THE SHAPE OF ETHICS:
KANTIAN MORAL AGENCY UNDER REASON HOLISM

Elijah Barton Barrish

Plan II and Philosophy Honors
The University of Texas at Austin
Fall 2017 - Spring 2018

____

Jonathan Dancy
Supervising Professor

____

Jon Litland, Ph.D.
Second Reader

"[T]he living all make the same mistake: they distinguish too sharply."

- Rainer Maria Rilke, *Duino Elegies*

Table of Contents

Introduction ... 1
- I. A brief outline ... 1
- II. The holism of practical reasons ... 2
- III. Shape and moral explanation ... 4

Kantian Moral Agency ... 8
- I. Kant's project: an argumentative summary ... 8
- II. Kant's account of reasons ... 15

Defeasible Moral Reasoning ... 22
- I. Stating moral principles ... 22
- II. Do moral principles tell the truth? ... 25
- III. Defeasible generalizations ... 29
- IV. Default reasons ... 33

Moral Explanation in the Kingdom of Ends ... 38
- I. Explanation in brief ... 38
- II. Preliminaries to explaining moral phenomena ... 41
- III. Modeling moral principles ... 44
- IV. The shape of ethics ... 47

Bibliography ... 50

Introduction

I. A brief outline

Consider two views of moral principles:

> Immanuel Kant: "Act only on that maxim which you can at the same time will that it should become a universal law."[1]
> Jonathan Dancy: "[M]oral judgment can get along perfectly well without any appeal to principles, indeed…there is no essential link between being a full moral agent and having principles."[2]

The views appear to conflict over the role of principles in moral judgment. If one takes universal moral laws to be principles, then Kant seems to require them for moral judgment. Kant does not insist that moral agents "appeal" to principles per se, but that their actions be based on maxims that could at the same time be willed as universal laws. Dancy, stating the thesis of moral particularism, suggests that agents do not need moral principles at all.

Despite this tension, the views are compatible—even complementary. A shortfall of Kantianism (it is thought) is its struggle to explain moral conflict. Against common sense, Kantianism seems to demand (e.g.) that moral agents never lie, even when it would be clearly for the worse to tell the truth. Moral particularism responds to these seemingly exceptional circumstances with aplomb, showing how reasons favoring an action in one case can in another case count against it.

Particularism struggles with the next explanatory step. When asked why reasons behave as they do, particularists can mention the salient features of a situation, but they lack an overarching theory. Thomas E. Hill, a Kant scholar, asks the particularists, "What explains why these particular facts, not others, are good reasons to act? What do the good reasons have in common?"[3] This roughly expresses what I call "the problem of shape." It is the problem of the apparent shapelessness of nonmoral facts with respect to the moral. Particularists, by and large, defend shapelessness, since their primary aim is to reject the exceptionless principles that would give shape to the nonmoral with respect to the moral. But (I will soon argue) Hill is right to find the particularists' non-response unsatisfying.

This issue motivates me to find particularism a theoretical home in the Kantian system. One might think of Kantianism as explaining the "back-end" of moral theory—its underlying machinery—while moral particularism explains the "front-end"—its behavior in concrete circumstances. This basic picture, I take it, prompted Robert Audi to combine intuitionism and

---

[1] Immanuel Kant, *Groundwork for a Metaphysics of Morals*, 4:402.
[2] Jonathan Dancy, *Ethics without Principles*, 1.
[3] Thomas E. Hill, *Virtue, Rules, & Justice*, 234.

Kantianism in *The Right and the Good*, and Mark Coeckelbergh to combine pragmatism and Kantianism in *Imagination and Principles*.

As with these previous attempts, the Kantian position defended here is not, strictly speaking, *Kant's* position, which is to say that in places textual evidence will speak against me. Kant is too rigid to fully accommodate particularism. But every marriage requires compromise, and the core Kantian commitments—centering pure reason and human autonomy—get preserved.

Particularism, meanwhile, will sacrifice Dancy's trenchant rejection of moral principles, instead accepting them as valuable heuristics for acquiring moral knowledge. Principles may play no *essential* role in moral agency, as Dancy says, but I have yet to encounter a moral agent who gets along well without them. Principles provide reference points for explanation, debate, and judgment in ethics. The question is how to understand the role of principles. Principles almost universally have exceptions. Yet moral agents still depend on them.

In the following two sections of this introduction, I will give the main argument for moral particularism based on the holism of practical reasons, then lay out "the problem of shape" in greater detail. The first chapter has two sections, the first argumentatively explaining Kant's project in ethics and the second showing how it can meet particularism's theory of reasons. The second chapter tackles the relationship between Kantian maxims and defeasible moral principles. The third chapter develops a theory of moral explanation for the Kantian moral particularist.

II. The holism of practical reasons

Moral particularism, as formulated in Dancy's *Ethics without Principles*, is the view that moral judgment does not depend on moral principles. Particularism follows from an analysis of moral principles. If principles express exceptionless laws governing our actions, as we tend to think that they do, then they have a fatal flaw. Principles typically link a general feature of actions (e.g. that they promote happiness) to a normative property (e.g. rightness). The general feature is supposed to guarantee the normative property. But this link has proven tenuous. The same feature can make an action morally obligatory or morally heinous, depending on context. This theory is known as "reason holism."

More specifically, it is the holism of *practical* reasons. A practical reason favors or disfavors an action (a theoretical reason favors or disfavors a belief). Holism is defined by Dancy as the idea that "a feature that is a reason in one case may be no reason at all, or an opposite reason, in another."[4] It opposes atomism, the view that a feature that favors in one case will also necessarily favor in another.

---

[4] *Ethics without Principles*, 7.

I will henceforth assume holism, but it shouldn't be a hard pill to swallow. Reflection on features of actions makes holism immediately apparent. Consider happiness promotion. We tend to think that if an action promotes happiness, the fact that it promotes happiness is a reason favoring it, morally speaking. But what if the action is torture and the happiness is the sadistic happiness of the torturer? One might think the action would be better if the torturer took no pleasure in his task. In the case of the sadist, happiness promotion seems to be a reason counting against the action.

That happiness promotion can make an action wrong challenges our preconceptions. Faith in exceptionless moral principles has a long history, recalling religious commandments and piggy-backing on Kant's influence over Western ethics. The cross-pollination of moral principles with legal ones reinforces the notion that our morals, like our laws, should be airtight. But for as long as principles have been around, they have also been challenged. Never has a principle escaped scrutiny; never has a principle not implied seemingly odd or unintuitive consequences, results that we wish would go away. Sometimes it seems for the better, or even obligatory, to lie, steal, and cheat. Philosophers might prefer a tidier moral plane, but here we are.

Under holism, the overall rightness or wrongness of an action—its moral valence—comes from the morally relevant features of the action all taken together. The features favoring or disfavoring the action are called contributory reasons: they push the agent one way or the other. These reasons compose a "resultance base," out of which emerges the valence.[5] Not every feature of the action makes it into the base. That the action kept a promise makes it into the base, but that the action occurred on Tuesday does not. Or, it doesn't unless the promise involved a commitment to take the action *on a Tuesday*; then the day of the action would be a relevant feature. The resultance base includes not just the features that directly favor or disfavor the action, but also features that make other features relevant or irrelevant: call these enablers and disablers—together, the conditions. It also includes features that alter the magnitudes of the contributory reasons—the modifiers. That the promise was to a friend instead of a stranger exemplifies an intensifier, strengthening a favoring reason. A feature that weakens another reason is an attenuator.

The three ways a feature can be relevant—as a contributory reason (favorer/disfavorer), as a condition (enabler/disabler), or as a modifier (intensifier/attenuator)—tell us what constitutes a resultance base. Moral valence, which can be thought of as the "direction" of the resultance base, indicates the action's overall rightness or wrongness. An action can have favorers and disfavorers; most actions do. We face conflicting reasons every day and must decide which reasons outweigh. Only features pointing in the same direction as the action's valence make it into the base, because it is from them that the valence results. Inspecting the architecture of the base will tell whether an action is overall right or overall wrong.

---

[5] Ibid., 85.

This completes my summary of practical reason holism and moral particularism. However, a litany of questions remains unanswered concerning the *nature* of practical reasons. I have described how reasons behave in a resultance base, but this is just a metaphor, just as the notion of a favorer "pointing" towards the action is a metaphor. As an analogy, recall a complaint about the correspondence theory of truth. In its most rudimentary form, the correspondence theory holds that a sentence is true when it expresses a state of affairs. But the theory (critics say) is silent on what a correspondence relation *is*. We can circle the true sentences without knowing what it means for a bunch of words to "correspond" to reality. Similarly, knowing what roles reasons play tells us the superficial behavior of reasons, but doesn't reveal their nature. Particularism's account of practical reasons answers the "what" question (what are the ways that reasons can contribute?) without answering the latent "why" (why do they favor or disfavor *here*?).

One of Kant's triumphs, in my view, is to answer the "why" question without stipulating a single teleological end. Kant says, at the beginning of *Groundwork*, that the only unconditional good is a good will. Other things can be good in some circumstances, bad in others; among these holistic items are intelligence, wealth, power, courage, and happiness.[6] The good will is one whose principles of willing (maxims) are formulated according to the Categorical Imperative (CI), the formal law of morality. Proper formulation of maxims, I will argue, requires careful attention to reasons. We have, then, a rough Kantian sketch of what a good reason is: it is one that appears in a properly formulated maxim. Insofar as Kant defends a "final end," it is pure rationality, which is just the capacity to perfectly evaluate reasons. But much more will need to be said on how this brief Kantian sketch of favorers and disfavorers can fit with particularism's other insights about reasons, including the features acting as modifiers and conditions.

III. Shape and moral explanation

Critics of moral particularism express concern about the "shape" of the moral plane. They worry that particularism implies that the moral is shapeless with respect to the non-moral. Moral facts are about right or wrong. Non-moral facts typically describe the observable world. Some philosophers think that every fact in the universe is reducible to non-moral, natural facts. But even among those who don't, there is typically a shared intuition that the moral *must* be patterned with respect to the non-moral.

What motivates this intuition? There are arguments against particularism that say patterns *must* exist, but, putting those arguments aside, I think the intuition is twofold.[7] First, because it is

[6] *Groundwork*, 4:393-4:396.
[7] The argument I put aside is presented by Jackson, Pettit and Smith in "Ethical Particularism and Patterns" (2000). I feel justified in putting it aside because of Dancy's excellent response in "Can a Particularist Learn the Difference Between Right and Wrong?" (1999).

widely accepted that the moral supervenes on the non-moral, it's thought that it must supervene in a pattern. An action won't go from right to wrong without some change in its non-moral features. But this doesn't mean that the moral is patterned per se; changes in the moral could correspond to jagged, haphazard, unpatterned changes in the non-moral. Critics of particularism seek a natural harmony that the world seems unwilling to provide.

