy of principles, somewhere in the explanation of why we should hold the protection of human life as overriding the obligation to not lie there would have to be reference to moral concepts: for example, that to lie is less wrong than to fail to protect human life and thus when the protection of human life is involved, one’s obligation to refrain from lying is overridden. In other words, somewhere along the way, we must be sensitive to the moral features – we must be able, for example, to weigh the wrongness of different alternatives. And without a sensitivity to the moral features that grounding features instantiate, it would be difficult (if not impossible) to do this. Of course, one might simply mechanically apply one’s principles simply because one has them – but I am assuming that this is not what we take mature moral agency to require.

Let us think about this with respect to another case: a case in which lying is morally neutral. Witnessing my children opening their presents from Santa Claus on
Christmas morning, it doesn’t occur to me to consider my obligation not to lie – even when my four year old asks point blank, “Momma, does Santa Claus really exist?”, to which I answer “yes, of course” and then go on in an elaborate story of his life at the North Pole. In this case, while there is certainly lying present, there is nonetheless no wrongness present. Yet, if our sensitivity is to the lying and not the wrongness, we cannot tell this case apart from any other. Thus, it must be a sensitivity to the wrongness (not the lying) that tells us whether or not some MP need be applied in a particular case. But, of course, insofar as the wrongness of the situation (or, in this case, the lack of wrongness) has already been sensed, any application of an MP (e.g., ‘lying is wrong’) seems superfluous.

Thus, it seems that when it comes to bridging the first gap, the principle-based theorist faces a dilemma. If the sensitivity being proposed is moral (i.e., sensitivity to the presence/absence of moral features), then the job that MPs are supposed to perform – namely, telling us what moral feature(s) are instantiated by the presence of certain grounding feature(s) – is no longer necessary. If, on the other hand, the sensitivity is non-moral (i.e., sensitivity to grounding features and the like), then the gap between the situation and the MPs still remains.

But, what about the second gap: what sort of sensitivity do we need to bridge it? To explore this question, let us return to the Bob, Joe, and Katie case. Let’s assume that Joe has employed the MP “lying is wrong” to determine that he must tell Bob the truth about Katie. Clearly, this is still a long way from an AMJA. So, what must Joe be sensitive to in order to judge and behave appropriately in his exchange with Bob? Ultimately, he must be sensitive to a whole host of morally relevant features present in
the particular situation. He must be sensitive to a wide range of psychological, emotional, and physical details about Bob, Katie, the man she was with – and himself. So, for example, he must be sensitive to Bob’s feelings for Katie, to Katie’s feelings for Bob, to Bob’s hopes for and worry about their relationship, to Katie’s intimate behavior with a married man, to the confusion Bob feels because he doesn’t know what is going on, to the agony Bob will feel if he tells him what is going on, to his own inability to handle emotionally sensitive issues delicately (since the worst thing he can do now is get nervous and just blurt things out), to Bob’s semi-drunken and emotionally unstable state, to Bob’s infamous temper which might cause him to do something rash, to the fact that the man Katie had been with has a wife and kids who could be impacted by the situation, to his own feelings for Katie that might be tainting his interpretation of what he had seen…and so on.

Of course, the principle-based theorist should be happy to grant the above – after all, what I just gave is a laundry list of (mostly non-moral) grounding features. But, importantly, this list isn’t enough. That is, it is not enough to be aware of the presence of such grounding features – one must be aware of the moral features they instantiate (or would potentially instantiate) as well. To see this, consider being aware of the presence of someone’s pain. In some situations, such as when it is being inflicted by someone for no good reason, the presence of pain may lead to a moral evaluation (e.g., “that is wrong”), while in others, such as when it is the result of having a rotten tooth removed by the dentist, it will not. In both situations, we are sensitive to the presence of someone else’s pain – so why does it lead to a moral evaluation in the former case, but not in the latter? At least part of the answer to this question must be an awareness of the moral quality
(e.g., the wrongness) of the pain in the first case, a quality that is not instantiated in the second case.

MMAs’ judgments and actions are guided by their awareness of the moral significance of the particular details present in the situations they encounter: they can sense which grounding features are (potentially) morally significant and which are not. If we consider Joe’s situation, then insofar as he is a MMA his recognition of what exactly to say and do in his encounter with Bob is born of this sensitivity to the particularities of the situation. When faced with Bob’s question, Joe does not need to first apply a MP and then from there figure out what it is best for him to do – indeed, neither lying nor telling the truth per se are options that Joe will consider. Instead, Joe will hone in on the quality of interaction with Bob (an interaction that will include a mixture of both truths and omissions) that is the most responsive to the moral subtleties of the situation.

The conception of the MMA that I am articulating may sound familiar – it has certainly shown up in the literature before. McDowell (1998), Garfield (2000), and Blum (1991) talk about moral sensitivity as a sort of perceptual skill: the ability to see the morally relevant features in a particular situation, to recognize what is required. To be sure, merely stating that MMAs form AMJAs because they see the right thing to do (i.e., they are sensitive to all the relevant features, to all the moral reasons, present in particular situations) is not enough. We need an account of what such a sensitivity would amount to and how it might be achieved. But any attempt at such an account will have to wait until Chapter 3.
II.2 ~ Conformity as normative authority

At this point, principle-based theories may drop MG entirely and contend that conformity to principles consists in normative authority (NA) alone. This amounts to the claim that it doesn’t matter how MMA form AMJAs (or what might be involved in forming them), so long as all AMJAs are ultimately rendered appropriate by MPs. Accordingly, we can say that,

\[(\text{NA}) \quad \text{A given moral judgment and/or action is appropriate iff it follows (according to norms of good reasoning) from a relevant MP(s).}\]

Let us examine whether this position is adequate.

(II.2.a) Is satisfying NA sufficient to be an AMJA?

In order for MPs to adequately track the appropriateness of AMJAs, there must be invariant grounding-to-moral feature(s) relations which they identify. As we have seen, such relations may obtain absolutely (as in concrete, specified, and abstract MPs) or presumptively (as in prima facie MPs), as well as at the action-guiding (concrete, prima facie, and specified MPs) or higher-order (abstract MPs) level.

It seems uncontroversial to say that concrete MPs fail in this regard. Consider ‘lying is wrong’. As we have already seen, it is simply not the case that one should always tell the truth or never lie. The same goes for other concrete MPs. It is not the case that one should always keep one’s promises. It is not the case that killing is always wrong – or even that killing in self-defense is always right. And so on, ad nauseum. Of course,

\[30\text{ Some theorists (e.g., Railton, 1984) have already suggested going this way in response to the impartiality criticisms discussed in Chapter 1.}\]

\[31\text{ So, on the face of it, concrete MPs can only at best be variant MPs, rough “rules-of-thumb” to be employed by those lacking the sensitive judgmental capacities required for AMJAs. The problem is that without invariant relations, there is no guarantee that our MPs will track the truth. While in some cases, it may be that “lying is wrong” turns out to be the case, it others it will not be.}\]
qualified MPs and abstract MPs are harder to dismiss via such straightforward counterexamples. Nonetheless, I think that these MPs face similar problems. So, let us look at each in turn.

*Qualified Principles*

As has already been mentioned, one obvious solution to the problem with concrete MPs is to treat them as *presumptive*, rather than absolute. Though it is always wrong to lie, break one’s promises, or kill someone, this wrongness only gives rise to a presumptive obligation which can be overridden by other considerations (specifically by the presence of other relevant prima facie MPs). So, returning to Hilda’s case, though she has the presumptive obligation to refrain from lying, the deranged intent of the Nazi soldiers and her further presumptive obligation to (say) never knowingly endanger human life effectively overrides her obligation to so refrain. As such, lying to the soldiers, though in some sense wrong (because it violates an obligation that invariantly obtains), is morally appropriate, all things considered.

The biggest difficulty with respect to prima facie MPs is that it doesn’t seem right to say that Hilda’s lying was morally wrong. It seems, rather, that Hilda had no obligation to refrain from lying whatsoever in this case. Indeed, if anything, it seems that in this case she had the (prima facie) moral obligation *to lie*. This is borne out by considering why Hilda’s behavior is morally praiseworthy. We do not merely judge her keeping her neighbors out of harm’s way praiseworthy; we judge her very lying to do so praiseworthy. Prima facie MPs tell us that lying is always (presumptively) wrong, and thus always “counts against” (in a moral way) – that is, the property being an act of lying

---

32 Recall that prima facie MPs tell us, in the words of Hooker (2000, 6), that “some properties, whenever they are instantiated, *always* count morally in favor of an action, and that other properties, whenever instantiated, *always* count morally against.”
always has negative moral valence. Yet, it seems that in Hilda’s case we want to say that her lying counted *for* her action (in a moral way), and the fact that her act was an act of lying in this case had positive moral valence.\footnote{In an unpublished paper “A Theory of Hedged Moral Principles”, Pekka Väyrynen argues that pro tanto (prima facie) principles can admit of *undermining* exceptions in addition to overriding exceptions. Undermining exceptions nullify the otherwise invariant grounding-to-moral feature(s) relation. This account, while perhaps more plausible than the one considered in the text, is still unable to account for the seeming rightness of Hilda’s lying: in my view, it isn’t enough to say that the wrongness of Hilda’s lying was undermined or neutralized – rather, we want to be able to say that it was *reversed*. No account of prima facie principles that I am aware of can account for this.} In short, Hilda’s behavior is praiseworthy *because* (not despite the fact that) it is an act of lying.\footnote{To clarify, this is not to say that in a given situation there cannot be both *reasons for* and *reasons against* a particular action. (Hilda’s case *is not* meant to be a counterexample to the possibility of moral dilemmas or moral regret.) Rather, this case is meant to be a counterexample to the notion that *reasons for* and *reasons against* are invariant: just because the fact that it involves lying might be a reason *against* an act in one situation doesn’t mean that it cannot be a reason *for* an action (or remain entirely neutral) in a different situation. This is the essential insight, I think, of the sort of holism defended by particularists: the very same properties and features that count morally in favor of an action in one situation may count *against* it (or be entirely neutral) in another. See, e.g., Dancy (2004). Note that if holism is correct, we can only make sense of how grounding features work in particular situations if we acknowledge that their contribution to the instantiation of moral features is variant.}  

This case puts the principle-based theorist in a bind. For, to acknowledge the rightness of Hilda’s lying would be to give up on the invariance of prima facie MPs. It cannot both be that “lying is (prima facie) wrong” and that “lying is (prima facie) right”: invariance, even of the prima facie variety, requires that it only be one or the other. Of course, principle-based theorists may argue that what this case really shows is not that they’ve lost their claim to invariance, but rather that either 1) we are simply cutting our principles too rough, and what we need are specified MPs, or 2) the invariance we seek is at a higher level of abstraction, and what we need are abstract MPs. Of course, these responses admit that prima facie MPs fail to satisfy NA: they fail to adequately track the appropriateness of moral judgments and actions. Still, it is worth considering whether specified MPs or abstract MPs can succeed where concrete and prima facie MPs have
failed (or where, at the very least, we are owed a better account of how they succeed). So it is to this that we now turn.

Specified Principles

While specification (of either form) certainly seems to be an attractive option for principle-based theories, it nonetheless seems that the endeavor to generate specified MPs faces a dilemma:

(Horn 1) We can make specified MPs general enough to function as robust principles, but then they will always be subject to unspecified exceptions (which is to say that they will fail to identify invariant grounding-to-moral feature(s) relations).

(Horn 2) We can make specified MPs highly specific so as not to be subject to unspecified exceptions, but then they will no longer be general (which is to say that they will cease to be principles).