The other part of the intuition is that some features of the natural world seem more morally *charged* than others. Honesty and promise-keeping feel ethics-laden in a way that color and the time of day do not. Lance and Little, two moral particularists whose understanding of moral principles I will later critically evaluate, put it like this:

> First, it is said, moral particularism seems to imply that there is no more intimate a connection between honesty and the good than between, say, shoe-lace color and the good… it seems odd to think honesty is only accidentally related to moral status. Second, it is claimed, the view seems to imply that there is no structure to a moral theory at all. Moral understanding must be simply a matter of accumulating one-off pieces of insight…[8]

Lance and Little commit a slip in constructing the particularist as claiming an only accidental connection between honesty and the good. This is a strawman, because honesty as a "thick property," combining descriptive and normative elements. But reason holism does mean that shoe-lace color and, say, giving money to charity are on equal footing.

Why should this be a problem? One might think that although philosophers would prefer a tidier normative plane, wishing so doesn't make it true. Faced with Thomas E. Hill's sharp questions about shape ("What explains why these particular facts, not others, are good reasons to act? What do the good reasons have in common?"), the particularist could dig in their heels and insist that good reasons don't have things in common; they depend too much on context. There is no principled difference between shoe-laces and giving money to charity, no deeper fact to call upon to explain why each reason contributes as it does in various circumstances.

But if this is all the particularist says, then they give up significant desiderata of moral philosophy. Robert Audi discusses two types of completeness to which moral philosophy aspires. The first, "normative completeness," is achieved when a theory provides "principles or standards in the light of which every overall (i.e. final) moral obligation can be plausibly exhibited as such."[9] The second, "epistemic completeness," is achieved when a theory can "take us from facts about human life…to what we ought to do."[10] I use the term *explanatory completeness* to sweep up both types, because what we want to do is fully explain why an action has a normative property given an action and moral valence, or fully explain what one ought to do given facts

[8] Lance and Little, "Defeasibility and the Normative Grasp of Context," 294.
[9] Robert Audi, *The Good in the Right*, 85.
[10] Ibid.

about one's circumstances. Audi's two types of completeness are not strongly distinguished; they converge on the same desideratum from opposite directions.

Like Audi's intuitionist, the particularist can claim first-order normative completeness, which involves explaining particular wrongs and rights in light of the facts at hand; a particularist will just indicate the morally relevant reasons. But the particularist lacks *full* normative completeness because they will not be able to explain *in light of what* the facts at hand produce a wrong or right valence.[11] Audi fears that though the intuitionist has a list of prima facie duties, they have few principles of combination for them. Still, the intuitionist is better positioned than the particularist, who doesn't even have a list of duties.

At first pass, full normative completeness might not seem unequivocally desirable. The particularist, after all, rejects codification of the moral plane. What I will show, though, is that there are models of explanation that do not depend on codified exceptionless laws or principles, as Audi assumes.

The problem is both theoretical and pragmatic. Because morality is shapeless, an agent can't pick up on patterns in the nonmoral plane that, repeatedly and reliably, underlie certain moral valences. They can't *explain to themselves* why an action they are considering is obligated. Going the other direction, an agent can't fully explain why an action is right or wrong (especially actions taken with complex considerations) in light of the world they experience. In complex cases, cases where we would like to explain moral phenomena as completely as possible, particularism comes up short.

I will offer an example to demonstrate the urgency of moral explanation. Some feel very strongly that abortion is immoral under any circumstances; others disagree. For clarity, let us naively ignore what gets lost in translation between the ethics of a *particular* abortion and the ethics of certain legislation. There are reasons favoring abortion and reasons counting against it. On the "against" side, consider one major potential disfavorer: that abortion involves the murder of a child. But this disfavorer requires some dubious enablers. It requires the additional fact that the fetus is child. If the fetus is not a creature of moral worth after all, then the disfavorer is disabled. Among the favoring reasons, there are facts about what might occur if the pregnancy is carried to term. The unwanted child's life might be sub-par. The mother's life might be thrown into disarray. A stack of reasons favors abortion, and one powerful reason may count against it.

How to evaluate the dilemma? Look first to the empirical evidence. The particularist has no prewritten moral rules pre-deciding the issue, so she ignores scientific advances at her peril. If prenatal science strongly indicates that the fetus, even at six weeks or so, feels intense pain during an abortion, then this would be a devastating disfavoring contributory reason.

---

[11] My language is ambiguous here between explaining *in light of what* a specific valence is produced, or explaining in light of what *any* valence is produced. The latter question asks after the source of normativity. How and why do some nonmoral circumstances make claims of positive or negative obligation on us? I have the former sort of explaining in mind (specific valence), but the Kantian system will help to answer the latter as well.

Sociological evidence on the negative impact of unwanted pregnancies on young women could convincingly favor abortion. Some other evidence is of dubious relevance. What if the father of the fetus strongly opposes termination of the pregnancy? Some would say that he has no say; it isn't his body. This feature may or may not be a contributory reason.

For believers in exceptionless moral principles (generalists), this dilemma is easier, though not trivially easy. If a generalist considers a fetus a person and holds an exceptionless law against killing, then all abortions would be unethical, even abortions to save the mother's life. Exceptionless rules make moral judgment deductive. Particularism has no obvious way to even continue debating the issue. James D. Wallace, in critically evaluating ethical intuitionism, puts it like this:

> If an argument could be produced to make a case for the superiority of one solution to a conflict problem over another, the argument must involve an appeal to one or more general considerations or principles that imply that one of the conflicting considerations takes precedence over the others.[12]

Expressing concern about intuitionism's limits, Wallace describes an explanatory shortfall. Intuitionists and particularists will insist that a moral phenomenon is a certain way, but, when pressed to explain why it is that way and not another, they can't depend on deduction. But deduction is not the only type of explanation. The more fully we explain, the better we understand our moral life. I tend to think that full explanatory completeness is a pipe-dream; our explanations are apt to bottom out somewhere. Conversations about complex moral dilemmas should probably never "end." But shapelessness forces particularists to end such conversations far too early. Embedding particularism in the Kantian system empowers particularists to trace the roots of reasons, normativity, and moral laws more deeply. Defeasible moral principles, together with a new theory of explanation, will give particularism the capacity to better explain moral phenomena.

---

[12] James D. Wallace, *Moral Relevance and Moral Conflict* (Ithaca: Cornell University Press, 1988), 22.

Kantian Moral Agency

I. Kant's project: an argumentative summary

The Good Will

I will briefly summarize what I take to be the core commitments of the Kantian moral system, highlighting where they contradict or resonate with particularism. The four areas of summary are the good will, practical identity, the first two formulations of the Categorical Imperative (CI), and maxims. In the next section, I will dive more deeply into maxims by way of reasons. Reasons are one of the main sticking points between particularists and Kantians, along with the possibility of universal moral law. Both points require expository work. The previously expounded caveat applies: *my* Kant is no one else's Kant, and certainly isn't Kant himself. I fully confess to cherry-picking the parts of Kant's works that work to the benefit of my argument. However, as I hope this exposition will show, I will not shy away from moments in Kant that should trouble the particularist, so long as they are important to the overall Kantian system. I take the Formula of Universalizability to demonstrate the sincerity of this aim.

Kant starts with the good will, which he determines is of "absolute worth", meaning that its value doesn't depend on anything else.[13] This is because, roughly, the will is what *assigns* value. As Korsgaard says, it is only because "we are self-conscious animals, capable of reflection about what we ought to believe and to do," that normative questions can be raised.[14] Korsgaard argues that the capacity of willing is what explains the existence of normative concepts. We might ask what sense to make of the use of "ought" in her statement, if we are using the term in explaining the source of normativity. Well, Korsgaard might respond, we ought to value the capacity that enables "ought-ness," namely, the will. We might still demand to know why, just because the will is precondition for valuation, it is rational to value the will? Were the will an odd machine that, when turned on, uncorked normativity into the world, would the machine value itself?

I won't keep going in circles; I cannot improve on Korsgaard's account. The notion of "absolute worth" was brought up to make just two points. The first is that, even if we are left somewhat puzzled about the source of normativity, we still enjoy greater normative completeness via the Kantian system. Kant can at least trace normativity back to the good will, if not further. That's much better than simply pointing to a principle, like the Principle of Utility, and saying that it must be right because it always holds; or pointing to a list of duties. The second point to make is that every action derives value from the will, even the will itself. Actions derive value *directly* from principles of the will, as opposed to (e.g.) their consequences.

---

[13] *Groundwork*, 4:394.

[14] Christine Korsgaard, *The Sources of Normativity*, 46.

Kant describes the will as "a kind of causality of living beings in so far as they are rational" because the will produces effects in the sensible world.[15] Rational humans cause change in accordance with their will; irrational humans (or animals) don't. When faced with temptations like happiness and wealth, a moral agent might act solely based on inclination unless the will is present to "correct and make generally purposive their influence on the mind, and with it also the whole principle of acting."[16] The will can *choose* what to act upon. So far as it is rational, untainted by animal desire, the will makes choices in a regimented way, according to principles that it selects for itself. Principles of willing are called "maxims." The good will is one whose maxims accord with the CI.

At first glance the image of a law-bound will whose laws in turn must accord with higher law, looks very strict. It looks as if the ideal agent would have a robotic will, programmed with a set of ideal maxims. But one thrust of my argument is that this depiction gets exaggerated. Each human will must legislate for itself, a robot writing its *own* code. It is, moreover, critically important than human beings consider themselves free (for reasons I won't go into). And although maxims must accord with the CI, this is a formal law that doesn't overdetermine the content of maxims. It should be liberating to learn that the only thing of absolute worth is the good will and the only externally-legislated "rule" (to keep it distinct from self-imposed laws) is the CI. Dancy once criticized principled ethics for discouraging agents from taking moral situations on a case-by-case basis, terming it the problem of "looking away."[17] Kant makes sure that agents attends to each of their actions just as much as they attend to their own rationality, because right actions are enjoined by self-imposed laws of rationality. Korsgaard writes on self-imposed maxims:

> From a third-person point of view, outside of the deliberative standpoint, it may look as if what happens when someone makes a choice is that the strongest of his conflicting desires wins. But that isn't the way it is *for you* when you deliberate. When you deliberate, it is as if there were something over and above all your desires, something which is *you*, and which *chooses* which desire to act on. This means that the principle or law by which you determine your actions is one that you regard as being expressive of *yourself*.[18]

The agent engages in a creative process when the act. They write their own will into coherence. Now, this view of human reasoning might be challenged as overly optimistic. Most of us struggle to see ourselves from "over and above our desires," as if we were an objective third party. But Korsgaard's view is aspirational. It is us at our best, at our most rational. And when we refuse reason, instead following inclination, regret often follows. We know, rationally, that

---

[15] *Groundwork*, 4:446.
[16] Ibid., 4:393.
[17] Jonathan Dancy, *Moral Reasons*, 63-64.
[18] *The Sources of Normativity*, 55.

we *should* do what we think is best. We know that we shouldn't eat our roommate's food, though our stomach may demand it at the moment. When we listen to the "should" and refuse the call of the stomach, we impose a law of willing on ourselves. We choose to act on our reason for not eating and in so choosing we become a more determined, steadfast person, and a better roommate. Each day we build our identity.

Practical Identity

Take two torturers. One is Jack Bauer. He tortures a suspected terrorist for information on an upcoming terrorist attack. Supposing that Bauer is reasonable to think that the suspect has valuable information, that torture will procure it, and that the information will stop a catastrophic attack, Bauer's action is morally right. He acts from good reasons. Engaging in torture causes Bauer feelings of regret, and likely PTSD, but it also expresses his sense of duty to the people he protects. Meanwhile, in Abu Ghraib, a CIA operative tortures someone suspected of wrongdoing, but does so for pleasure. The operative claims that the suspect might have important intel, but he doesn't really care either way. He tortures because he wants to torture.

The Kantian and the particularist correctly judge the torturer receiving pleasure from his task to be more morally repugnant than the reluctant torturer. Pleasure acts as a wrong-maker for torturing. Kant need not insist that torture is always wrong because its wrongness will always depend on the principle of willing enjoined to it. Wrongness comes not solely from the action but from the reasons behind the action. Korsgaard's concept of practical identity empowers the Kantian to embrace this nuance.

Practical identity is "a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth taking."[19] It is fashioned with each maxim we adopt, each choice we make in the exercise of our own will. One's practical identity is complex. In addition to conceiving of ourselves as rational beings, we also conceive of ourselves as fathers and mothers, doctors and lawyers, advisers and advisees. Though these descriptions are a posteriori and contingent (one might quit doctoring and take up lawyering), they place one in a network of duties. Doctoring imposes obligations that lawyering doesn't. No role can contradict one's self-conception as a rational will—our rationality is the aspect of our identity that is indispensable—but they add to it, coloring in the lines that the pure will draws.

One's identity must cohere. It will not tolerate persistent contradiction. Yet sometimes contradictions do emerge. They often emerge as moral dilemmas. Korsgaard gives the example of a person whose practical identity includes conceiving of themselves as a soldier and as a human. The person is ordered to massacre innocent civilians. In this case, one faces a "destabilizing effect," as a rogue obligation tugs against the integrity of one's identity. But one's identity as a human is far more important than one's identity as a soldier. As Korsgaard says, 'some parts of our identity are easily shed, and, where they come into conflict with more

---

[19] *The Sources of Normativity*, 101.

fundamental parts of our identity, they should be shed."[20] It isn't pleasant or easy to shed layers of practical identity, to give up conceptions about ourselves. But sometimes we need to do so out of duty. This goes to the heart of civil disobedience, renouncing the identity of citizen—a weighty sacrifice—in favor of one's human identity.

Bauer, like all compelling dramatic characters, suffers contradictions of identity. On the one hand, Bauer is committed to protecting the United States from terrorism. His first duty is to his country; one of his most important identities is as an officer of the law. But sometimes Bauer needs to work outside the law to protect his country. He is a committed father, but has to make hard choices when terrorists take his daughter hostage. *24* is hardly Shakespearean tragedy, but Jack Bauer illustrates how a man utterly driven by duty nonetheless engages in creative moral decision-making. He must pick the right reasons for action, pick some duties over others. He, like all Kantian agents, is autonomous. Forced to reckon with the complexities of the world and without external laws to obey, Bauer is like a nation unto himself.

Maxims and the Categorical Imperative

The Categorical Imperative is the formal law of practical moral reasoning. It is an unconditional requirement on the wills of all rational actors. That said, the CI is not the sort of law that legislates over *actions*. Instead, it legislates *laws*; it is a meta-law. It vaguely determines the form maxims take, but much is left up to the agent. Maxims, the principles of willing, must *at minimum* include an action and a reason. Maxims may also include a "trigger," some condition making a maxim relevant to a situation, as in the maxim "Whenever one has an opportunity to lie, one shouldn't do so, because it violates trust." This sample maxim has three parts: {A, T, R}, for {action, trigger, reason(s)}, though in other maxims {T} is sometimes left empty. There can also be more than one reason, as I argue in the next section.

Regardless, the maxim seems to express a regular association between A and R. At first glance this appears problematic for the Kantian particularist. If the association between action and reason(s) is not just regular but *necessary*, having the character of a *law*, then not only would the maxim contradict the holism of reasons, but also push against the creative, self-legislating character of Kantian moral agency I have been defending. It would mean that humans are more robot than country, more programmed than programm*ing*.

There are a few things to note here. First, note that the problem I'm discussing is not about the Kingdom of Ends, wherein supposedly all beings follow the laws that would hold if everyone were treated as an end. The problem, rather, is that one's *own* laws become tyrannical. For instance, if one self-imposes a maxim that says, "I will not lie, because lying is a betrayal of trust," then it seems like in future instances, one must not lie—never lie, even when new circumstances seem to make it for the better. Maxims start as creative and freely chosen, but end up weighing one down, like a treaty that outlasts its purpose when the world changes.

---

[20] Ibid., 102.

However, Kantian maxims display a non-monotonic logic. Non-monotonic logics capture defeasible reasoning, which occurs when an inference is conditionally asserted, true in a case but vulnerable to changing circumstances. If maxims were monotonic, that would cut against autonomous moral agency. When an agent applied a maxim like the example prohibiting lying, that would be the end of their reasoning. Then each time they were in a situation where lying was a possible action, their own maxim would exert a repressive force, and they would have no option to lie for other reasons. But that makes no sense. A maxim includes an action *and* reasons. Just because X reasons prohibit lying does not mean that Y reasons must also prohibit lying. Just because I adopt a maxim that prohibits lying to get ahead at work, does not mean I have a general prohibition against lying *as such*. This view simply misunderstands the structure of maxims by overemphasizing the action and underemphasizing the reasons.

Moreover, as we have seen with practical identity, sometimes a person has duties that come into apparent tension. They have good reasons to lie in one aspect of their life (perhaps they are employed as a spy), and good reasons not to lie in other aspects (they are also a spouse). This tension need not be destructive. One isn't (usually) simultaneously acting as spy and spouse; the roles and duties don't (usually) overlap or collapse into each other. Good moral agents compartmentalize, meaning that they recognize when they are called upon to perform different duties and don't conflate them. Only an irrational, confused individual would act the spy at home and the spouse at work. There is nothing wrong with the maxim, "I will lie when it will prevent terrible harm." Nor is there anything wrong, prima facie, with the prohibition against lying as a betrayal of trust. Yet if maxims display a monotonic logic, they are incompatible. This view is simply at odds with the aspiration of autonomy, rationality, and self-legislation.

The task taken up by the next chapter of this thesis is to examine different non-monotonic systems that could work for Kantian maxims. In the meantime, I hope I have built sufficient intuition that merely the existence of maxims does not contradict the particularist project. Just the opposite—maxims are in-built to give agents increased sensitivity to reasons.

Kant offers several formulations of the CI. The different Formulae are meant to be equivalent in terms of the actions they necessitate, though they may express different content. The first one, the Formula of Universalizability, says, "Act only according to that maxim whereby you can at the same time will that it should become a universal law." The second one, the Formula of Humanity, says, "Act in such a way that you treat humanity, whether in your own person or in the person of any other, never merely as a means to an end, but always at the same time as an end." These two formulations will be my primary focus in this chapter. Another formulation, the Formula of the Kingdom of Ends, says, "Act according to maxims of a universally legislating member of a merely possible kingdom of ends." It will only come up in my third chapter.

The Formula of Universalizability has to do with contradiction. The Formula does not require that maxims be literally universal to hold. Each will wills autonomously. But the

Formula, following the CI's lead, compels maxims to take a certain form, and this is the form of non-contradiction. Maxims must be, at least in theory, amenable to universalization without resulting in contradiction. Kant says:

> Some actions are of such a nature that their maxim cannot even be *thought* as a universal law of nature without contradiction, far from it being possible that one could will that it should be such. In others this internal impossibility is not found, though it is still impossible to *will* that their maxim should be raised to the universality of a law of nature, because such a will would contradict itself.[21]

In this passage, Kant suggests two types of contradictions. The first type is "contradiction in conception." A maxim that demands two logically contradictory actions, X and not-X, would fail this test. This test has more limited application, so I pass over it. The second type of contradiction, much more salient to everyday reasoning, is "contradiction in the will." This type of contradiction occurs when a maxim would undermine itself if universalized over every will. The "contradiction in the will" test also has several contentious interpretations, but, following Korsgaard, I will mainly discuss the "practical" interpretation, on which a maxim contradicts itself if its universalization would in some way act against the purpose expressed in the maxim. It imagines, what if everyone did the same action for the same reasons as found in my maxim? Would the reasons, which point to ends or purposes for which I act, be undermined? One might object, for instance, to Jack Bauer's maxim for torture, which includes reasons of patriotism and defense-of-country, if it were demonstrated that widespread torture incited much more violence and spurred terrorist recruitment. As Korsgaard says, a person acting in accordance with a conflicted maxim is acting "as if she both did and didn't will the end."

The Formula of Universalizability engenders wide and strict duties. Wide duties have to do with an agent's virtue, counseling us to adopt good reasons for acting. Practicing beneficence (as a reason from which one acts) is an ethical duty because if beneficence were universalized, then the tide would lift all ships. Receiving beneficence empowers everyone to act in accordance with their own wills, not limited by scarcity of means or deprivation of character. It is hard to be an ethical consumer. I would like to own an electric vehicle, but they're costly. If actors in the government did a better job subsidizing alternative energy (enjoined by a maxim of improving lives by avoiding the worst impacts of climate change), then perhaps my *own* will, my *own* maxim of beneficence, could take freer reign over my life, and I could buy a Tesla to reduce my personal carbon footprint. Wide duties, then, range over the reasons one has for acting.