Regarding Horn 1, the point is that no matter how much we specify our principles in advance (while remaining non-ad-hoc and generalizable), there is bound to be unspecified exceptions, situations in which the grounding-to-moral feature(s) relation

\[\text{35 Of course, this assumes that specified MPs can be legitimately crafted from concrete MPs. Some philosophers (Mitchell 1963, Montague 1986) have argued that specified MPs are misconceived: they perform no better than their more simple brethren at providing us with AMJAs. To see this, consider laws of nature. Laws of nature are true only if understood as containing } \text{ceteris paribus conditions. So understood, these laws do not describe and are not true of complex natural phenomena (since } \text{ceteris paribus conditions rarely – if ever – actually obtain). They can serve an explanatory function only insofar as they are understood without their } \text{ceteris paribus conditions – and under these circumstances they are, strictly speaking, } \text{false. More complex laws (modified – as specified MPs are – to take into account complex phenomena) are likewise true only if understood } \text{ceteris paribus – and thus, likewise fail to explain complex natural phenomena. And they too can only serve an explanatory function insofar as they are not understood } \text{ceteris paribus – at which time, strictly speaking, they are } \text{false. In other words, there is no way to combine the more general laws of science into super (i.e., more complex) laws which are both true } \text{and capable of explaining complex phenomena. As such, there is nothing gained in the creation of more complex laws over simpler ones (as, for example, attempting to revise the law of gravitation in order to take into consideration those situations containing charged bodies; see Cartwright 1980). Analogously, there is nothing gained in the creation of specified MPs – they are neither } \text{truer nor more explanatory of the actual complex moral world.}\]
identified do not hold. Of course, this problem occurs only if we take seriously the
intuition that the kinds of moral features that will be instantiated in a given situation
depends (to a large degree, if not entirely) upon the particular set of grounding features
that are present in that situation.\footnote{This intuition is essentially the one lying behind the holism thesis defended by particularists (see Footnote 34). I am not going to discuss or defend holism in detail – that has been done quite nicely (and quite extensively) elsewhere (see, e.g., Hooker & Little 2000; Dancy 2004).} This intuition leads us to Horn 2: the only way to make
MPs exceptionless (that is, the only way to identify a genuinely invariant moral-to-
grounding feature(s) relation) is to make them situation-specific. The problem with this,
of course, is that MPs must be non-ad-hoc and must apply to more than one situation in
order to be considered genuine principles. So, we are caught in a tension between
generating specified MPs that are exceptionless and, at the same time, non-ad-hoc and
generalizable.

Now, to defend this last claim—namely, that there is reason to think that making
specified MPs exceptionless requires that each specified MP be tailored to only one
situation—it is necessary to show that each moral situation is more or less unique with
respect to the combination of grounding features present. This looks to be supported by
the following claim:

\begin{equation}
(12) \quad \text{The likelihood of any two situations (much less enough situations for robust generalization to be plausible) being sufficiently alike in their grounding features to instantiate the exact same moral feature(s) (in the same way, for the same reasons) is vanishingly small.}
\end{equation}

In other words, while lots of situations may involve lying, this is never the only
grounding feature present. A potentially infinite number of other features may be present
– all of which can influence the moral feature(s), if any, that will be instantiated. To the
extent that such grounding-to-moral feature(s) relations are situation specific, the
specified MPs designed to capture them will be as well.

In other words, the argument for (12) goes as follows:

(P1) In any given situation, what moral features are instantiated depends
upon the potentially relevant grounding features that are present
(intuition discussed above).

(P2) The number of potentially relevant features that could be present in
any given situation is very large (if not infinite).

Therefore,

(C) The likelihood of any two situations (much less enough situations for
robust generalization to be plausible) being sufficiently alike in their
features to instantiate the exact same moral features (in the same way,
for the same reasons) is vanishingly small.

Remember that the reason why principle-based theories turn to specified MPs is
because of the intuition characterized in (P1).\textsuperscript{37} If (P2) is also true, it means that even
specified MPs are doomed to failure because it pushes them onto Horn 2: in order to
capture genuinely invariant grounding-to-moral feature(s) relations, specified MPs will
have to be so specific as to no longer be non-ad-hoc and generalizable. Thus, the weight
of the argument for (12) rests on (P2).

\textsuperscript{37} see Scanlon (1998). Another objection to Scanlon’s specified principle approach (raised by Dancy 2004)
is that such MPs must package every morally-relevant feature into the structure of the MP – however, many
of these features are not themselves moral reasons (rather they are defeaters, enhancers, attenuators and the
like of the reasons that are present). As such, the MPs contain information that MPs should not contain –
while the information is relevant to the moral status of the particular situation at hand, it is not relevant to
the job MPs are supposed to perform.
What reasons do we have for believing (P2)? Consider for a moment all of the different sorts of grounding features that carry at least potential moral relevance in a given situation: people’s past and present intentions, their past and present beliefs, desires, goals, access to information, psychological and physical conditions, educational training, relations to others, treatment of others, treatment by others, positions in society, and so on – and this is true for every individual involved in the situation, both those who are acting and those affected by the action(s). Also relevant are features such as past and present physical living and economic conditions, political/legal structures, cultural/religious/familial norms, available technology, available finances, general level of safety, general level of welfare, general level of education, general level of excess, and so on. The potential consequences – given all of these other factors – are potentially relevant, as are the potential implications for moral character that each option represents. And this only scratches the surface: in any single situation the possible combinations of all of the potentially relevant grounding features are astronomical: as Little (2000) has pointed out, even the color of one’s shoelaces could be morally relevant in the right sort of circumstances. And each situation-specific combination of potentially relevant grounding features instantiates its own moral feature(s).\(^\text{38}\)

Consider again the case of Joe, Bob, and Katie. If Joe’s moral judgments and actions in response to Bob’s request for advice about Katie are appropriate – that is, if they are fully sensitive to the (relevant) moral features present in the particular situation – then they will change as the situation (including the people within the situation)

\(^{38}\) Of course, in the face of such complexity, it is not only the viability of principle-based theories that comes into question, but the viability of any theory – indeed, the very possibility of moral knowledge comes into question. Dancy (2004) does a nice job of addressing the worry of skepticism in Ch. 8 of his book (the details of which I will not go into here).
changes. These changes could be substantial or minute: perhaps Joe is exhausted from illness or has had a little bit too much to drink; perhaps he becomes aware that has feelings for Katie (or Bob) and would be pleased if their relationship ended through no fault of his own; perhaps Bob is under a lot of stress at work; perhaps Katie has already consulted Joe in private about not telling Bob about what he saw because she wants to be the one to do it; perhaps the married man she had dinner with is psychologically unstable and could snap if Bob confronts him; perhaps they live in a culture that will ostracize Katie for her infidelity to Bob or that castrates married men for being unfaithful to their wives; perhaps Katie has already decided not to see the man again; perhaps Bob has a serious heart condition; and so on. In the end, the number of potentially relevant grounding features that could be present is tremendous and any number of alterations will affect what the AMJAs would be. Consequently, the moral status of Joe’s moral judgments and actions depend (indeed, must depend) on a whole host of complex grounding (i.e., personal, psychological, physical, cultural, etc.) and moral features. If this is correct, then it seems that (P2) must hold. Given this, accepting the truth of (12) seems reasonable.

An advocate of specified MPs might protest that there are at least some grounding-to-moral feature(s) relations that, if adequately specified, will be invariant while at the same time being at least modestly generalizable. Take, for example, the narrowly specified MP ‘lying-done-for-the-sole-purpose-of-causing-others-unjustified-harm is wrong’ or ‘killing-that-involves-the-agonizing-death-of-innocent-people-for-the-sole-purpose-of-pleasure is wrong’. These principles are of the form: ‘if lying is accompanied

---

39 This is assuming, of course, that the changes that occur are (at least potentially) morally relevant. There are obvious reasons to make this assumption.
by the additional grounding features of *being done with the sole intent of causing harm* and *being an instance of unjustified harm*, then it is always wrong’; and ‘if killing is accompanied by the additional grounding features of *being the cause of agonizing pain in innocent people* and *being done with the sole pursuit of pleasure*, then it is always wrong.’

At first glance, such narrowly specified MPs look quite formidable indeed. Certainly, we must strain our imaginations to find convincing counterexamples (though this might simply be an indicator of the limits of our imagination and *not* of conceptual space). Let us assume (for the sake of argument) that such narrowly specified MPs succeed in identifying invariant relations, and consider whether this vindicates principle-based theories.

My central objection to such narrowly specified MPs is that they artificially build their invariance into them. That is, they do not pick out grounding-to-moral feature(s) relations, but rather moral-to-moral feature(s) relations. What we require of MPs is that they identify *grounding* features (e.g., lying, killing, hitting a child, promise-breaking, etc.) that instantiate *moral* features (e.g., rightness, wrongness, cruelty, maliciousness, etc.). Saying that unjustified harm is wrong is bordering on the tautologous. Of course unjustified harm is wrong: it is *unjustified*. And while it may always be wrong to do that which is wrong, cruel, malicious, it is nonetheless still an open question what specific grounding features, in specific situations, will count as wrong, cruel, or malicious. Thus,

---

40 Compare Hooker’s (2000 5) example that “one should always refuse to perform any act that would eliminate forever all consciousness in the universe.” Even here, Hooker admits that if eliminating forever all consciousness in the universe was the only way to prevent an eternity of universal misery (which means you would be doing it out of love or some other morally admirable intentional state), it would no longer be wrong.
telling us we must never commit unjustified and malicious acts misses the point – we are interested in which acts (under which circumstances) those might be.

To put the point another way, the invariance of narrowly specified MPs such as the ones listed above is generated by their use of “thick” moral concepts (e.g., *unjustified harm*, *agonizing pain*, *innocent people*, etc.). Such concepts are “morally loaded”: i.e., they have a certain positive or negative moral valence (e.g., a ‘rightness’ or ‘wrongness’) built into them. As such, these principles are analytic: the predicate (wrongness) is contained within the subject (unjustified harm). Essentially, what they are saying is nothing more than ‘doing that which has wrongness as a part of what it is to be that concept is (always) wrong’.

Consider the principle ‘lying is wrong’. This principle could easily be rephrased in the following manner: ‘lying is cruel’, ‘lying is unjustified’, ‘lying is selfish’, and so on. The important point is that these moral terms (e.g., wrong, cruel, unjustified, selfish) are all being predicated *of lying* – lying is the thing we are being told we ought not do for moral reasons (i.e., that it is wrong, cruel, selfish, etc.). Cruelty in ‘lying is cruel’ functions as a predicate that expresses a wrong-making property that is possessed by the subject. It seems inappropriate, therefore, for moral terms to be used as the subject (and not the predicate) of the MP. To predicate wrongness of cruelty is redundant.

Apart from this (I think serious) objection, we ought to worry about the adequacy of a principle-based theory that relies on MPs such as these. For such a moral theory would leave most of ethical life *unprincipled*. Most of our AMJAs would not be covered by the narrowly specified MPs identified above, a fact that is important for the following reason. Roughly, these MPs appear to define what we might call the “outer boundaries”
of the moral domain, in the sense that they seem to delimit those areas that most moral agents, mature or otherwise, would never even think to tread. As such, they are silent with respect to the vast terrain that most of us traverse in our daily moral engagements.

Of course, even if principle-based theories were to admit defeat with respect to fully invariant *general* propositions, they might still argue that fully invariant descriptions of grounding-to-moral feature(s) relations are possible – even if they are situation-specific. While such descriptions are not MPs (since they do not satisfy the non-ad-hoc and generalizability conditions give at the outset), they would nonetheless be propositions that identify invariant grounding-to-moral feature(s) relations. For example, it is at least conceivable that we could come up with a complete description of, say, the situation to which Joe responded with an AMJA, thereby capturing all of the grounding-to-moral feature(s) relations present in that situation. As such, we could then rightfully say that, given the particular grounding features present, Joe’s moral judgments and actions were indeed appropriate.

If such situation-specific propositions were indeed possible, they would be largely worthless to us. For one, as already mentioned, they wouldn’t be MPs. They wouldn’t be non-ad-hoc and generalizable. Nor would they be *prescriptive*. For clearly it could not be said that it was *because of* their compliance to this proposition that Joe’s moral judgments and actions were appropriate. For a MP to render an AMJA appropriate, it would have to have been identified (or at least identifiable) *independently from* any particular moral judgment/action as the moral standard to which such moral judgments/actions must conform. That is, the appropriateness of moral judgments and actions is supposed to be
measured against the normative power of an independently articulated, theoretically grounded MP that can be specified in advance.

Indeed, the normative power of these propositions, if there be any, would be entirely dependent upon the appropriateness of the moral judgments and actions which they describe. Thus, such propositions could not be prescriptive – only descriptive.

Abstract Principles

One reason why NA cannot be satisfied at the concrete level may be that the invariant relations that MPs are supposed to track are abstract, in the sense given to abstract MPs at the beginning of this chapter. In other words, perhaps the goal of MPs is to identify those moral feature(s) that must be invariably instantiated while remaining silent with respect to which grounding feature(s) will, in any given situation, actually do the instantiating.

Consider the principle ‘maximize utility’. According to this abstract MP, what determines the appropriateness of any given moral judgment or action is whether or not it successfully maximizes utility.\footnote{Alternatively, rule-utilitarians would hold that a given moral judgment or action is appropriate just in case it follows rules that are designed to do so.} In other words, a moral judgment or action is appropriate just in case it succeeds in achieving or promoting the maximization of utility.

But it is not entirely clear that achieving or promoting the maximization of utility is by itself sufficient to render a given moral judgment or action appropriate. Consider, for example, the (by now classic) cases in which conformity to the principle of utility seems morally suspect: e.g., harvesting the organs of one healthy person to save the lives of ten socially important people who are in need of organ transplants, or enslaving 5% of the population (who, we can suppose, have been picked by lottery and/or genetically
altered so as to be happy with their station in life) for the betterment of the remaining 95%. Similar sorts of counter-examples face other abstract MPs. The result is a forceful challenge to the idea that mere compliance with abstract MPs always renders moral judgments and actions appropriate: again, certain forms of compliance strike us as morally questionable.

One might respond that there are at least some abstract MPs that do not fall prey to the problem I have just noted, despite my suggestion to the contrary. For example, several other abstract MPs might be suggested (in imperative form): ‘always treat others as ends and never as means’ (Kant, 1956), ‘always do that which is consistent with everyone’s self-love’ (Garrett, unpublished draft), ‘do onto others as you would have them do onto you’ (a.k.a., the “golden rule”), ‘always uphold justice’ (a principle Kohlberg would have liked), and so on. While each of these is certainly a legitimate moral objective (generally speaking), there are at least two reasons to think that we ought to be skeptical that such MPs can ground AMJAs.

First, I think that even if we were able to fashion an abstract MP that was not open to counterexamples, it would nonetheless provide the wrong sort of normative authority. Consider: a young man who gives up his seat to an ailing elderly woman on the bus; a young woman who, though cajoled by her boyfriend to steal untraceable cash from her boss to help fund their vacation, refuses; a little boy who puts up signs around his neighborhood about a lost puppy he has found (in an effort to find the owner even though he would happily keep the puppy for himself). Do we really want to say that the source of

---

42 Alternatively, one might contend that the principle of utility itself is not prone to the objection given above, for all the counterexamples show is that this MP must be reformulated. However, I am not aware of any such successful reformulation, and consider the objection to at the very least place the burden squarely on the shoulders of defenders of this MP.
the moral appropriateness of each of these moral judgments and actions is the same? That is, do we really want to say that the appropriateness of each of these AMJAs is located solely in the fact that they all upheld justice (maximized utility, were consistent with everyone’s self-love, or treated others as ends and not means)? I think that the answer to this question is no. The appropriateness of each of these actions lies in the details of the situations themselves: for example, the appropriateness of the young man’s act is located (roughly speaking) in the compassionate way he responded to the needs of the elderly woman he encountered on the bus; the appropriateness of the young woman’s act is located in her refusal to succumb to the temptation and pressure to perform an action she recognized to be wrong; the appropriateness of the little boy’s act is located in his empathic consideration for the puppy’s original owner and his ability to overcome his own selfish desire to keep the puppy for himself.

In short, to hold that the appropriateness of these disparate moral judgments and actions all equally lies in their promoting or achieving a given abstract MP is to ignore many morally significant differences between them. Perhaps it is the case that each of these actions in some way satisfies the end specified by our abstract MP(s): perhaps the young man, the young woman, and the little boy did all succeed in maximizing utility or treating others as ends or behaving consistently with everyone’s self-love. But, nonetheless, it seems strange (if not misguided) to say that their moral judgments and actions were appropriate for that reason (and that reason only) – and, not only this, but that when considering these judgments and actions, we must first determine whether or not they satisfy the MP(s) before we can judge them to be AMJAs.43

---

43 This objection is related, I think, to the alienation objection: it just seems to place our moral attention in the wrong place, alienating us from what matters. Now certainly Mill (1979) stated that, for the most part,
Secondly, it strikes me that there is an important difference between that which is the ultimate end of morality and that upon which moral judgments and actions are rendered appropriate. Recall that AMJAs must successfully respond to a wide variety of grounding-to-moral feature(s) relations. That something is (or would be) cruel or selfish; that something would ease someone’s pain or be considerate – these are the sorts of features that MMAs must be sensitive to. And the appropriateness of AMJAs lies in their responsiveness to such features. On the other hand, that something would maximize utility or be consistent with everyone’s self-love are not the sorts of features MMAs can be sensitive to – if they are features that particular situations can even possess (hence the seemingly intractable problem of “utility calculus”). So, while MMAs sensitivity and responsiveness to the moral qualities of situations may ultimately result in a net promotion or achievement of our abstract MPs, this in no way commits us to the position that such MPs are the sole arbiters of appropriateness.

(II.2.b) Is satisfying NA necessary to be an AMJA?

Is satisfying NA the only way to render moral judgments and actions appropriate? If there is an alternative way to render appropriateness, then no. But, what might such an account look like? The proposal I would like to consider – namely, ‘moral reliabilism’ – goes as follows:

(13) A given moral judgment and/or action is appropriate iff it is formed through a reliable process(es).

we didn’t need to worry about whether our judgments and actions ultimately satisfied the MP; that it was perfectly acceptable for us to do things for other sorts of reasons – because these reasons ultimately mapped onto (more or less) the MP that is the true source of normative authority. And this is all well and good – but, to my knowledge, it remains an untested and undefended assertion.
Recall that AMJAs are moral judgments and actions that are fully sensitive to the moral (reason-giving) features present in particular situations. Accordingly, let us take the relevant class of processes to be those that are (more often than chance) fully sensitive to the (relevant) moral features present in particular situations. This means that a moral judgment or action is rendered appropriate by the fact that it is formed via processes that are (more often than chance) fully sensitive to the (relevant) moral features present in particular situations.

If this is correct, then in order to understand the appropriateness of AMJAs we must identify those processes that are reliably sensitive to the (relevant) moral features present in particular situations. It would follow that MMAs employ such processes in forming moral judgments and actions, and that understanding mature moral agency thus requires understanding these processes and how they develop.

Shafer-Landau (2003) argues that the empirical study of moral exemplars is one way to identify and study such moral processes. However, the difficulty with this approach, as Shafer-Landau acknowledges, is the problem of circularity:

[W]e can know which processes are reliable [only] if we know who is and isn’t a moral exemplar. We can know whether a person is an exemplar only if we know whether her judgments are reliable. But, if we knew that, then reference to the exemplars would be otiose. (297).

I think (as does Shafer-Landau) that there are ways of short-circuiting this circularity in such a way as to render it innocent. Be that as it may, relying solely on the empirical study of moral exemplars may not be sufficient – if for no other reason than

---

44 Indeed, some very interesting empirical work along these lines has been done. See, for example, Walker, et al.’s (1989, 2004) work, along with Colby & Damon (1992) book Some Do Care, as well as Baltes’ (1997) empirical studies of wisdom.
they are hard to find. It is our good fortune, then, that there is other empirical research to which we can turn: in particular, research on the development of expertise.

Understanding the processes by which experts efficiently, accurately, and reliably track relevant features of various domains may provide us with valuable insight into the sorts of processes that might be employed by MMAs in forming AMJAs. I will explore and defend this claim in Chapter 3, where I argue that what has been learned through studies of expertise provides us with at least the beginnings of a viable non-principle-based theoretical conception of mature moral agency.
CHAPTER THREE

~ A Skill-Based Account of Mature Moral Agency ~

“…he might have said, if any man could have got him to talk about it, that like the morning dove, the bittern, the Indian, he had a sixth sense. What he thought of as his sixth sense was in fact only what his five senses agreed on and communicated to his mind, acting together, like an intelligence agency, to sort out, accept or reject, and evaluate the impressions that came to them.”


“People whose actions are generated by adherence to rules are like beginners learning a motor skill.” - Francisco Varela (1999, 29-30)

I. MORAL RELIABILISM

According to principle-based accounts, normative warrant for one’s moral judgments and actions comes from the inferential connection they possess to moral principles (MPs): the moral status of any particular judgment or action is (and must be) ultimately derivable from either one or a set of MPs. Such MPs themselves derive their justification either through their non-inferential self-evidence, derivability from other premises, or through their inferential compatibility with our larger belief system. Given the problems that principle-based theories must confront (as discussed in Chapter 2), the assumption that moral judgments and actions are appropriate only to the extent that they stand in the appropriate inferential relation to MPs seems problematic. Thus, it is important to look elsewhere for an alterative account of appropriate moral judgments and actions (AMJAs).

45 Just an introductory note of warning (and apology) – the purpose of this chapter is to sketch out what I take to be a viable alternative account, many of the details of which are yet to be worked out (and which would require much more space and elaboration than can be provided here to be worked out). As such, there is going to be more gesturing (i.e., “hand waving”) than in Chapter 2.
One such an account (though certainly not the only account) is the reliabilist account (e.g., Shafer-Landau 2003). According to the reliabilist account, AMJAs are those moral judgments and actions that are formed via reliable (moral) processes, such processes being those that lead (more often than not) to true verdictive moral beliefs (translated, for our purposes, into appropriate moral judgments and actions). MMAs are those moral agents who employ such reliable processes – processes that are reliably sensitive to the (relevant) moral features present in particular situations – in order to form moral judgments and actions. Given the critique in Chapter 2, it is not likely that these reliable processes will (normally) require guidance from MPs: though they may be sensitive to universal grounding-to-moral feature(s) relations in the sense of grasping the *necessity* of the moral features present in a given situation, this does not require a sensitivity to the sorts of invariant generalities supposed to be captured in principled form.

So, what might such reliable processes be? Proposals for the sorts of processes I have in mind have come in a number of different forms. Some have attempted to cash out such processes in strictly perceptual terms (e.g., Blum 1991, Fortenbaugh 1964, McDowell 1998, Starkey, unpublished draft). MMAs perceive what is morally required of them – that is, they can *see* what needs to be done. According to McDowell, MMAs have a ‘reliable sensitivity’ which is essentially the ability to perceive the requirements of

---

46 In his book, *Moral Realism* (2003), Shafer-Landau gives an extended defense of the viability of a reliabilist approach to moral epistemology, which includes addressing concerns about the externalist requirement, causal worries, and concerns about the individuation of processes – all concerns faced by reliabilism more generally. Since I have neither the time nor space to go into a defense of reliabilism here, I refer all interested readers to Ch 12 of Shafer-Landau’s book.