Strict duties are meant to be necessary and universal, often prohibiting actions rather than necessitating them. They overlap with Kant's doctrine of rights. The distinction between wide duties and strict duties is one I intend to sidestep as much as I can. This isn't just because wide duties are easier to accord with moral particularism, though they are. It is because no one has

---

[21] *Groundwork*, 4:424.

given a cogent account of strict duties. Strict duties, like wide ones, can conflict.[22] So what is the substantive difference between strict and wide duties? I consider it as a matter of degree, thinking of duties in the light of practical identity. Violating strict duties is destructive to practical identity. Strict duties usually accompany betrayals of one's rational nature: acts of deception, breaking promises, exiting from norms and institutions—in other words, tearing down the web of relationships that make someone who they are. It isn't (for me, the particularist) inconceivable that reasons would obligate someone to violate a strict duty, but such an action is never taken lightly. It always imposes a cost, accompanied by powerful regret, which I will discuss when I look more closely at moral dilemmas.

The Second Formulation, The Formula of Humanity, does not supply a test like the Formula of Universalizability. Or rather, the test it supplies sets a low bar. It obligates moral agents to not use other people *merely* as means. It is rare, though, that someone is *just* a means to an end. Even in exploitative relationships, the exploited usually benefit in some way, though not proportionately or justly. Factory labor sometimes offers much-needed employment to impoverished workers. Still—factory labor doesn't meet the spirit of the Second Formulation, which speaks to the sort of aims one should have when one constructs maxims. One should aim to promote humanity as an end. For the corporate boss, the end of factory labor is profit. Perhaps employees aren't *mere* means, but they are *mostly* means, with their value at the periphery of the boss's consciousness. The Second Formulation instructs the boss to change her maxim. The boss must act in accordance with a maxim that recognizes the intrinsic value of the employees' lives. One might look to Costco as an example of a company whose leadership advocated raising the minimum wage and doled out raises during the 2009 recession.

The Second Formulation imposes a goal that all agents should strive for, but the goal isn't an "ultimate end" in the teleological sense. Teleological claims nauseate most moral particularists, who think that humans are obligated to act for all sorts of different ends, depending on circumstances. Sometimes what is right isn't what promotes flourishing or whatever *telos* for humans is proposed. The Kantian claim that each person ought to be treated as an end isn't a *telos*. It just falls out of the theory of normativity. It is just in virtue of rationality that we value anything, and so (goes the argument) rationality deserves to be valued in itself. Our reasons must *tilt* towards humankind, but only because humans are rational. The point is almost tautological: reasons should value rationality; that is their normative nature. As Korsgaard says, "moral conduct is rational conduct."[23] But *how* we value rationality may depend heavily on circumstances. Sometimes, to value someone as an end, we must also cause them some small harm or even deceive them.

The Formulation also calls upon moral agents to treat *humanity* as an end. Not individual humans, but *humanity*. There are at least two possible readings here. On one reading, humanity

---

[22] See Hruschka, 1991.
[23] Christine Korsgaard, "Kant's Formula of Universal Law," 4.

refers to the sort of exercise and experience of being human. Individual people may possess humanity as an attribute; brutal dictators might be called *inhuman*. Specifically, because dictators do not act in accordance with the CI, they are irrational; they are inhuman just insofar as they are irrational. On the other reading, *humanity* sweeps up all of humankind, the species, the civilization, the community—some collective project. I am undecided between the two readings. What I would emphasize, though, is that on either reading moral agents can sometimes treat individuals merely as ends, so long as they never treat humanity merely as an end. This justifies killing Hitler to prevent the Holocaust, either because Hitler lacked the attribute of humanity or because killing him respected our collective humanity as an end, even if it partially treated humanity (qua Hitler) as a means.

Contrary to popular belief, the CI hardly lays out a rigid plan for moral agents to follow. By its own lights, the CI *can't* determine every action beforehand. If it did, then there would be no maxims left to write. It is therefore not desirable for Kantianism to achieve perfect epistemic completeness, which occurs when a theory can tell its token moral agent what to do in every circumstance. Such a theory would be tyrannical. It would leave no space for the agent to *choose* what to do; instead they would just consult the rulebook. By definition, such a theory can't be moral. Such a theory doesn't allow space for agency, for freely choosing to act according to reasons enjoined by a properly formulated maxim. Such a theory, in short, refutes itself. Kantianism preserves moral agency. The Formulae are not a two-step program for perfect maxims. They impose primarily formal requirements, with only the barest hint of what reasons one should act from. In the disoriented space between the Formulae and human action, the will designs its own principles.

II. Kant's account of reasons

Other than universalizability, the main concern about the possibility of Kantian moral particularism is apparent disagreement over reasons. For particularists, reasons play the starring role. The argument for moral particularism falls out of the theory of reason holism. More to the point, particularists build the notion of obligation out of reasons. On the particularist view, a moral agent has reasons for acting: reasons favoring and disfavoring the action under consideration, as well as features of the situation strengthening or weakening, enabling or disabling the favorers and disfavorers. What the agent has a duty to do is what the reasons favor overall. One imagines a sort of vector calculus of reasons, with reasons pointing the agent in different directions but summing up to a single action.

Kant seems to start with duty and laws, and to not talk about reasons at all. A moral agent has an obligation to act in accordance with the Categorical Imperative. On this picture, what matters for the agent is a law drawn from pure reason, which would apparently rule out the context-sensitivity that particularists emphasize. A philosopher can wedge reasons into Kant's moral structure by looking to maxims, which must include reasons for acting. But, as Scanlon

says, “the idea of a reason and of the strength of a reason have at most a derivative role.”[24] Dancy similarly claims that it would be expected for Kantians to be “suspicious of the very notion of a contributory reason.”[25]

I see two possible tensions between Kantians and particularists over reasons. They can both be handled by closer attention to the composition of maxims. Textual evidence will show that Kant *did* talk about reasons, just in different terms. The two tensions are:

i. The supremacy of reasons. Particularists start with reasons and build up to obligations. Kantians work from the top down, with reasons only lurking in the background.
ii. The behavior of reasons. Particularists claim that reasons interlock as contributors, modifiers, and conditions, and that they also have different strengths. Partly in virtue of working from the top down, Kantians seem to flatten reasons out, not accounting for differing strengths and roles.

Supremacy

The particularist worries that Kantian ethics cannot recognize the holism of reasons because Kant does not talk about reasons. This initial worry is a little thin, though. If Kant didn’t talk about reasons, then this might create space to introduce reasons to the Kantian system. So let’s get more concrete about how Kant’s approach might rule out a late turn to reasons. Perhaps the real worry is that Kant implicitly defends a theory of practical reasoning in which reasons themselves do not strongly feature. If this is so, then perhaps the Kantian moral agent can’t “sum” reasons as the particularist would have it. Instead, the Kantian just looks to laws. It’s true that the Kantian agent’s *obligation* comes from laws. The properly formulated maxim, the principle of willing, forces the agent to act. But the Kantian agent still acts for *reasons*. Evidence appears in maxims themselves, which must include reasons. The main difference from the particularist’s picture is that, for the Kantian, reasons are combined in these obligating principles of willing, whereas, for the particularist, combinations of reasons directly produce an overall obligation for the agent. The Kantian can’t really adopt the particularist’s picture of combinations of reasons obligating agents because then maxims become redundant to the picture, an antiquated bit of machinery in the modern era of reasons.

But it would make little sense for maxims to be the whole story. Maxims must be well-formulated to accord with the Categorical Imperative. To be well-formulated, maxims must include at least an action and a reason (or reasons) for acting. What reason (or reasons) the maxims include matters. The reasons must, according to the Formula of Humanity, respect humans as ends. They must also not enjoin an action that thwarts its own purpose. Reasons aren’t just window-dressing. The complaint that Kant ignored reasons is not wholly fair.

---

[24] T. M. Scanlon, “How I am not a Kantian.”
[25] *Ethics without Principles*, 67.

Historical context will bolster the intuition that Kant can deal with the particularist's challenge. As Jens Timmermann points out, Kant had Alexander Baumgarten in mind when he denied the possibility of conflicts of duty: "In the *Initia*, Baumgarten unashamedly speaks of obligations that differ with regard to their normative force. Depending on the strength of the *causa impulsiva* that generates it, there are major and minor, stronger and weaker obligations (S16). There are, accordingly, 'collisions' of obligations."[26] In *Metaphysics of Morals*, Kant rejects Baumgarten's thesis, arguing that "since duty and obligation are concepts that express the objective practical *necessity* of certain action, and two rules opposed to each other cannot both be necessary at the same time – rather if it is one's duty to act according to one of them, to act according to the opposite one is not only no duty, but even contrary to duty – a *collision* of *duties* and obligations is not even conceivable…"[27]

That Kant responded to Baumgarten indicates that Kant was well aware of the position that reasons had to be summed to find the overall. Kant denies this picture, because Baumgarten says that each *causa* generates an obligation, but obligation implies necessity, and two mutually exclusive actions can't be simultaneously necessary, as would happen when *causae* conflict.[28] But what Kant denies is only conflicts at the overall level. Kant rejects that obligation can be anything other than overall—but this doesn't mean he must reject the "causes" underlying obligation. The particularists actually do something similar. They would deny that an individual reason obligates an agent when other reasons must be considered. Instead, reasons form a "resultance base" from which results the overall obligation.

Kant had an equivalent to Baumgarten's "causa impulsiva." In the same passage from *Metaphysics of Morals* quoted above, Kant continues, "[T]wo *grounds* of obligation (*rationes obligandi*), one or the other of which does not suffice to obligate (*rationes obligandi non obligantes*), can be combined in a subject and the rule he prescribes to himself, but in that case one of them is not his duty." Kant introduces and defends the concept of *rationes obligandi*, glossed by Jens Timmermann as *grounds* of obligation. "Rationes" is also the Latin word for *reasons*.