47 What I mean is that the moral features that are instantiated by the grounding features present are *necessarily* instantiated and would necessarily be instantiated in any situation in which exactly the same grounding features (and no others) were present (hence, grounding-to-moral feature(s) relations are universal in this very limited sense). The sort of sensitivity I have in mind is thus not the same as Rossian intuitive induction (i.e., intuining the general invariant truths from the particular case).
particular moral situations. Importantly, this entails that MMAs are able to perceive the situation *as it really is*: they accurately perceive the moral features that are present. Along these lines, Fortenbaugh writes,

> To the coward what is not formidable appears formidable, and what is slightly formidable appears exceedingly formidable. For the rash person the appearances are reversed. But to the courageous man things appear *as they really are*... (1964 78, emphasis added)

Others have argued that the reliable processes we seek are cultivated emotions and/or sentiments (Jacobson 2006, Prinz forthcoming) – MMAs are those moral agents that experience the appropriate sorts of emotions (e.g., approval, disgust, indignation) in response to encountered moral features. Still others cash them out in terms of skill and “know-how” (Dreyfus & Dreyfus 1991, Jacobson 2006, Varela 1991; cf. Annas 1995) – as Jacobson writes, “One sees *what* to do in some situation, because one *knows how* to perform an activity skillfully” (391). MMAs know how to skillfully perceive and respond to moral reason-giving features.48

My discussion of mature moral agency will attempt to incorporate insights from all of these accounts, with an emphasis on the moral perception and the skill-based accounts. I will argue, based on what I take to be the core features of our current empirical models of expertise, that mature moral agency requires (at least) the following two reliable processes: what I call trained perception and automated responsiveness (or “know-how”).

48 Julia Annas (1995) likewise conceives of virtue as “a disposition to act for reasons”, though she thinks of this more as a grasping of general principles that are then applied to particular situations through practical reasoning than as a perception of particular moral features. Her recent paper (given at the Pacific APA, 2006), however, seems to be moving away from this a bit – she explicated a phenomenology of the virtuous agent in terms of “flow” (which I will likewise discuss momentarily).
II. EMPIRICAL MODELS OF EXPERTISE

Bruce Weinstein (1993) and others have argued that expertise comes in two distinct forms: *performative* and *epistemic*. The former concerns expertise in or at a given domain while the latter concerns expertise about that domain. In other words, performative experts demonstrate their expertise through their skillful engagement in and with the domain in question: they create masterpieces in art, play Bach masterfully, perform incredible feats of physical and mental skill, and so on. Epistemic experts, on the other hand, are able to dazzle us with their knowledge: they can give us the artists and dates for every great masterpiece ever produced (as well as why they were great); they can tell us in fantastic detail all about the major periods of classical music; they can discuss the mechanics of ski jumps and professional golf swings – though they may themselves be unable to sculpt, play an instrument, or ski.

It is most likely the case that MMAs are both performative and epistemic experts. Nonetheless, I take it both that it is possible for moral agents to be the *latter* sort of expert without being the former (most academic ethicists being a case in point) and that, more importantly, what most interests us about mature moral agents is their performative expertise – i.e., their ability to skillfully (and appropriately) engage with others in the moral domain to form AMJAs. As such, my primary focus will be on this particular form of expertise.

Recently, psychologists and cognitive scientists have begun to explore the processes underlying performative expertise (Baltes 1997, Bloom 1986, Dreyfus & Dreyfus 1986, Ericsson & Charness 1994, Ericsson & Smith 1991, Sternberg &

---

49 Much of the discussion in this section, though since revised and re-worked, was pulled from a previous paper of mine (“Expertise, Skill, and Know-how”) that was co-authored with John Bengson. I am indebted to John for his input.
Grigorenko 2000). Their collective research efforts unite on the fact that whatever the domain – e.g., riding a bike, playing the piano, driving a car, playing chess, martial arts, nursing, and so on – expertise requires a developmental progression from novice to expert, a progression which occurs only through extensive training, instruction, imitation, dedication, and practice.

Dreyfus and Dreyfus (1986) have proposed a five-stage model of this development from novice to expert. For novices in a given domain, tasks to be performed must be decomposed into basic context-free features that are recognizable without the benefit of experience. Novices are given simple rules and/or decision procedures for determining actions on the basis of these features. These rules and decision procedures are insensitive to situational elements, allowing the performer to ignore the situation-specific details of the task (details they are not yet equipped to handle). Through practical experience with actual concrete situations, advanced beginners start to become aware of how these situation-specific elements impinge upon the task at hand and thus begin to acquire various maxims that refer to and incorporate the most frequently encountered situation-specific elements, enabling them to deliberate about what actions to take on the basis of both features and (limited) context. Further experience in the domain brings an expansion of task options and, with it, an explosion of such situation-specific information. Faced with the need to systematize decision-making, competent performers may identify general guidelines or principles with which to organize incoming information, reducing the amount that must be considered during any given task. They engage in the detached assessment of those elements that are salient with respect to the chosen guideline, followed by a rule or maxim-guided action.
Throughout this developmental process, something important begins to occur: performers’ actions begin to be reliably coupled with an emotionally tuned experience of the outcome (i.e., their emotions begin to track the subtle details of their successes and failures), until they begin to experience their decisions as natural, spontaneous “choices of action”. This emotional tagging helps to facilitate learning (aiding with both memory formation and recall), connecting patterns of input with various components of cognitive and behavioral output. For example, these emotionally tagged experiences begin to develop into automated response motor sets (i.e., the beginnings of “know-how”).

*Proficiency* is marked by the early stages of trained perception: that is, proficient performers begin to identify meaningful action-guiding patterns without needing to decompose such patterns into context-free features or other simple components for deliberation. They can comprehend the domain environment without deliberative effort insofar as they directly (i.e., non-inferentially) perceive what is going on. It is not until the level of expertise, however, that such direct perception of meaningful patterns and features leads reliably to spontaneous (and appropriate) action. *Experts* are able to directly comprehend not only what is going on, but also (because their body of know-how has become robust enough to be adaptable to most contingencies) what to do.

Two of the examples discussed by Dreyfus & Dreyfus (1986) are chess and driving. In learning to play chess the novice starts out using a numerical value for each piece to calculate moves according to some basic rule: e.g., ‘always exchange if the total value of pieces captured exceeds the value of pieces lost’. Chess masters, on the other hand, have developed an immediate and non-deliberative “compelling sense” of the best move: they see the game as a dynamic, shifting set of holistic patterns whose meaning
they are able to perceive and respond to rapidly, allowing them to play at the rate of 5-10 seconds/move without significant degradation of performance. Consider also driving, where the novice goes from following a rough calculation for shifting (e.g., 20 miles/hr = 2\textsuperscript{nd} gear, 30 miles/hr = 3\textsuperscript{rd} gear, and so on) to seeing and responding to the changing environment (speeding up, slowing down, shifting, etc.) without thought or need for deliberation. What is notable about each of these cases is that, for experts, perceiving and responding appropriately have become interconnected. There are few deliberative pauses between what one perceives and what one does: rather, it is as if they have become two aspects of the same activity.\textsuperscript{50}

The rich experience possessed by experts provides them with the skills necessary for a flexible and adaptive responsiveness to the domain environment: one that is both more efficient and more accurate than non-experts. Accordingly, experts rarely need to detach in order to analyze, problem-solve, or deliberate: their trained perception and automated responsiveness enable them to respond directly and skillfully. This is not to say that reflection and deliberation are no longer necessary (nor important) – it is merely the case that the need to do so happens less often. There are even moments in which experts must digress back to structured rule/maxim application because their experience has not yet provided them with the skills to handle what they have encountered. The expectation, in such situations, is that their response will most likely be less than ideal.

While other models of expertise may differ from Dreyfus and Dreyfus’ model in the details, they all nonetheless appear to agree that expertise requires the development of (at least) the following two processes:

\textsuperscript{50} Importantly, this interconnection of perception and action comes without loss of appropriateness: perhaps the novice could also act spontaneously, but such action would frequently result in impaired performance; whereas for the expert, the opposite is typically the case.
Trained perception: the process by which (complex) patterns of stimuli,
composing or instantiating meaningful features, are directly
(i.e., non-inferentially) perceived.

Automated responsiveness (or “know-how”): the process by which engrained
(automated) motor sets or programs are activated and
adapted on-line to appropriately “fit” the situational
features that originally activated them.\(^{51}\)

It is important not to confuse automated responsiveness with mere engrained,
habitual responses: it is much more powerful than that.\(^{52}\) Automated responsiveness
allows experts to respond very rapidly (and often without need for deliberation) not only
to situations that they have encountered in the past, but to novel and unexpected
situations as well. They can adapt to new contingencies (within a certain range) both
efficiently and effectively, maintaining accuracy to an astonishing degree.

Empirical research has demonstrated the importance of both trained perception
and automated responsiveness to expertise. Concerning trained perception, research
suggests that one of the key differences between experts and non-experts is that only the
former fully understand both what to attend to and what that to which they are attending
means. As an example, researchers using eye tracking technology have found that experts
in medicine, art, chess, and cartography are much more efficient and selective in their eye
movements than non-experts (presumably because they know to what they should attend)

\(^{51}\) This process typically occurs without any need for deliberation on the part of the MMA (translated into
cognitive science terms, there is typically no executive control intervention from the pre-frontal cortex, but
rather involves “older” centers of the brain, such as the basal ganglia).

\(^{52}\) That is, automaticity is not mechanical or rote, but intelligent. Ryle (1949) distinguished between “blind
habit” and “intelligent action”: only the latter describes expert practice. In an attempt to clarify the distinct
processes by which habit and intelligent action are formed, Ryle made a corresponding distinction between
drill (conditioning by brute repetition) and training: “we build up habits by drill, but we build up intelligent
capacities by training” (1949, 42).
and what they conclude on the basis of their perception is much more accurate. So, “subjects who had extensive art training tended to concentrate on finding thematic patterns among compositional elements…[and] ‘untrained viewers failed to recognize the perceptual organizing functions of symmetry’” (Solso 1997, 147). Likewise, “The patterns of eye movements produced by expert radiologists as they make a diagnosis from an X ray is far more parsimonious than that of novices” (143, see also Krupinski, et al. 1993, Nodine et al. 1993, Yarbus 1969).

Expert bird-identifiers (“birders”) have been shown to “recogniz[e] not just individual features but clusters of features” that enable them to automatically identify the species of particular birds. “[A] quick glance, hearing the song, may be all that is needed for immediate identification.” On the other hand, novices must “focus their attention deliberately and laboriously on noticing the identifying characteristics…and mapping them onto a particular species” (Samuels & Flor 1997, 2). Likewise, expert air traffic controllers do not infer the location of planes from “blips” on a screen. While non-experts are hampered by the need to extrapolate from the “blips” to the actual location of the planes they represent, experts see the “blips” as planes – and thus see their location and speed, enabling them to respond rapidly and appropriately to the planes’ positions and trajectories (Dreyfus & Dreyfus 1986). And, finally, chess experts do not infer that such-and-such a move would be best from the particular locations of the individual chess pieces. Instead, individual pieces are seen as holistically ordered arrangements, or meaningful patterns, often referred to as “lines”, enabling experts to respond quickly with the best next move and to accurately predict future positions (Stillings, et al. 1995). Chess masters are also able to reconstruct the configuration of the pieces on the board even after
only seeing the pattern for a few seconds: however, when presented with meaningless configurations of chess pieces (i.e., configurations that wouldn’t occur in an actual game of chess), their ability to perceive and respond to the pieces as a holistic pattern breaks down, leaving experts no better at reconstructing the configurations than novices (Ericsson & Smith 1991).

Of developmental relevance is also the research on infant perception—particularly their perception of human faces. It has been found that one month old infants scan people’s faces sporadically, attending mainly to the outer edges, but by the second month their eye movements begin to fixate onto the central features of the face (eyes, nose, and mouth), forming an efficient triangular pattern of eye movement that focuses on the area of the face most crucial for person recognition, emotion detection, and theory of mind development. Autistics (both children and adults) fail to employ such triangular patterns, continuing to engage in sporadic facial scanning, perhaps explaining why they lack emotion detection and basic theory of mind skills (Pelphrey et al 2002, Maurer & Salapatek 1976). We know also that skilled readers do not need to read one letter or word at a time, as lower-level readers do. Strings of letters are seen as words and strings of words are seen as meaningful “phrases, sentences, and ideas” (Bloom 1986, Samuels 1976, Schiffrin & Schneider 1977, LaBerge & Samuels 1974, Willows 1974). Indeed this is the secret behind speed reading, which trains readers to see “blocks” of text as one entity, enabling them to take in even entire paragraphs at a time.

Though the above research has as its emphasis the development of trained perception, it is worth emphasizing the importance of automaticity (in the form of
automated responsiveness) to expertise as well. As the research suggests, experts exhibit greater automaticity in the sense that they cease to need to depend upon inferential rules or decision-procedures to determine action – they can respond automatically. Such automaticity is achieved over time through extensive instruction, training, and practice. Experts “over-learn”, so to speak, the skills required in their domain until they can be performed with little effort or deliberation (hence the expression, “I’ve done this so many times, I could do it in my sleep”). Chess experts, for example, are able to play competitively against other experts in games of “lightning” chess (2-5 minute games), even when asked to add numbers that are given to them at the rate of approximately one number per second in their heads (Dreyfus & Dreyfus 1986).

For experts, perception and action become linked in a way that allows for spontaneous, non-deliberative (and yet appropriate) responsiveness: one that adapts to appropriately “fit” the particular details of the domain environment, much as we adapt our step to appropriately fit the upcoming changes in the physical landscape, even when we are not deliberately focusing on our movement. Indeed, deliberate focus often screws things up: when focusing on our step, we often become clumsy and hesitant, as perhaps best exemplified by the awkward movements of a self-conscious teenager aware that a group of girls is watching him. In line with this point, empirical research has demonstrated that experts typically perform better on tasks when they are performed automatically rather than deliberately, even in far from ideal circumstances. For example, in a series of experiments conducted by Beilock, et al. (2004), not only were expert golfers more successful than novices at hitting a predefined target under a number of

---

different conditions, but their performance actually improved in those conditions that were apparently antagonistic. Specifically, expert golfers performed significantly better when under a strict time constraint (3 seconds) or when asked to simultaneously perform tone-monitoring tasks than when they were encouraged to “single-mindedly” concentrate on what they were doing with neither a time constraint nor a distractor-task involved. Novices’ performance, on the other hand, both significantly improved when given ample time to concentrate and significantly worsened when placed under the time constraint or given the distracter-task.

Perhaps this is because deliberation takes time, thereby slowing things down. As Benjamin Bloom (1986, 74) observes, “automaticity in reading, speech, driving, piano playing, skating, or dancing is far more precise and accurate than the same processes would be if they were done with full conscious control”, a sentiment which is in line with reports from experts in many different domains – perhaps most famously from athletes that, when they are at the top of their game (“in the zone”, as they say), report acting without deliberative effort, often with no sense of being in control of their actions.

As one expert martial artist reports,

There is no choosing. It happens unconsciously, automatically, naturally. There can be no thought, because if there is thought, there is a time of thought and that means a flaw…If you take the time to think ‘I must use this or that technique’ you will be struck while you are thinking (quoted in Dreyfus & Dreyfus 1986, 35).

In summary, empirical research shows that becoming an expert requires the development of both trained perception and automated responsiveness with respect to

54 Deliberative and automatic activities are also processed differently in the brain (Kandel, Schwartz, & Jessel, 2000).
particular domains. These two processes work together, becoming “more and more refined *dispositions* to respond to the solicitations of more and more refined perceptions of the current situation.” (Dreyfus, unpublished). This gives rise to what Sam Menard (in Vardis Fisher’s *Mountain Man* – quote above) called his “sixth sense”, which is nothing more (or less) than a robust disposition to accurately perceive and appropriately respond to the relevant features of a given situation: to be *fully sensitive* to the reasons that are present in the case.

**III. DEVELOPING A MORAL SENSE**

So, it seems that there is evidence enough to give credence to the idea that (performative) expertise is essentially a robust disposition composed of the reliable processes of trained perception and automated responsiveness. But, why should we believe that this has any real implications for ethics – why treat mature moral agency as a form of expertise? The main reason to do so is because trained perception and automated responsiveness are the sort of reliable processes we need for an account of mature moral agency: the sort of reliable processes that will lead to AMJAs. So, let’s consider each in turn.

**III.1 ~ Moral perception and AMJAs**

It is clear that refined moral perception (the moral equivalent of trained perception) is at the very least *necessary* for AMJAs. As we discussed in Chapter 2, possessing a host of MPs and good deliberative skills will do a moral agent no good unless she (accurately) perceives the moral situations that confront her as *moral* situations. What is more, she must be able to (accurately) perceive the moral features that

---

55 Of course, this is not to say that expertise does not also require a substantial body of declarative knowledge as well. It almost certainly does.
are present in these situations in order to determine whether one principle versus another should be applied. But, there are some who argue that moral perception is sufficient for AMJAs as well. McDowell, for one, explicitly argues against the need for principled generalizations, concluding that “[o]ccasion by occasion, one knows what to do, if one does, not by applying universal principles but by being a certain kind of person: one who sees situations in a certain distinctive way.” (1997 162, emphasis added)

There is an important worry for this view, however, involving exactly how we are to understand the use of perceptual terms in the moral context. Jacobson (2006), for one, takes McDowell’s use of perceptual terms to be purely metaphorical. He argues that, strictly speaking, we do not see the demands of moral situations, we “see” them. And we do so by means of feeling them (that is, we can “make sense of the idea that a way of seeing a situation might involve the ability to feel some way about it” [394]): thus, the skill model we are after, according to Jacobson, involves a training of the moral sentiments.

I don’t think this view is quite right. While I certainly agree with Jacobson that the emotions (or sentiments) play a crucial role in AMJAs (remember that emotions serve as crucial experience markers in the development of expertise generally), nonetheless I take it that one does not feel moral features – one perceives them. What one feels is instead a reaction to that perception. Imagine, as an illustration, witnessing a group of young boys torturing a cat: it makes sense to say that one immediately perceives the

56 As Blum (1991) writes, “An agent may reason well in moral situations, uphold the strictest of standards of impartiality for testing her maxims and MPs, and be adept at deliberation. Yet unless she perceives moral situations as moral situations, and unless she perceives their moral character accurately, her MPs and skill at deliberation will be for naught and may even lead her astray.” (701)

57 Of course, one immediate objection to this interpretation of the use of perceptual terms in moral evaluation is that it assumes that there really are moral features (or moral properties, if one is more comfortable with this terminology) out there to be perceived. As a moral realist, I’m not particularly bothered by this assumption – though arguing for it falls outside the scope of this paper.
cruelty of their action towards the cat while feeling, in response, extreme disapproval (as well as something like moral indignation towards the boys and intense sympathy for the cat). Of course, I think it is likely that moral perception, insofar as it involves the perception of features which are reasons for certain moral judgments and actions, is necessarily linked (at least in MMAs) to emotion: that is, barring some non-normal conditions such as chronic depression, it necessarily elicits certain emotional (motivational) responses.\(^{58}\) Thus, in a very real sense, MMAs do feel cruelty, insofar as they reliably respond emotionally to its presence. But, to be able to respond emotionally to its presence, they must accurately perceive it first. Given this, I think that moral perception should be treated as distinct from moral sentiment.

Consider, for example, Blum’s (1991) discussion of John and Joan, who are both riding a subway train: it is a fairly full train, they are sitting next to one another, and there is a woman standing nearby holding two full bags of groceries. She is clearly in a bit of distress, the weight of the bags rapidly becoming difficult for her to manage comfortably. John, while generally cognizant of the woman’s presence, does not notice her distress; Joan, on the other hand, perceives the woman’s discomfort clearly. As Blum (I think rightly) argues, the difference in John’s and Joan’s perception of the woman is of moral significance: John fails to perceive a morally relevant feature (the woman’s distress) that is quite salient for Joan. Joan’s perception of the woman’s discomfort seems distinct from the way she might feel about it – though, if Joan is an MMA, her sentiments will be in harmony with her perception. But, this is so because her moral judgments and action will

\(^{58}\) What I’m saying here, while apparently internalist, is not inconsistent with an externalist account of reasons – I simply take it that for MMAs (though perhaps not for all moral agents) that one has an (external) reason to act \textit{just is} an (internal) reason to act.
naturally be in harmony with her perception, her sentiments being inextricably linked with her moral judgments and action.

This point brings to light the importance of moral perception when it comes to the formation of AMJAs. Consider the moral judgments and actions with regards to the woman that John and Joan are likely to form: John perceives no reason to come to the woman’s aide (and thus, will likely fail to), whereas Joan’s perception of the woman’s distress “already provides her with a reason for action” (703). That is, Joan’s perception of the woman’s distress renders her, to use McDowell’s turn of phrase, “susceptible to reasons”.\(^{59}\) This fact is important: Joan’s perception of the woman’s distress (assuming that it constitutes the only relevant moral features in the situation at hand) is both necessary and sufficient for her to be sensitive to moral reasons. Joan’s perception, in other words, enables her to be fully sensitive to the (relevant) moral features present in the case. Thus,

(14) Mature moral agency requires *refined moral perception*, which involves the (accurate) perception of the (relevant) moral features present in particular situations – features that are reasons for (insofar as they recommend or obligate) certain AMJAs.\(^{60}\)

Of course, one could argue that the woman’s discomfort, while certainly a feature of the woman that one can perceive, is not itself a moral feature. After all, I could

---

\(^{59}\) Importantly, McDowell – like Jacobson – is prone to treat his use of the term perception as strictly metaphorical. At one point, he cashes it out in terms of susceptibility to reasons (1997). But as the above discussion suggests, I think we can adhere to a more literal (though perhaps not phenomenological) sense of perception and still get susceptibility to reasons out of it.

\(^{60}\) This is a sort of ‘seeing as’ account: I see your actions as cruel, I see the beating of the woman as unjust. Of course, this is not meant to imply that my perceptions are not veridical (though ‘seeing as’ accounts are often employed to indicate disconnection from truth – e.g., seeing the gray boulder as an elephant). I supposed that a factive account (I see that your actions are cruel) could always be given (if need be) in order to ensure the appropriate connection to truth.
perceive a person’s discomfort that arises from finally confessing an illicit affair to a spouse or from having a rotten tooth drilled at the dentist’s office and yet not perceive any moral reason-giving features (at least, not any that I must be sensitive to). So, there seems to be an important difference between the perception of discomfort and the moral perception of discomfort – how do we make sense of this?

It is important to remind ourselves that grounding features (such as discomfort) do not always instantiate moral features (and when they do, they are not always the same ones). And so, it is not surprising that there are times when the presence of another’s discomfort instantiates moral features and times when it does not. What we can say is that another person’s discomfort is a potentially morally relevant grounding feature: whether or not it instantiates (or contributes to the instantiation of) a moral feature will depend upon the presence (or absence) of a whole host of other potentially morally relevant grounding features. And when it does instantiate a moral feature – that is, when it goes from being potentially morally relevant to actually morally relevant – then MMAs (having, as we established in Chapter 2, a sensitivity attuned to moral features) will reliably perceive it as such.

**III.2 ~ Automated responsiveness (“know-how”) and AMJAs**

Mature moral agency requires refined moral perception – MMAs must perceive the moral features instantiated in particular situations and form accurate moral judgments on the basis of them. But accurate judgment is only one component of AMJAs – MMAs must act appropriately as well. And this is where automated responsiveness comes in.

(15) Mature moral agency requires *morally attuned responsiveness*, which involves spontaneously generated (appropriate) reactions to the perceived
moral features of particular situations – features that are reasons for (insofar as they recommend or obligate) certain AMJAs.