The first part of this quote will please the particularist; the second part will cause confusion. First, Kant suggests that grounds of obligation can be *summed* to give the moral agent an overall obligation. This is a curious sort of statement. One might think that if X were a ground of obligation, then by definition it would obligate by itself; otherwise it wouldn't be a ground of obligation, but just a free-floating fact. Kant could mean either of two things. On the one hand,

---

[26] Jens Timmermann, "Kantian Dilemmas?," 39.
[27] Immanuel Kant, *Metaphysics of Morals*, VI 224.11-18.
[28] He says that two conflicting rules cannot both be necessary *at the same time*. This seems to suggest defeasible moral rules that hold only sometimes but not always. Timmermann remarks in a footnote that the German passage is not obviously about a temporal notion of 'at the same time' simultaneity, that perhaps the passage is about concrete obligations not holding at the same time, rather than abstract laws; but defeasible laws would fit too. More on this later.

perhaps he points to the way that facts can enable or disable, strengthen or weaken the contributory reasons. Just because one promised to take the dog for a walk on Tuesday doesn't immediately obligate one to walk the dog; it also *has to be Tuesday*. This is one way for two grounds, which on their own do not obligate, to obligate when combined. The other thing Kant could mean is that sometimes reasons seem to exert a minor moral pull on an agent, but the reasons aren't so tyrannical to obligate the agent if they have conflicting practical, non-moral reasons. Perhaps you really ought to hold the door open for someone, but if you are in a rush, you aren't *obligated* to do so. However, if the person has a disability and would have serious trouble opening the door by themselves, then you are obligated. This could also be what Kant means when he says that sometimes two grounds must be combined to obligate an agent.

On either interpretation, the quote makes clear that reasons (grounds of obligation) form unities and that one sometimes acts for the unities, not for any single reason. The particularist should welcome this move by Kant. It creates an aperture in Kant's theoretical edifice for facts acting as modifiers and conditions, facts that form a sort of "resultance base," to use Dancy's terminology. Dancy himself does not have a clear idea for what to say about defeated reasons for doing competing actions.[29] Does each favorer for an action automatically count as a disfavorer for any competing action? Would the absence of a more pressing duty count as a favorer for one's duty? Perhaps Kant offers a tidy solution. Maxims define what an agent values. They reveal the agent's rational process. Whenever one constructs a maxim, one implicitly rejects the competing alternatives and recognizes the absence of more pressing duties. The absence of competing duties fits here in a way that it doesn't fit the notion of a "resultance base." But not much hangs on this.

The second part of the quote threatens to undo this leap forward. It says that when two grounds of obligation are combined to obligate an agent, one of the grounds is not the agent's duty. On one reading, Kant means that the *other* ground *is* the agent's duty. But this doesn't make sense. Kant has just admitted that neither ground by itself is enough to obligate the agent. What would it mean for one ground to *be the duty* but to not obligate? I think, perhaps, that Kant just intends to say that no single ground is the duty. Or at least that is what I am going with.

Behavior

I have argued that when the particularist worries that Kant ignores reasons, they are being ungenerous. Kant had a theory of reasons based on *rationes obligandi*, and though they don't obligate by themselves, they generate obligation when placed in the context of a maxim properly formulated for the Categorical Imperative. But if reasons don't obligate by themselves, how are they reasons at all? Reasons are supposed to be favorers and disfavorers. Moreover, they are supposed to be subject to intensifiers and attenuators, enablers and disablers. Or at least this is the particularist's picture.

---

[29] *Ethics without Principles*, 93.

Severing the connection between reasons and obligation is not so bizarre, so long as reasons keep their normative force, which seems inbuilt to the notion of a reason. Timmermann offers this pithy argument: "'Ought implies 'can,' but 'reason' does not."[30] Whenever one has an obligation to do X, it means that one can do X. If one couldn't do X, then it wouldn't be a necessary duty, for what is necessary is also possible. But one can have a reason to do something impossible. I would love to teleport to class—I would less often be late. This is a good reason to teleport. But teleportation is impossible. A crucial difference emerges between "favoring" (which Kantian reasons do) and "obligating" (which they don't).

That said, four types of reasons-behavior require further explanation by the Kantian: strengths of reasons, modifiers, conditions, and switched valences. To start, the concept of relative strengths of reasons is not alien to the Kantian. Kant sharply distinguished between strict and wide duties, with the strict supposedly always outweighing the wide. In one example, he says that even if someone is dying, one can't go into debt to save them because owing money is strictly prohibited. Obviously I would deny this. However, the strict vs. wide distinction shows that obligating grounds can have differing strengths. Some grounds are built into maxims of strict duty, seemingly absolute, while others are built into maxims of wide duty, which have (more) exceptions.

In the passage from *Metaphysics of Morals* I have been quoting, Kant continues, "When two such grounds conflict, practical philosophy does not say that the stronger obligation retains the upper hand (*fortior obligatio vincit*), but the stronger *ground of obligation* retains the field (*fortior obligandi ratio vincit*)." This sentence reinforces the distinction between grounds and obligations. It shows two more things as well. First, it reveals that grounds of obligation, like reasons, can pull an agent in different directions. The sentence also shows that reasons can be fully outweighed. One ground can vanquish another, push it off the field of reasons entirely. This is presumably what happens when one reason outweighs and is mutually exclusive with another (as opposed to Kant's metaphor earlier in the passage of two grounds "combined in a subject and the rule he prescribes to himself").

The particularist wants to say that even if a reason is outweighed, it still retains some normative force. Following one of Kant's examples, a person has reason to give money to their benefactor and reason to help their impoverished parents, but can't do both. Still, both present valid reasons. Kant captures this notion—the lingering normative force of defeated reasons—with grounds of obligation, even if he will not allow the grounds to obligate in themselves. A passage on moral conflicts from the notes of Vigilantius, one of Kant's lecture students, is helpful here.

> Now one can never say in such cases that it is absolutely impossible to fulfil both duties; and these duties remain even if they are not fulfilled because, as was said above, laws and rules cannot

[30] Timmermann, 57.

> contradict each other, but there is an opposition of the grounds of one duty against those of the other…[31]

Duties "remain" even if the agent can't do both, though the agent only has an obligation to do one. At first, it appears that obligation and duty split apart. But this can't be. More likely, Vigilantius means to note that there are still grounds for duty, presenting a sort of token-duty that would press the agent into action if not for the mutually exclusive, stronger ground. It is only due to contingent circumstances that one is the duty and not the other; if the person had adequate funds, *both* would be duties. Timmermann agrees, "if I were richer than in fact I am, the very same grounds of obligation—to help my benefactor, and to help my parents, both valid in their own right—would not conflict with each other. Neither would be defeated… Grounds of obligation depend on the precarious availability of means to generate duties, but unlike those duties do not disappear when they cannot be acted upon."[32]

Thus, the Kantian effectively captures different strengths of reasons and reasons pulling a moral agent in different directions. This makes everything that follows much simpler. First, it clarifies the notion of enablers and disablers. The presence of an enabler allows a contributory reason to exert a claim on an agent. In the example we have been working with, my wealth is an enabler (or disabler, as the case may be). If I have just enough money to help my parents but not also to help my benefactor, then helping my parents gets enabled as a duty. If I have money to do both, both are enabled; if I have too little money to do either, both are disabled. This is a very clear illustration of how "ought" requires "can," but reasons don't.

When a reason switches valence, this is because the ground of obligation is in the presence of some other feature that affects its operation. For instance, if my parents have been horrible to me, then perhaps gratitude is not what they deserve. It could be a wrong-maker, simply inappropriate to the situation. Kant has already stated that sometimes a moral agent combines two grounds of obligation in a single maxim.

Intensifiers and attenuators work similarly. There is a reason why Kant contrasts helping one's benefactor to helping one's parents in the example. Both might be thought of as belonging to the same type or class of moral actions, those of (in Rossian terms) gratitude. But one case presents the stronger ground for obligation. There is no commitment, in Kant's moral universe, that makes this suggestion untenable now that we have *rationis obligandi* as a resource. In fact, I think it makes a great deal of sense in light of the concept of practical identity. As I have elucidated, one's practical identity embeds one in a web of duties that depend on the roles one plays. My role as a son to my parents is certainly crucial to my practical identity, more crucial than my role as a grateful beneficiary to a benefactor. We see that sacrificing one's role as a son would be far more damaging than sacrificing one's role as a beneficiary.

---

[31] Cited in Timmermann, 56.
[32] Ibid., 57.

Perhaps the particularist's real concern is that there's no way to figure out which reasons are stronger or weaker (or which parts of one's practical identity are more or less integral). In fact, Kant has more to say here than most particularists. Particularists often rely on *intuiting* which reason outweighs which. The Kantians can at least look to the Formula of Humanity and see respecting humankind as an overarching goal, shaping all calculation of reasons. At least the Kantian particularist is no worse off than the intuitionist particularist.

Defeasible Moral Reasoning

I. Stating moral principles

In the previous chapter, I argued that Kant's moral system reveres above all else the autonomy of moral agents, the self-legislating will, and the creative process of forming an identity. Insofar as the will is rational, it wills according to self-imposed principles. These principles, called maxims, express connections between reasons and actions. Their moral worthiness depends on properly formulating the connections; poor maxims reason poorly. Therefore, in formulating the connection, maxims need to attend to all the complexities of reasons that particularists emphasize, like differing strengths, flipped valences, modifiers and conditions. Maxims also display a non-monotonic logic. One can hold two maxims that apparently counsel different actions (e.g. one enjoining lies, one prohibiting lies), so long as they are formulated with different reasons. An agent can't have two maxims that include the exact same reasons but different actions, as this would simply demonstrate irrationality. In this chapter, I will attempt to get much more specific about the defeasible logic that maxims display.

Before getting to defeasibility, it is worth clarifying what a maxim *is*. I start with the paradigm view that maxims are moral principles. But they are special moral principles, and philosophers disagree about what moral principles are, too. On some other views, moral principles are epistemic objects that moral agents carry around like congressmen carry around pocket Constitutions; they pull them out when the situation calls for it. An example might be the Golden Rule, given to Nazareth in the Torah: Do to others what you want them to do to you. The agent summons the principle up in considering an action and uses it as a benchmark for judging whether the action is right or wrong. McKeever and Ridge calls moral principles of this type "guides."[33]

On another view, moral principles express universal facts that *make* actions right or wrong. Gideon Rosen proposes that a moral law has the form $\mathbb{N}\forall x(\phi(x)\rightarrow Fx)$, which from left to right reads as "It is normatively grounded that for all x, if x ϕs, then x is F."[34] Rosen believes that moral facts are grounded in the features of an action conjoined with a law relating those features to normative properties. On this view, the laws "*make it the case*" that actions are right or wrong, in the way that a traffic law makes it the case that speeding is illegal. A driver looks to traffic signs to know the speed limit, so laws loosely serve an epistemic role, but the laws also make speeding illegal. McKeever and Ridge call these principles "moral truth-makers."[35]

Neither view of principles is one Kant can fully countenance, but both illuminate the nature of maxims. Maxims are epistemic objects insofar as they are freely chosen. A motif in

---

[33] Sean McKeever and Michael Ridge, *Principled Ethics*, 8.
[34] Gideon Rosen, "What is a Moral Law?", 147.
[35] *Principled Ethics*, 12.