There are few things more beautiful to perceive than someone who is gracefully in sync with their environment – whether it is a ballet dancer moving delicately across the dance floor, a pianist from whose fingers flow the poignant notes of a difficult melody, a highly-skilled telemarker dipping into arcing turns down a steep mountain face, or a child finally getting into the “groove” of walking. This masterful synchronicity – this apparently seamless connection between the organism and its environment\(^6\) – is an essential characteristic of automated responsiveness.

It seems reasonable to expect that MMAs will also demonstrate a similar synchronicity. Both the psychological and philosophical literatures on “wisdom” and on religious and moral exemplars all emphasize this sort of seamless connection: a connection born out of an enlightened detachment that requires a high level of awareness, a tranquility of mind, and an absence of internal conflict (Csikzentmihalyi 1990, Curnow 1999, Liu 2003, Maslow 1968). What is so striking about such exemplars is their apparent lack of effort and uncertainty; the clarity and heightened awareness they demonstrate in their daily interactions (Curnow 1999). Such heightened awareness and balance allows exemplars to perceive the features and objects of their environment as they really are and to respond to them accordingly, without hesitation or uncertainty.

Take, for example, the following passage:

---

\(^6\) I say ‘organism’ here instead of ‘person’ because it seems reasonable to suppose that some version of this synchronicity has existed between living organisms and their environments for many millennia – and has, perhaps, been essential to the evolutionary success of today’s living organisms (see Maturana & Varela, 1987 for discussion).
Consider a normal day in the street. You are walking down the sidewalk thinking about what you need to say in an upcoming meeting and you hear the noise of an accident. You immediately see if you can help. You are in your office. The conversation is lively and a topic comes up that embarrasses your secretary. You immediately perceive that embarrassment and turn the conversation away from the topic with a humorous remark. Actions such as these do not spring from judgment or reasoning, but from an immediate coping with what is confronting us.\(^{62}\)

The skill required to navigate a conversation away from embarrassment without anyone noticing, to comfort someone who is suffering, to speak truthfully about a painful or delicate issue, to find the appropriate time, place, and manner to confront someone about their ethically questionable behavior – or to manage any number of other sensitive moral issues that arise in our daily encounters – is no different in kind than the other sorts of performative skills we develop (and occasionally master). It requires practice, diligence, and experience to “get it right”: i.e., to successfully do what is called for.

Of course, it takes practice, diligence, and experience of a certain sort. One cannot get up at 6-o’clock every morning and run 10 miles in order to train for moral maturity (though to the extent that such activities “build character”, they are not completely irrelevant). It involves, rather, the development (and fine-tuning of) one’s empathic sensitivity to others and one’s self-reflective awareness of oneself, as well as an appreciation of how these two interact in particular situations. It involves a certain clarity of mind and emotion: one must be attuned to things as they really are (not as one wishes

\(^{62}\) Varela (1999), p. 5.
or fears them to be).\textsuperscript{63} It involves being able to \textit{sense} what is morally required, given the details of the situation.

This is the “immediate coping” referred to in the quote above. It is the ability to allow one’s actions to be \textit{sculpted} to fit the situation, almost as if by the situation itself. As Mandelbaum writes:

I sense the embarrassment of a person, and turn the conversation aside; I see a child in danger and catch hold of its hand...in such cases I am reacting directly and spontaneously to what confronts me. ... [I]t is appropriate to speak of "reactions" and "responses" for in them no sense of initiative or feeling of responsibility is present. ... [W]e can only say that we acted as we did because the situation extorted that action from us.\textsuperscript{64}

In the moral domain, such attuned responsiveness is often talked about under the guise of \textit{moral character} or \textit{virtue} – i.e., certain dispositions (or traits) possessed by the MMA that express themselves in the form of appropriate moral action. Possessing the trait of courage is to respond courageously in those situations where courage is called for; possessing a sympathetic disposition is to respond sympathetically to others in moral situations (and so on).\textsuperscript{65} And, since Aristotle, such virtuous traits/dispositions have been thought to naturally and spontaneously lead to appropriate behavior: MMAs do not need to put effort into acting appropriately – \textit{they just do}.\textsuperscript{66}

\textsuperscript{63} In other words, as McDowell points out in his discussion of moral properties as secondary qualities – just as one should fear only that which \textit{merits} fear, one should (dis)approve of only that which merits our (dis)approval.
\textsuperscript{64} Mandelbaum (1955), p. 48-49.
\textsuperscript{65} Such a view is pretty standard fare for virtue ethics (for an overview see Crisp & Slote, 1997).
\textsuperscript{66} This is not meant to imply that my account must necessarily be a virtue ethics account (at least as commonly conceived). For one, my account requires a sort of uni-virtue: MMAs do what is morally
This is a view shared by the eastern philosophical and religious traditions, who likewise emphasize the gradual attainment (through extensive training and practice) of morally attuned responsiveness:

Buddhist sainthood is like all other sainthood in its vision of the perfect life as a life of spontaneous goodness. It is one in which the good or right way of living is as natural as breathing itself, an integral part of the saint’s being. (King 1964, 30)

This is why morally attuned responsiveness is a necessary component of mature moral agency: there must be a tight link between moral agents’ refined moral perception and their moral judgments and actions. What MMAs perceive as morally called for must naturally (and without undue hesitation) lead to the appropriate response. MMAs know what to do, know what action best fits the situation before them – just as the master gymnast, skier, and artist knows. This does not necessarily mean that they could articulate this knowledge ahead of time: rather, they know what to do insofar as they are able to sense what is called for in the moment. And, as in Mandelbaum’s quote above, their actions often flow from them almost without a sense that they are the ones performing them. They flow naturally from their cultivated automated responsiveness. Perhaps this is why Mencius viewed “only people who act from dispositions they have at the very appropriate (whatever that might be) and such AMJAs are not the result of a varied set of dispositions so much as a set of skills that give rise to one unified disposition (the disposition to do what is morally appropriate). Not unlike Dewey, whose ethical views were certainly consistent with (though were never explicitly referred to as) virtue ethics, my main concern is with AMJAs, not with the virtues.

67 In other words, though Blum (1991) was right to point out the independent moral value of refined moral perception in and of itself, nonetheless, McDowell (1997) was also right to insist that the virtuous agent’s perception must be intimately tied to her actions if we are to consider her virtuous. As McDowell writes (referencing Aristotle), “If someone needs to overcome an inclination to act otherwise, in getting himself to act as, say, temperance or courage demands, then he shows not virtue but (mere) continence.” (145)
moment of action as a result of a long process of cultivation [as meriting] the name of truly virtuous” (Varela 1991, 30).

Such dispositions give rise to the experience of doing what one does because it is what one must do – this ‘must’ being felt, not as an externally imposed weight of obligation, but as a spontaneously (yet necessarily) generated expression of one’s developed (and attuned) connection to one’s environment. As Nietzsche wrote,

Artists have here perhaps a finer intuition; they who know only too well that precisely when they no longer do anything ‘arbitrarily’, and everything of necessity, their feeling of freedom, of subtlety, of power, of creatively fixing, disposing, and shaping, reaches its climax… (1989, 156).

This is the ethical “flow” of mature moral agency, no different in kind from the “flow” spoken of by athletes, musicians, and artists: the experience of the unified (and yet almost “authorless”) quality of trained perception and spontaneous responsiveness whose highest aspiration “is to be responsive to the needs of the particular situation” (Varela 1991, 71).

IV. OBJECTIONS

So far, I’ve provided a rough outline of an alternative account of AMJAs that relies on the insights of empirical research on expertise (as well as empirical and philosophical insights on wisdom and the nature of exemplars) to posit two reliable processes that may prove to be central to mature moral agency. With only such a rough outline provided, the questions, challenges, and concerns that remain to be addressed are substantial. I will now take a few moments to address but two: the problem of normativity and the problem of moral discourse.
Problem of normativity: Ethics does not have the kind of standards-driven feedback loops necessary for (performative) skill development.

Problem of moral discourse: A skill-based account of ethics neglects the fundamental role of moral discourse in ethics, which requires being able to make and argue for/against (true) general claims.

IV.1 ~ Objection 1: Problem of Normativity

The problem of normativity could be interpreted in two ways: the first has a fairly straightforward answer, while the second does not. As such, I will start with a brief discussion of the former and end with a more extended discussion of the latter.

On one reading, the worry about standards-driven feedback loops is simply this: look, it is easy to know when you are screwing up when you are skiing or riding a bike – you fall down. And it is also easy to know whether you are performing well or not. In these sorts of activities, there are clear “winners” and “losers”, clear masters and novices. But, how could that be the case for ethical practice? How can we tell who is “performing well” and who isn’t?

At a certain basic level, the answer to this question is simple – ask any parent. Ask any teacher or police officer or guy off the street. They will tell you who is performing well and who isn’t. They will tell you who is kind and who is generous, who is cruel and who lies and cheats to get ahead. Though clearly not as objectively immediate as the experience of slamming face first into the snow, human communities are networks of cultural and moral expectations that provide developing moral agents
plenty of standards-driven feedback. And such standards, while they may be no less are also certainly no *more* “man-made” fabrications than chess, golf, backgammon, or basketball: we decide what counts as success and failure and we measure progress (and expertise) accordingly.

*Faux Virtue*

But, this answer isn’t entirely satisfactory because under the surface, the problem of normativity is actually a fairly thorny one. And this fact is perhaps best illustrated by means of an example:

Imagine Jeremiah, a boy who grows up in a backwoods sort of rural community in Utah, a community completely cut off from the outside world that holds extreme views about both women and race. Jeremiah is raised to believe that his greatest moral duty is to subjugate the local “colored folk”; to beat them, work them hard, and keep them on the brink of starvation, in order to facilitate the great purging of the sins of their people (sins that are evident by God’s blackening of their skin) so that their souls might have a chance to go to heaven. Likewise, he believes that women are passive (though unwitting) vessels for evil spirits and thus should not be allowed out of the house on their own, lest they become possessed by them. The best way to keep them from being possessed by evil spirits, his elders have explained to him, is to keep them “filled” with the Lord’s children. And the more women a man can so protect from the evil minions of the devil, the better.

Throughout his life, Jeremiah receives the strong and univocal feedback from his elders required for his moral and spiritual development. Jeremiah grows into a devout man responsible for the lives of many colored men, women, and children, as well as a houseful of wives. He faithfully engages in those practices that he believes to be for the
spiritual benefit of all those under his charge, demonstrating the high levels of skill indicative of a future elder of his community. He is cherished and respected by his people as a deeply moral and spiritual man. Yet who among us (those who live outside of this isolated community) would be willing to grant Jeremiah the status of a moral exemplar? To many of us, he seems more like a moral tragedy.

The challenge represented by Jeremiah’s case is best captured in the following questions: Do Jeremiah’s beliefs about and his sensitivity to the requirements of being a member of a spiritually superior race give him any genuine reasons (in an objective moral sense) to beat or starve another human being? Indeed, are there any genuine moral features present, such that beating and starving another human being would qualify as an AMJA? Most of us would want to insist that – no, certainly not. But it is not entirely clear how we can do so.