Kant's *Groundwork* is that agents "represent" laws to themselves.[36] One *decides* to put down the gun and offer peace; one *chooses* to help the old lady in the street. Just as Nazareth represents to himself the Golden Rule, Kantian agents represent to themselves maxims; they aren't hidden in the underlying machinery of normativity but rather put front and center.

On the other hand, maxims, like Rosen's moral laws, give actions their normativity. A right action is one enjoined by a properly formulated maxim. Perhaps moral valences are at least partially *grounded* in maxims, though I won't explore the idea here. But maxims aren't facts *out there*, as Rosen would have it. They aren't like the law that everything is identical to itself, though they may be just as necessary. They are internal to each self-legislating agent.

The combination of these two perspectives on principles—that they both represent laws to agents in an epistemic role *and* make actions right or wrong in a metaphysical role—together gets at what it means for an action to be enjoined by a maxim. This is the sense in which maxims are moral principles.

A moral principle, let's say, is a statement of the form "what ϕs is F," where F is a moral property. I remain silent on the content of ϕ: it could be anything, since anything could be a reason for acting. Moral principles can take a non-normative property or *other* normative properties as ϕs. "Good acts are right" is a moral principle—a crucially important one in the history of philosophy—that takes one normative property (goodness) to another (rightness). The principle that "an honest act is right" also takes a partly normative property to rightness. "Stealing from charities is wrong" takes a non-normative property to a normative one.

The notion of a moral property ("F") requires clarification. Here are the properties I have in mind: right, wrong, rightmaking, and wrongmaking. "Rightmaking" refers to features that act as contributory reasons favoring actions. The principle "honesty is rightmaking" says that actions that tell the truth will be morally more favored in virtue of that feature (though the actions could still be overall wrong in virtue of a different feature). I include rightmaking and wrongmaking because they, like right and wrong, are *thin* normative properties: they don't describe. Allowing principles that merely regulate how features contribute (rather than regulating overall rightness or wrongness) also opens the door for principles like, "You shouldn't lie." These principles do not necessarily imply that all acts involving a lie are wrong; they merely assert that lying will always be a wrong-making feature of all those actions.

Maxims meet this format because ϕs can be reasons for action, and, by expressing whether the agent will or will not take an action, maxims indicate F, whether the action is right or wrong. I schematized maxims as {A, T, R}, for {action, trigger, reason(s)}. Take the maxim, "When up for a raise, I shall not exaggerate my accomplishments, because doing so is unfair to my colleagues." This maxim could just be expressed as, "Whatever involves exaggerating my accomplishments for a raise is wrong on account of its unfairness." That last bit—"on account of its unfairness"—often does not appear in non-maxim moral principles, but moral principles vary

---

[36] *Groundwork*, 4:401.

widely. Many principles do not have the "what ϕs is F" form, though we still think of them as moral principles. "Do not kill" comes to mind. One can rewrite this statement into "whatever actions involve killing are wrong," and most principles will be susceptible to rewriting in this way. The arguments about moral principles that follow will apply to moral principles of other syntactic forms as well.

Moral principles have a suppressed universal quantifier. The principle "killing is wrong" covers every act of killing; it means "all acts of killing are wrong." If something is forbidden, it is implicitly forbidden *always*; the concept of forbiddenness already has universal quantification folded into it.[37] Against this one might think that principles are instances of indefinite statements, like "birds fly." "Birds" is a *bare plural*, a noun without quantification, just like "honest actions." And if a speaker asserts that birds fly, they do not intend to say that all birds fly; they only describe birds categorically, perhaps describing what the prototypical bird does. But I do not think this is plausible for moral principles. When I say "honest acts are right," I am not just describing the category of honest acts; I am *prescribing* a principle that one ought to act in accordance with. It follows from "honest acts are right" that you should do honest acts, but it hardly makes sense that birds *should* fly: are penguins defective? Rightness just differs from flying. Ought-to-be-doneness is not indefinite, though the syntax makes it look so.

This goes for maxims as well, and I think it explains much of the confusion surrounding them. As I argued in the previous chapter, it's often thought that maxims rigidly determine an agent's moral life, but instead they obey a defeasible logic. However, to be defeasible, maxims must also have universal quantification. If they were simply indefinite, then they would never be *defeated*, just as "birds fly" is not defeated by the existence of penguins. But maxims *are* defeated. They are defeated when one must act on one duty at the cost of another. Strictly speaking, one can't have two conflicting duties, because the necessity of the stronger one destroys the weaker. But in this case the weaker "duty" is defeated by way of its grounds of obligation never producing an actualized obligation.

This happens *all the time*, which is why our practical identities are so complicated (and why we can often feel moral regret). Suppose that two maxims, A and B, take a different stance on an action based on different reasons. A prohibits stealing for trivial reasons and B obligates stealing to feed one's family if they would otherwise starve. Someone has the option to grab a bag of goodies that holds bread, meat, and a Rolex watch. In this case, one maxim will defeat the other. If the person's family is starving, they should grab the bag, because saving one's family is more integral to practical identity than not being a thief. One hasn't done wrong in this case

---

[37] Wrongness is more straightforward than rightness in this regard. Actions which are overall wrong are *prohibited*; you ought not to do them. Actions which are overall right may either be *permissible* or *obligatory*. It is overall right that I wave at the mailman because it brightens his day. But do I have to do it? This is a puzzle for a different essay. The indeterminacy around rightness does not bear on my argument. Whether rightness expresses permissibility or obligation, it *universally* expresses this property in a given principle: "such-and-such action is always permissible"… "such-and-such action is always required."

because one duty destroys the other. However, the *ground* of obligation (the presence of the watch) is still there, and so one might still feel lingering regret over the situation.

Readers of Kant have an unfortunate tendency to conflate monotonicity with universal quantification. They perceive that maxims are universally quantified, that there are no in-built exceptions. But then they take this to mean that maxims are never defeated. This simply isn't so.

II. Do moral principles tell the truth?

The rest of this chapter discusses different ways to understand defeasible moral principles. First, I will develop my own view, which is that moral principles are *explanatory*. Maxims explain one's will to oneself. They organize sets of reasons in relation to actions, and insofar as one is rational, they are binding. Maxims' defeasibility comes from the fact that explanations are non-monotonic; each maxim only includes a finite set of reasons, and new ones could change the equation. Explanation is the appropriate way to understand maxims because they serve both the epistemic role of standards *and* the metaphysical role of, say, scientific laws. Scientific laws explain and underlie the natural phenomena we observe, but their scope, too, is limited. To develop this view, I call on Nancy Cartwright's philosophy of modern physics, a discipline that seems to require scope-limited laws. One peculiar consequence of my view is that moral principles come out *false*, because counter-examples (outside the scope of the laws) falsify them. Yet they still explain in a local way. This turns out to be the most elegant way to cash out defeasibility, as I will argue in the last two sections of this chapter.

Cartwright argues in *How the Laws of Physics Lie* (1983) and *The Dappled World* (1999) against the view that natural laws in physics are both true and explanatory. She says about physics, "if the theory is to have considerable explanatory power, most of its fundamental claims will not state truths…this will in general include the bulk of our most highly prized laws and equations."[38] This is because of the inverse relationship between the truth of a law and its scope. This paradigm, which I term the Paradigm of Anti-Facticity (PAF), can be stated as follows:

> PAF: The more states of affairs a law covers, the less likely it is to be true. The fewer states of affairs a law covers, the more likely it is to be true.

Cartwright argues that F=M*A is false in certain circumstances. She says that Newton's laws do not explain very large systems, where relativity takes over, nor very small systems, the domain of quantum physics. Newton's laws only hold when restricted to a scope. They hold in high school physics problems when the student is instructed to ignore friction, minor relativistic effects, fluid dynamics, etc. If those factors get stirred back in, the laws cease to hold, strictly

---

[38] Nancy Cartwright, *How the Laws of Physics Lie*, 78.

speaking. Laws hold, generally, in the models they are meant to explain. A model is a constructed system of finite scope.

Some laws of physics may or may not hold—hence the word "likely" in PAF. Quantum mechanics may or may not hold around black holes. We just don't know; their effects are too slight. Meanwhile, at the level of elementary particles, we usually do not find Newton's laws; only electromagnetism, strong nuclear and weak nuclear forces. These different domains are not *natural* per se. They are observed. When we note phenomena, we create models for them with laws that explain them. Inferring from one model to another is unwise. Cartwright therefore defends a "patchwork" view of physical laws. She resists the temptation to infer from one degree of granularity to another. Each represents a different domain governed by its own laws, written just to explain the phenomena observed there. The "truthiness" of laws becomes suspect outside limited circumstances.

I hold approximately the same view about Kantian maxims. A person, Kant says, is an "organized being."[39] When humans lifted their gaze to the stars, they constructed a model to fit observed phenomena and wrote Kepler's laws to explain them. Likewise for moral phenomena. Each maxim is an instance of the rational human organizing reasons with respect to actions. Maxims are the explanatory laws of the will; they say how the purely rational will responds to certain reasons. Maxims have a representational quality, too; as Kant says, it is "the representation of the law in itself" which is "the determining ground of the will."[40] The moral field is complicated, and maxims are a way to divvy it up, to represent aspects of it to oneself in the form of reasons and actions. Newly introduced reasons may change the appropriate direction of the will, which is why maxims, like Cartwright's natural laws, are defeasible. Each maxim only accounts for a finite number of reasons, but as reasons can be weakened, flipped, disabled, outweighed, etc., the maxims are, strictly speaking, false.

Where I differ from Cartwright is in her quasi-realist pragmatism. Cartwright's theory implies that natural laws get constructed in a sort of ad hoc way. They are written to fit empirical evidence, rather than the other way around: empirical evidence fitting predetermined laws. Kant's project must remain primarily a priori, such that maxims are *real* in a deeper sense than natural laws. The precise metaphysical distinction here doesn't matter much to me. What I take from Cartwright is the view that defeasible laws (of any variety) can be false yet explanatory. This allows Cartwright and myself to make sense of the lasting usefulness of laws with exceptions, and at the same time to develop a fitting model of explanation.

Two things may be said against Cartwright's view. The first is simply that human intuition recoils at a "patchwork"; the far more appealing picture portrays everything as governed by one set of laws—most reasonably the laws of quantum mechanics. That picture is not supported by empirical evidence. There isn't a mechanism for particle physics to explain

---

[39] *Groundwork*, 4:395.
[40] Ibid., 4:401.

relativistic effects. We have yet to prove a "theory of everything." A second objection is that F=M*A should not be taken literally. The law, *prima facie*, holds for all bodies in space and time and describes an overall force, but perhaps natural laws are meant to be taken in a more limited, pragmatic way. This is rather like the argument for moral principles as indefinite statements and earns the same response. There are no exceptions written into these laws and they aren't hedged. Reading the law as true would also be unnecessarily generous. Even without a theory of everything, Newton's laws play an indispensable role in human life, guiding everything from building bridges to playing sports.

Cartwright's views are as controversial in philosophy of science as moral particularism is in ethics. The two positions have a lot in common. Both hold that the laws that are supposed to govern their fields are exception-riddled. Both strive to develop a theory of explanation that doesn't rely on such laws. For Cartwright, this is the "simulacrum" account of explanation, a sharp turn away from Hempel's Deductive-Nomological model. For moral particularists, moral explanation likewise can't be a process of deducing facts from laws, so particularists like Dancy rely on a broad notion of unprincipled moral judgment. Moral explanation is the subject of the next chapter. Here, I will argue that the parallels between Cartwright's and Dancy's views on laws illuminate a new direction for moral particularism.

Previous particularists have taken principles to be either hedged and truth-telling, or non-explanatory and not truth-telling. Cartwright argues that physical laws are not hedged (they are mostly false) yet nonetheless important in the explaining physical phenomena. Moral principles should be thought of in the same way. If moral principles are hedged, then they tell the truth. A hedged principle has a clause, suppressed or explicit, that admits the failure of the principle in certain circumstances. One way to hedge is with a *ceteris paribus* (CP) clause: "other things being equal" and other phrases of this sense. Whenever a CP principle fails to obtain (e.g. when an honest act goes wrong, given the CP principle that honest acts are right), the failure gets swallowed by the hedged nature of the CP principle. Counter-examples don't falsify hedged generalizations. Unhedged moral principles, on the other hand, are falsified by their many counter-examples. Assuming universal quantification, the many times lying has been the right thing to do falsify the principle that "honest acts are right."

Moral principles, maxims especially, should not be read with a suppressed *ceteris paribus* clause in this way. Maxims are principles of willing that one represents to oneself. When so represented, they do not include a CP clause. If they did, they would lose their forcefulness. How would one know if the ceteris *were* paribus? Without further specification of the "other things," the CP clause is useless. With further specification, the CP clause becomes redundant to the principle.

We must reckon with a degree of ineluctable uncertainty when dealing with false principles. One motivation for holding moral principles to be exceptionless is that it makes moral judgment watertight. That hardly means judgment becomes easy. It can be extraordinarily hard to apply an exceptionless principle. Take the Principle of Utility (PoU): that the right action is

the optimific one. Utilitarians face a complexity problem. Determining which actions produce the greatest total happiness in the long-term is a difficult empirical question. Nonetheless, utilitarians take comfort in knowing that the PoU affords them a certain watertightness. The right action can be deduced from the principle in conjunction with facts about the world. If utilitarians had enough information to deal with complexity, they could supply a mathematical proof that their actions were right. But when we recognize moral principles to be exception-laden, absolute certainty about moral facts goes away. One can never deduce from faulty principles.

But watertightness is overrated. Explanatory power is a far more useful property, and it competes directly with watertightness. Principles that explain a lot are very rarely watertight. Try to imagine a true, exceptionless principle. We have already listed many exceptions to the principle that "honest acts are right." What if we quarantined all those exceptions into a clause that makes the principle more specific? "Honest acts are right, so long as they do not harm others, do not break a promise, do not point the madman to the hidden gun, … etc." The list of exceptions would be very long, though adding this clause would restore to the principle the ability to support deduction. If one had all the salient information about an action, then one could go down the list of exceptions and, if none appeared in the case of the action, then the action would be right. This level of surety is seductive. It is appealing to think that if one had exactly the right principles and knew everything there was to know, one could never err. The alternative—a vision of stumbling aimlessly through the moral landscape, hoping against hope that one's actions are right—is repugnant.

That is not the alternative on the table. Just because our principles have exceptions, it does not mean our moral judgment is blind. Humans aren't calculators; they can judge non-algorithmically, or in algorithms that exceed the capacity of language to represent. James Wallace identifies the supposed alternative as "a false dilemma. It is based upon the notion that the only alternative to doing exactly what one is told by a rule is to act arbitrarily."[41] Suppose that all moral principles are false. Doesn't the principle that "honest acts are right" still have a hand in one's judgment that telling the truth in Nazi Germany would be wrong? The false principle guides one's judgment by partly explaining what is exceptional about Nazi Germany. In fact, one doesn't really know whether the principle is true or false except by performing moral judgment with the principle *in mind*. And it is the principle in its best-known form, "honest acts are right," that one has in mind, not a principle with a CP clause. Moral judgment not only tolerates false principles but depends on their supply. Without "honest acts are right," we would be in the woods.

This, roughly speaking, is the sense in which false principles can be explanatory. Principles range over a specific model. Natural laws are true for the scientific models they explain; e.g. the laws of special relativity range over models with masses large enough to generate non-negligible relativistic effects. A maxim ranges over a model of the will that

---

[41] James Wallace, *Moral Relevance and Moral Conflict*, 65.

features only the reasons specified in the maxim. In these circumstances, the maxim explains that a set of reasons necessarily obligates or prohibits an action. The maxim is locally true. It is a sort of idealization of moral agency, imagining one having a limited supply of reasons to deal with, imagining circumstances conducive to perfect rationality.

In other circumstances, when foreign reasons invade, the maxims may not hold; they may be defeated by another maxim. Still, they serve an explanatory purpose, somewhat like McKeever and Ridge's "guides." The moral agent recognizes that their maxims are locally true but globally false. Part of the challenge of moral judgment is utilizing false-yet-explanatory principles to navigate circumstances where one isn't sure which maxims apply.

The principles that fail to obtain in the most circumstances are often the most fundamental, like Ross's *prima facie* duties. That they are so often falsified just emphasizes their role in moral judgment. By contrast, the "truer" a principle is, the narrower it is. Cartwright says about physical laws, "a law that actually covered any specific case, without much change or correction, would be so specific that it would not be likely to work anywhere else."[42] Similarly, true moral principles would only apply in cases where it was obvious that no exceptions pertained. In cases of moral conflict, when one duty overwhelms another, the principle would not obtain, nor would it obtain when features pop up that could switch its moral valence. But it would be wrong to say that such circumstances make the principle *irrelevant*. It is exactly in these unusual circumstances that one needs the principle most. Where honesty is wrong, the principle that "honesty is right-making" is what must be, in some sense, explained *away*. In explaining an unusual moral phenomenon, one depends on the principles the phenomenon belies.

If moral principles do not tell the truth, then why the powerful intuition that they do? It is hard to admit that the principle "honest acts are right" is a lie. I am not suggesting non-cognitivism; just the opposite, I think that moral principles have truth conditions and often fail to meet them. But they feel rather "truthy," as if, as Dancy says, their failure to tell the truth always leaves something to explain. In the following section, I examine one view of why principles have this property. Lance and Little, two moral particularists, cash out moral principles as having a suppressed premise that limits the circumstances in which the principle is supposed to obtain. Even in exceptional circumstances, a lingering *trace* of the principle's true self remains.

III. Defeasible generalizations

Moral principles are defeasible generalizations. Principles have "defeat" conditions: cases where they do not obtain. The moral principle "honest acts are right" is defeated by the case in which telling the truth leads to terrible harm, e.g. pointing the madman to where the gun is

---

[42] *How the Laws of Physics Lie*, 112.

hidden.[43] The question at hand is how to cash out truth conditions for defeasible generalizations. Particularists agree that principles are full of holes. Does this make them *false*? I think so: any universally quantified statement must hold for all the objects in its domain, and moral principles don't. But others have tried to make moral principles true by interpreting them in unusual ways. In this section, I evaluate such an attempt by Lance and Little.

Explanatory power motivates Lance and Little to read moral principles in a manner that would make them true. Like me, Lance and Little acknowledge the holism of reasons but aren't prepared to renounce principles entirely. Admitting principles' falsehood would, they fear, render the principles explanatorily inert. Instead, Lance and Little argue that principles are a type of defeasible generalization with a suppressed at the front that says: "in privileged conditions." This is supposed to get principles that are "both genuinely explanatory *and* ineliminably exception-laden."[44] In two steps, Lance and Little permit generalizations to do explanatory work even when the generalization does not obtain, i.e. in non-privileged conditions. First, they connect defeasible generalizations to *kinds*. The generalization that "a match lights when struck" says something about the behavior of a *match*, a match-as-kind, not of a phosphorus-tipped stick. It is essential to a *match* that it lights when struck. Second, they develop the notion of normative privileged conditions for a phenomenon. When we declare that a match lights when struck, we think of a dry match in the presence of oxygen, not a sodden match in a vacuum. When a match is struck and doesn't light, this exception to the generalization must be occurring under exceptional conditions. Lance and Little say that this is subtly different from the notion of *defaults*, which I will discuss later.

Lance and Little's claim that moral principles are exception-laden is phrased to distance them from other theories of defeasible generalizations. One is that generalizations give a statistical account of a phenomena; e.g. "matches light when struck" just signifies that that's what matches do most of the time. But then the generalizations would be purely descriptive, not explanatory. There are also numerous counterexamples where a principle seems true but not in virtue of any statistical content. Lance and Little give the example of the generalization that fish eggs develop into fish. Most eggs actually get eaten or die or otherwise fail to develop; yet the principle still holds. A second theory is that generalizations are enthymematic, meaning that they contain suppressed premises like "ceteris paribus." Lance and Little reject this idea because it implies that exceptions to the generalization must be quarantined or expunged, instead of principles being "ineliminably exception-laden."

---

[43] Principles about *rightmaking* and *wrongmaking* (as opposed to right and wrong) are a little different. They are defeated when the feature under examination switches valence or stops contributing at all. That honesty is rightmaking is not defeated while honesty maintains a positive moral valence, even if the actions it enjoins are overall wrong. This is the case that W. D. Ross foresaw with prima facie duties. They can pull moral agents in different directions, with one duty outweighing another even if both duties carry consistent valences.
[44] Lance and Little, "Defeasibility," 299.