In a discussion of a skill-based account of moral virtue, Jacobson (2006) puts the worry this way:

The fact that the kind person has a reliable, perception-like sensitivity to the demands of kindness does not yet imply that she knows there is a reason to be kind. Knowing how to apply a concept that purports to be reason-giving is a very different thing from knowing what one actually has reason to do.” (394)

Indeed, there have been many supposed (faux) virtues throughout history that today strike us as questionable (if not downright preposterous): the need to die or kill another (or rape his wife) to defend one’s honor, to be a warrior and die gloriously in battle, or to engage in some extreme version of feminine chastity, to name only a few.68

---

68 The Bible (along with, I’m sure other ancient religious texts) is known for having some bizarre/atrocious examples of morally required behaviors: e.g., sacrificing animals, buying and selling slaves, etc.
Such faux virtues purport to be reason-giving, but the reasons they purport to give are, in fact, spurious. Unfortunately for the skill-based account, however, it would appear that Jeremiah’s devotion to racist and sexist practices fits the skill model as well as does another man’s generosity and kindness: their development requires the same sort of practice, imitation, and habituation guided by feedback from socially established standards. As Jacobsen notes, “the skill model thus gives out here, perhaps a crucial moment too soon. Faux virtues, which descry spurious reasons to act, fit that model as well as do real virtues.” (395)

How, then, is a skill-based account of moral maturity supposed to help us distinguish genuine reasons for action (genuine moral features) from spurious ones? The hard-line realist could certainly respond that moral agents from these past traditions and backwoods cultures were (and are) simply mistaken. They have developed sensitivities to the wrong sorts of features: features that they took to be (or to instantiate) moral features, though they were (and did) not. As such, they simply aren’t mature moral agents – rather, they are faux moral agents (i.e., agents possessing faux virtues, sensitive to faux moral features). As hard-line as this seems, it is certainly no more hard-line than the view that most of what people have believed to be true about the world is, strictly speaking, false (including, most likely, most of what we believe today). Indeed, such a view actually puts ethics in good company.

Unfortunately, this is a difficult response for a skill-based account: the main reason for posing a skill-based account is because trained perception and automated responsiveness are the sorts of reliable processes we need for a reliabilist account of mature moral agency. But the above response pushes us into having to choose between
two equally troubling options: either 1) moral agents who are sensitive to faux moral reasons do not \textit{truly} possess the reliable processes (trained perception and automated responsiveness) required for mature moral agency, or 2) they \textit{do} possess trained perception and automated responsiveness, which means that they are not, by themselves, sufficient for mature moral agency: they do not, by themselves, guarantee sensitivity to \textit{genuine} moral reason-giving features.

The first of these options would certainly be an attractive way to account for faux moral sensitivity, but it is not clear how a skill-based account would make good on it: trained perception and automated responsiveness are the sorts of processes that one either does or does not develop and I don’t know what it would mean to have a perceptual skill that \textit{looked} like trained perception or a performative skill that looked like automated responsiveness, but wasn’t. As such, it seems that the skill-based account is forced to the second option: trained perception and automated responsiveness are not, by themselves, sufficient for mature moral agency. Their development does not (and cannot) guarantee sensitivity to genuine moral reason-giving features.

Of course, the same is true in non-moral skill domains as well. Consider chess, for example: a person could be educated and trained to be extremely skilled at a game she \textit{thought} was chess (though it in fact was not). This person could then enroll in a chess competition, only to find out that the game she had learned to play is nothing like the game that all the other people at the competition are playing. Her trained perception and automated responsiveness have, unbeknownst to her, been attuned through training to a completely different game (call it \textit{x-chess}). Thus trained perception and automated responsiveness, by themselves, do not (and cannot) guarantee sensitivity to genuine
chess-relevant features: they do not guarantee that a person will perceive the right sort of patterns or respond in the right sorts of ways.

Perhaps, then, we have given up on the first option too quickly. It is true enough that trained perception and automated responsiveness are the sorts of skills one has either developed or not (they’d be hard to fake) – but this does not yet establish that faux moral agents possess reliable processes. Indeed, trained perception and automated responsiveness are only reliable to the extent that they have been developed to track and respond to the right sorts of inputs. In other words, both must be developed in the right way in order to be reliable – which, in the moral domain, means that both must be appropriately attuned to genuine moral features. Thus, although communities such as the one Jeremiah grew up in are certainly possible (and probably even historically documented), they do not represent a challenge to the ability of trained perception and automated responsiveness to accurately track genuine moral features because theirs, contrary to what they might believe, have not been trained to do so. To the extent that these processes have been developed at all, they have been attuned to other sorts of (culturally and religiously) relevant features – tragically, from a moral perceptive, misguided. This is no different, in point of fact, from a thermometer whose temperature tracking mechanism has been mis-programmed or has been otherwise damaged.69

The upside to this, of course, is that while trained perception and automated responsiveness are not sufficient for mature moral agency, refined moral perception and morally attuned responsiveness are. The crucial question of interest, then, is how trained perception and automated responsiveness become appropriately attuned to genuine moral features.

---

69 Nor would we challenge the reliability of thermometers generally just because of the faultiness of a few.
Feedback Loops

As philosophers as different as Hume and Mencius (Liu, 2003) have noted, our humanity (as much, if not more so than our rationality) is what makes us a unique species. Unless things go terribly awry, we are each born with basic sentiments and emotions that guide and inform our experiences of the world. Of these, the most basic of all is sympathy. The capacity for sympathy – as well as for empathy and love – for our fellow creatures serves as the developmental foundation of both human dignity and decency.

However simple and fragile these early capacities, they serve as the initial constraint and guiding force for moral development. They serve as the early sensors (or “feelers”) of our developing moral awareness, indicators of the presence of certain moral features: most strongly, the “badness” of harm and suffering. Children are quite sensitive to cruelty and kindness, to generosity and selfishness, especially as these are linked with harm to themselves and others. Even before they can walk, young children emit distress responses to signs of suffering and by 18 months will take decisive (if at times, egocentric) actions to comfort and alleviate the suffering of others.

Added to the basic constraints provided by our biological, physiological, and emotional constitutions are the early feedback loops of our social environment. While often better at facilitating sensitivity (often under the guise of morality) to such things as cultural/religious expectations, legal/societal norms, and the like, unless completely off

---

70 Of course, this makes it sound like it is our emotions that make moral features salient enough for us to perceive them, whereas earlier I stated that it is our perception that makes them salient enough to emotionally respond to. Perhaps it works both ways – either way, I think it is right to say that moral perception and emotion are distinct.

track, such early feedback loops represent an important beginning to a moral agent’s moral training: they function to further sensitize developing moral agents to the presence of moral features, as well as their relationship to (and responsibility for) those features. Feedback on the appropriate types of responses (beyond those spontaneously generated by the sentiments) begins here as well.

From there, two sorts of feedback loops become important: *internal* and *external*. External feedback loops come in the form of mentors and outside ethical guidance. In all time periods, the wisdom required for such guidance has been rare: perhaps at no time more rare than it is today. Mentoring and ethical guidance comes in many forms, the most potent of which is living exemplars (and the next most potent of which may be past and/or fictional exemplars). In cultures committed to promoting moral excellence, such exemplars are socially prominent and openly revered: in cultures such as our own, they are not.

It is important that such external guidance not be reduced to a form of “moralization”: mature moral agency cannot (ultimately) be achieved through the rote passing down of ethical *do’s* and *don’ts* (though early training usually begins there). The primary function of external feedback loops is to *promote awareness*, to draw moral agents’ attention to the moral features present in particular situations and to help them to recognize why such moral features have been instantiated, as well as to attune their emotional responses to their presence. Ultimately, the goal of such external feedback must be to promote the development of the agent’s own internal feedback loops: without this development, mature moral agency cannot be achieved.
The psychological literature refers to the creation of mature internal feedback loops as “moral identity formation” (see Blasi 1980, Erikson 1968, Lapsley & Narvaez 2004; very similar, in many respects, to Dewey’s (1960) discussion of moral character formation), which typically begins to occur (if it occurs at all) during late childhood to early adolescence. The formation of a moral identity requires an awareness of the centrality of moral concerns to one’s life and one’s endeavors. It requires the self-identification of oneself as an active seeker of virtue and maturity. This is no different, really, than commitments made to becoming master gymnasts, artists, or pianists: the cultivation of a skill is placed at the center of one’s core of values, making it an essential part of how one defines oneself. Likewise, forming a moral identity places moral concerns at the center of one’s value system, creating the self-reflective structure necessary to monitor one’s progress towards the goal of moral maturity.

There is a paradoxical quality to this process, however: internal feedback loops, if formed improperly, can become warped by the very self awareness they were created to monitor, generating false moral identities. This is why eastern philosophers warned, first and foremost, against self-righteousness. In this regard, Mencius compares the truly virtuous man to the “village honest man” (translated by some as the “bourgeois righteous simulation of excellence”72). The village honest man desires to be seen as virtuous: he is a man “liked by the multitude and is self-righteous” (29). Psychologists such as Maslow (1968) and Moshman (2004) also warn against this and other forms of moral self-delusion: false moral identities function, not as psychological stabilizers that allow for the development of moral awareness, clarity, and detachment, but rather as psychological barriers that protect the individual from having to encounter the world as it genuinely is.

72 Varela 1991, 29.
(thereby making people like Jeremiah possible). Such barriers are commonly psychologically unstable and hard to maintain, often resulting in nervous breakdowns, manias, and other forms of collapse (Moshman 2004).

Thus, moral identity formation can function as an appropriate feedback loop for the cultivation of moral maturity only insofar as it is motivated by a desire not to be or to appear moral, but by a desire for moral awareness and clarity – i.e., a desire to clearly perceive what is morally called for (and to be able to respond accordingly). As such, the end product of the cultivation of a genuine moral identity is the development of the sort of disposition (i.e., one composed of refined moral perception and morally attuned responsiveness) that largely removes the deliberating, desiring “self” (or what is often referred to as the “ego”) from view.

This is a central insight of the Stoics and Buddhists alike: one must achieve a level of objective detachment from one’s own needs, desires, self-interests, fears, and so on in order to perceive the world as it really is (Curnow 1999). As Maslow (1976) writes, “the essential difference between [those who have achieved healthy moral cognition and those who have not] is that [the latter] always refers back to, and is guided and colored by, the needs and desires of the perceiver. On the other hand, [the former] respects the independence and autonomy of the object, which is perceived in its own right…there is a bracketing out of the perceiver, as far as possible, and the object is permitted to be itself” (251). The clear implication of this view is that “only when there is detachment from the ego and self-interest can the world be perceived as it really is” (Curnow 1999, 193).

In summary, though genuine moral awareness is fragile, there are multiple layers of physiological, emotional, psychological, and social constraints and capacities that
facilitate (if not ensure) its development. In reality, deeply morally misguided communities like the one Jeremiah grew up in are surprisingly hard to create and/or maintain. This is not to say that the tragedies perpetuated in communities such as this are not entirely too commonplace – it is rather to say that such tragedies are rarely (if ever) perpetuated with a genuine moral conviction. Where there is genuine emotional and psychological health, there is genuine sympathy, compassion, and generosity. Though such basic responsiveness is not yet moral maturity, it is the essential starting point. And people cannot achieve the sort of genuine awareness, clarity of perception, and responsiveness required for mature moral agency and yet remain blind to the genuine moral features present in the world around them. They cannot fail to perceive the cruelty of their own and others’ actions, the wrongness of their own and others’ behavior. Though this does not entirely eliminate the possibility of faux moral agency and faux virtue, it nonetheless establishes a secure foundation from which to address it.

IV. 2 ~ Objection 2: Moral Discourse

Perhaps one of the most important contributions of principle-based theories to ethical philosophy is a solid grounding in an objective source of normativity that speaks a universally spoken language: rationality. It one manner or another, it is rationality that grounds the normativity of the ethical concepts employed by principle-based theories (however different from one another they might be). They all employ “the light of reason” (whether deliberative or intuitive) to guide the discovery (or construction) of general MPs which are to be employed to identify values, actions, and ends that are rationally defensible (i.e., those that ideally rational agents – or rational agents in ideal conditions – would have, would want us to have, and/or would all agree upon).

73 George W. Bush notwithstanding.
An important consequence of giving ethics this rational foundation is the capacity for moral discourse – the ability to give/take reasons and to make (true) general moral statements. This leads to the second objection I would like to consider, which is that the skill-based account neglects the fundamental role of moral discourse. After all, unlike other sorts of performative skills, ethics essentially involves the giving/taking of reasons, deliberation, and justification – and, as such, it requires being able to talk about moral situations in general terms; to make appeals to general moral claims. We employ such claims (such as ‘lying is wrong’) in many important contexts – pedagogical, justificatory, and explanatory (to name only a few). Yet, the particularist skill-based account I have advocated raises the serious worry that such statements are, strictly speaking, \textit{false} – which means that our appeal to them is mistaken. What, then, is to become of moral discourse?

This is, again, a huge (and Thorny) issue: I will narrow it a bit by restricting my discussion to the status of general moral claims. There are two ways, I think, that the truth (and usefulness) of general moral claims in moral discourse can be salvaged. Specifically, I will discuss two sorts of general statements and the role they might play in moral discourse: 1) defeasible generalizations (or generics) and 2) variant MPs (or ‘rules-of-thumb’).

\textit{Defeasible Generalizations}

Our lives are filled with general moral statements: ‘lying is wrong’, ‘share your toys’, ‘be kind to others’, ‘don’t take that which is not yours’, ‘be polite’, ‘do onto others as you would have them do onto you’, and so on. And there is much to be said for the role that such statements play in our moral engagements: they function as crucial
guidelines and sources of knowledge for moral assessment, even though they admit of exceptions (and thus, must ideally give way to the sorts of skills discussed in this chapter).\textsuperscript{74} One way of understanding such general statements is to put them in the same class as other (non-moral) general statements that we frequently employ: e.g., we employ generic statements (that we take to be true) such as ‘birds fly’, ‘trees have leaves’, and ‘dogs have four legs’, for example, even though there are many kinds of birds (e.g., penguins, emus, and ostriches) that don’t fly, as well as plenty of examples of trees with no leaves and three-legged dogs.

Such statements are referred to as \textit{defeasible generalizations} (Little, 2001) or \textit{generic statements} (Gelman 2003, Väyrynen 2004). Lyons (1977) suggests that, semantically speaking, generics can be translated roughly as meaning \textit{generally, typically, characteristically, or normally} – though not \textit{necessarily}. As such, they are not taken to identify invariant relations: rather, they are taken to express defeasible conditions of \textit{normalcy}. Importantly, such normalcy is of a \textit{modal}, rather than a merely statistical or extensional nature: a generic like ‘birds lay eggs’ can be true even though less than half of the actual bird population does so and ‘turtles live to a ripe old age’ can be true even though only a tiny fragment of the turtle population actually manages to survive the first few weeks of life. This is the case, many believe (see Gelman 2003), because generics identify essential – or \textit{non-accidental} (as opposed to merely accidental)

\textsuperscript{74} Considering the model of expertise development proposed by Dreyfus & Dreyfus (1986, discussed above), this should be unsurprising. As a completely anecdotal example of this, I have often explained to my son the wrongness of his behavior with statements such as ‘lying is wrong’. While this strategy worked quite well for a while, eventually he began to notice and point out to me numerous examples where lying \textit{isn’t} wrong, so I finally had to sit him down and ask him to reflect on a number of different situations and tell me whether, in each situation, lying was okay or not okay. Perhaps not surprisingly, he had a lot of the same reactions that I did. When I asked him \textit{why} it was okay to lie in some situations, but not in others he thought for a moment, shrugged and said “I don’t know – it’s just that in some of the situations it doesn’t seem like a bad thing to do, while in the others it does”.

– features or properties of things. The statement ‘turtles die young’ would thus not be a true generic (even though it would be statistically factual) because it is only accidentally the case that most turtles die young. And, while laying eggs and flying are not properties possessed by all birds, for those that do possess them they are possessed non-accidentally. In summary, the importance of such statements is that they convey information that, while perhaps often false (or non-applicable) at the level of particular cases, is nonetheless true as a generalization.

What role do such defeasible generalizations play in moral discourse? The same, I think, as they do in non-moral discourse: they identify non-accidental, though nonetheless variant, features and properties (in this case, of potentially morally relevant grounding features). To say that ‘lying is wrong’ is thus to say that wrongness is a moral feature that at least some instances of lying non-accidentally instantiate. This sort of a statement communicates something that a statement like ‘wearing red shoelaces is wrong’ cannot – even though it may very well be the case that in a particular situation(s) the wearing of red shoelaces functions as a morally relevant grounding feature (and in some possible worlds may even do so more frequently than lying). Being wrong just doesn’t seem to be the sort of property that the wearing of red shoelaces could non-accidentally instantiate (no matter how many times it happened to do so): there is nothing about the nature of red shoelaces that would make the wearing of them (non-accidentally) wrong. For lying, on the other hand, there is.

Importantly, this is not to say that MMAs would fail to be sensitive to the role of red shoelaces in instantiating wrongness, should they happen to have one: MMAs are sensitive to the presence of moral features regardless of the nature of the grounding
features that instantiate them. Likewise, MMAs are sensitive to the presence of exceptions. As has already been demonstrated, the truth of generics (be they moral or non-moral) is often irrelevant to the facts actually present in particular cases. The fact that ‘turtles live to a ripe old age’ does not stop all but a few young turtles from being eaten by seagulls and other hungry predators at a particular hatching and the fact that ‘dogs have four legs’ does not stop Fluffy from being a three-legged dog. Nor does ‘lying is wrong’ stop many instances of lying (such as Hilda’s) from not being wrong.

Indeed, this fact raises an important worry for the use of generics in moral discourse. While the fact of most turtles’ young deaths is an accidental fact that has no bearing on the (non-accidental) truth of the statement ‘turtles live to a ripe old age’ (and, likewise, the fact that Fluffy has three legs is an accidental fact that has no bearing on the truth of ‘dogs have four legs’), it seems that this is not the case for something like lying. Take for example, the fact that Hilda’s lying to the Nazi soldiers was the morally right thing for her to do. Unlike its non-moral brethren, this fact does not seem to be a merely accidental fact that has no bearing on the truth of the statement ‘lying is wrong’. It seems false to say that in Hilda’s case (or any other cases in which lying is right) the rightness of lying in that case was accidental: indeed, the rightness seems to be just as much a non-accidental feature of lying (when it is instantiated) as wrongness does. This brings the legitimacy of ‘lying is wrong’ as a true generic into question (and, as we’ve already seen, appeals to statistical frequency – e.g., lying is more frequently wrong than right – or any other such consideration are irrelevant).  

Little (2001) tries to appeal to the notion of “deviancy” to make sense of the truth of the generic ‘lying is wrong’: the fact that lying is (non-accidentally) right in Hilda’s case, then, would not over-turn the truth of the generic because the conditions that made her lying right were morally “deviant” – (i.e., the connection that typically holds between lying and wrongness failed to hold because something was morally amiss or
Perhaps what *can* be truly said (in generic form) about lying is that it is *morally relevant* (in a way that red shoelaces and other such things are not).  

That is, it is true that there is something *about lying* that makes it (non-accidentally, though not invariantly) a good candidate for moral relevancy. If lying is present in a particular situation, then it will *normally* (though not always) play a role in the instantiation of whatever moral features are present: as such, it is one of those grounding features that, when present, it is worth paying attention to.

Thus, while it is questionable whether generics of the form ‘lying is wrong’ are true, generics may nonetheless be legitimately used to identify those (types of) grounding features that are non-accidentally morally relevant. Given this, there are probably a whole host of true general moral claims that can both be employed in our moral discourse and can serve as valuable tools for moral education and negotiation (though not, as many people argue, in the form of MPs).

*Variant MPs*

Given the above discussion, perhaps statements such as ‘lying is wrong’ should be understood not as defeasible generalizations (or generics), but rather as variant MPs (or *rules-of-thumb*). The important difference between these two is that while the normalcy conditions of generics are determined modally, for variant MPs they are strictly
defective in that particular situation. As appealing as this approach is (the extremity of Hilda’s case making it seem even more appealing), I don’t think it works: there are many cases in which some form of lying or another is appropriate even though the situations it which it occurs would hardly qualify as morally deviant (e.g., Santa Claus or playing various forms of games).

Or, an alternative would be to recognize both ‘lying is wrong’ and ‘lying is right’ as true generics.  

As Asher and Bonevac (1997) point out, the ‘normally’ of this example could be cashed out in one of two ways: 1) as “holding under normal circumstances” or 2) as “holding ceteris paribus (all things being equal)”. Though logically similar, these approaches have somewhat different implications.

Of course, this is a very cursory treatment of a very important subject, upon which many good philosophers have written. Unfortunately, I do not have the space here to give it the treatment it deserves.
*statistical:* rules-of-thumb are only useful as heuristics to the extent that they identify relations that are typically encountered in *actual* cases.

As such, an update to the definition of variant MPs is warranted:

(2*) A MP is a genuine variant MP iff (a) its grounding feature(s) is variably morally relevant or (b) its grounding feature(s) closely (and reliably) approximates a genuinely invariably morally relevant grounding feature(s) and (c) the grounding feature(s) most frequently instantiate morally relevant features of a particular *valence (in the actual world).*

If in the actual world lying frequently instantiates wrongness, then the variant MP ‘lying is wrong’ will be both *true* (as a variant MP) and *useful* – for certain pedagogical, justificatory, and explanatory purposes. Of course, it is important to remember that with regards to particular situations, such statements do not (and cannot) tell the whole story: that wrongness has been (or will be) instantiated in a particular case cannot be fully accounted for non-holistically. In other words, it will never be solely and completely because there was lying involved that a particular action was wrong (or, for that matter, right): it will depend upon the presence (or absence) of a whole host of other situation-specific grounding features. Nonetheless, the value of such general statements as heuristics should not be underappreciated – not only for their educational value, but also for their capacity to provide guidance in moral discourse. As with any statistical information, while knowing the odds cannot provide us with certainty, it can nonetheless give us a rough estimation of what to believe and what to expect. In addition, such
information gives us a fairly good idea of where deliberation, discussion, reflection, and negotiation should begin (though not, as some would have it, where it should end).

V. CONCLUDING THOUGHTS

If I were to attempt my hand at any sort of “take home message” from all of this, it would be to impress upon the reader the rich and daunting complexity of our moral lives. It some respect, it should come as no surprise that we cannot get off so easy as to formulate a handful of robust MPs to guide us successfully through the rough terrain. And while my concluding remarks about generics and variant MPs provide reason to be hopeful that (true) generalizations are both possible and useful, I think the onus is nonetheless on us to begin to pay closer attention to the hard road ahead. As philosophers, psychologists, and educators, the job of exploring, understanding, and developing the sorts of processes that I’ve gestured at here will be challenging (at best). We know so little about what allows for the skill, grace, and beauty of expertise generally speaking, about what processes and mechanisms are involved and about how to best facilitate their development: we know even less about their moral equivalent.

The recognition of the ultimate inadequacy of MPs pushes us to delve further, to come to grips with the extent of our moral infancy and responsibility. Knowing what is morally called for can be no longer be simply a matter of what can best be argued for: it becomes a matter of psychological and emotional health, of self-awareness and reflection, of balance and practice and life-long commitment.

It is the recognition that to be an ethicist, one must be willing to be ethical: not in some “moralistic” sort of way, but in a Sam Menard “sixth sense”\textsuperscript{79} sort of way. One must be willing to live in sync with one’s moral environment – fully open, aware, and

\textsuperscript{79} This is a reference to the quote at the beginning of Chapter 3.
engaged with oneself and with the world. As Alice Borchardt wrote at the beginning of her novel, *The Dragon Queen*, “…and we sought truth, enlightenment, love, and beauty in each other’s faces, hands, hearts, and bodies – not in the shadowed, crackling pages made of paper and parchment” (p. ii).
References


Dreyfus, Hubert. Unpublished draft. “A Phenomenology of Skill Acquisition as the basis for a Merleau-Pontian Non-representationalist Cognitive Science”.


King, Winston. 1964. In the Hope of Nibbana. La Salle, IL: Open Court.