I agree with Lance and Little on the "fundamentally exception-laden" nature of principles and on the importance of respecting the existence of those exceptions.[45] As Michael Scriven expounds, being reasonable about defeasibility means "providing an anatomy where our predecessors thought only a pathology was required." But Lance and Little are misled by their wish to make principles true and this makes their "anatomy" defective. In my view, they make three mistakes. The first is to take moral principles as enthymematic, meaning that they have suppressed clauses. The second is basing natural kinds of objects on generalizations about them, which reverses the proper metaphysical picture. The third is to cash out principles as normatively dependent on teleology. The more elegant solution is to let moral principles be false. To capture reality in principles is akin to the task of the Danaïdes, condemned in Greek mythology to fill sieves with water; exceptions will always pour out.

The enthymematic

It is somewhat mysterious to me that Lance and Little argue against the enthymematic school, because their own approach belongs to it. Defeasible generalizations, they want to claim, have a suppressed "operator" that goes in front reading "In privileged conditions…" This operator restricts the scope of the generalization to normatively determined conditions. Although Lance and Little reject the "fundamental premise that real explanation happens only where exception no longer resides," they believe that exceptions are explained by *reference* to the privileged conditions. Explanations only function with a great deal of help from the exceptionless.

I simply find suppressed premises implausible. When we think about privileges, we do not think about a CP clause. Such clauses would have to be not only suppressed in the written word, but mentally suppressed. How, then, do we know that the clauses are there? Only by *ad hoc* analysis, observing that principles are usefully employed in moral judgment despite their exceptions. But this analysis hardly guarantees the presence of such hidden clauses. If another theory allows that exceptions can be explanatory without hidden clauses, it is to be preferred.

Metaphysics

One unexpected implication of Lance and Little's theory is that natural kinds become oddly parasitic on the laws that describe their behavior: "what it is to say that *K*s are essentially defeasibly *F* is to say that what it is to be a *K* is to be the sort of thing which functions either as *F*, or as something that serves, in context, as a suitable variant on the kind of thing that is in a privileged context and is *F*."[46] Also, "any *K*, whether in privileged conditions or not, is of a kind that is constitutively such that in privileged conditions it is *F*." Lance and Little offer the

---

[45] Ibid., 297-298.
[46] Lance and Little, "Where the Laws Are," 166.

example of chairs being the kind that is defeasibly for sitting on (some chairs are just there to be admired).

This might make some sense for chairs, but in a moral context Lance and Little might put the cart before the horse. Consider the kind "honest statements"; is the fact that honesty is defeasibly rightmaking really *constitutive* of this kind—is this defeasible normative property "*of the essence*" of honest statements?[47] Normative defeasibility could play some classificatory role, but I am reluctant to give it that sort of metaphysical priority. It seems more likely that what classifies honest acts together is that they are all honest, and that they are defeasibly rightmaking emerges from them being honest but isn't essential to it.

Normativity

My sharpest disagreement with Lance and Little come from how they cash out the "exception-laden" nature of defeasible generalizations. As I have said, they argue that defeasible generalizations describe the behavior of a kind—like matches—under privileged conditions. When the kind diverges from this behavior (e.g. the match refuses to light; it's too windy to maintain a flame), it is due to it being in non-privileged conditions. But the relationship between privileged and non-privileged conditions isn't a binary, either/or relationship. It follows a gradient. Deviant cases carry the "trace" of what its behavior would have been under privileged conditions.[48] For example, there is a charming chess variant called FischerRandom (also known as Chess960), which is played just like regular chess except that the starting positions are different, with the non-pawn pieces arranged in randomized order on the usual first and eighth rows. There are 960 starting positions available, but after the pieces are placed the games proceed normally. Lance and Little's point is that deviances like FischerRandom are only fully understood in relation to normal chess, chess in the privileged position of a rook on the A- and H-files, then the knights, the bishops, and the king and queen in the center.

I have spoken in counterfactual terms ("would have been"), but Lance and Little do not make explicit that the "trace" itself just *is* this modal property. About pain Lance and Little claim that "it is only because pain is paradigmatically bad-making that athletic challenges come to have the meaning they do…"[49] So pain-kind carries a trace of its usual bad-making property, and cases where an instance of pain is good-making communicate special significance because of the trace. But I am not sure what the trace *is*. We are supposed to "grasp" it; grasp what exactly? It's not like everyone who encounters a deviant case of a kind knows that it is deviant; they may not have encountered the correct generalization before. Lance and Little rightly say that deviant cases, like FischerRandom chess, are often explained in relation to the "normal" cases from which they deviate. But they don't have to be. One finds people playing games by "house rules,"

[47] Ibid., 165.
[48] Lance and Little, "Defeasibility," 300.
[49] Lance and Little, "Where the Laws Are," 166.

following the norms they grew up with; perhaps they don't collect money when they land on Free Parking in Monopoly. Many amateur chess players do not know how to properly long-castle, and if I take their pawn *en passant* they might think me to be a cheat. Privileged conditions must be relative to the norm-maker. And although we can allow relativity in the case of Monopoly and chess, in morality this is problematic. We want honesty to be the same concept-kind for everyone; we hope that honesty's generalized valence—regardless of how we spell out that notion—is to be right-making. So the norm-maker had better be the same for everyone.

For this reason, Lance and Little say that the norm-maker is, broadly, humans. They reject the possibility of "privilege-in-itself," and instead claim that privileged conditions reveal a "subtle dependence on human interests."[50] Lance and Little leave open the question of what these human interests are, but it is hard to avoid the conclusion that they encompass eudaimonia, the Aristotelian term for "human flourishing." The seductive power of teleology is familiar in moral philosophy. But eudaimonia is incompatible with moral particularism.[51]

However, Lance and Little do tap into one crucial property of principles: the "trace." Any theory of moral principles must explain how even exceptional cases seem to refer back to the principles. The theory analyzed in the subsequent section, Horty's default reasons, does not capture this property. My theory, explicated in the next chapter, will claim that moral judgment is performed in the shadow of moral principles, even when the principles are false. Principles create space that enables analysis under false conditions. When we believe a moral principle, we suspend disbelief. But even reading an engrossing novel, we don't forget about the real world. In fact, if it is a very good novel, it may even clarify the trials that we encounter in everyday life.[52] This is my view of moral principles, including Kantian maxims.

IV. Default reasons

Defeasible generalizations are intimately connected with default reasons. A default is what Lance and Little call a "justified 'start-here' position."[53] In Lance and Little's theory, a moral agent's default always accords with the defeasible generalization they hold, so long as the agent lives in privileged conditions and *knows* that they do. If the generalization, "birds fly," is held by an average agent, then their default should be to think that birds fly; but if the agent has been confined to an emu farm, then their default should probably be that birds don't fly. When they encounter an emu, their thought process shouldn't be: "This is a bird. My default is that it flies. But it is in emu, so it doesn't fly." If they are rational, they should simply think: "This is an

[50] Ibid., "Where the Laws Are," 167.
[51] Christine Swanton persuasively argues this point in *Virtue Ethics: A Pluralistic View* (2003).
[52] Joyce, "Moral Fictionalism," *Fictionalism in Metaphysics*, ed. Mark Eli Calderon (Oxford: Clarendon, 2005).
[53] Lance and Little, "Defeasibility," 301.

emu, so it doesn't fly." They skip the default-defeating step because they aren't surprised to encounter a flightless bird on an emu farm. Within this context, nothing about a flightless bird needs to be explained. The default would only hinder successful reasoning. One could object that overcoming such a default takes no mental work at all for the agent, and that once she leaves the emu farm she will want her normal default, "birds fly," returned to her. Well, that is exactly what will happen. One forms defaults, *a posteriori*, based on empirical observations of one's context. Any correct account of defaults will accept this contextualist claim.

This contrasts with the invariant view of defaults. One might think that honesty, for instance, is invariantly default right-making. (Of course, the contextualist could cop to this. Some reasons will have the same default in any context; what matters is that the default is not preordained by the nature of the reason). Notice where the adverb goes: what is invariant for this view is the default, not the valence. W. D. Ross defends invariant defaults by defending honesty as invariantly right-making. If a reason always points one way in virtue of its nature, then its default would likewise be invariant. Dancy recognizes the holism of reasons, but, like Ross, may inadvertently defend invariant defaults. Dancy says that a default reason is a "consideration that is reason-giving unless something prevents it from being so."[54] But when a default changes (e.g. when honesty goes from default right-making to default wrong-making), it is exactly because something prevents the reason from acting as it did previously. The difference is that it is prevented *systemically*, not just in a single instance; it is prevented by altered circumstances for the moral agent. That default reasons "arrive switched on" serves as an apt metaphor so long as one means that the reasons arrive *to the agent in a context* switched on, and not that they arrive *in existence* switched on and pointing to an unchanging valence.

Muddling the issue is the referent for "default." Dancy uses the phrase "default reasons" because defaults *are* reasons. When he claims that honesty is default right-making, I think he means that the property of being "default right-making" belongs to honesty. The default is fully vested in the feature to which it belongs; it is an invariant property of the feature. Thus honesty arrives "switched on" to the right-making valence because honesty, in any context, *is* honesty, and so it always possesses the same default. But Lance and Little divest considerations like honesty from their defaults. They would say, at most, that right-making "default belongs" to honesty. I think this makes more sense, partly just because we use the word "default" instead of "disposition." If defaults were states of considerations or properties, we would call them dispositions: tendencies to manifest in certain ways, though not always. But dispositions have grounds. Glass is always fragile and has this property in virtue of its delicate molecular structure. Dispositions are not context-dependent, and I can't think of what it would mean for them to be so. Honesty would need to have the property of sometimes being default right-making, sometimes default wrong-making, depending on the context. But presumably the nature of honesty itself doesn't change—it is always the same consideration. Indeed, that fact plays a

---

[54] *Ethics without Principles*, 112.

critical role in delivering the thesis of holism. Nothing *within* honesty produces this internal change in its default status. Context is the catalyst. It makes much more sense to vest epistemic agents with default *beliefs* about considerations than to claim that the considerations have intrinsic default valences.

Defaults also play a role in moral explanation. Dancy says that when a consideration behaves in alignment with its default, "there is nothing to explain," but when a default is broken, the situation must be unusual and demands some sort of explanation.[55] I agree with this claim. The term "unusual" reaffirms why defaults should be understood as contextual. There is nothing unusual about encountering an emu on an emu farm, nothing to be explained. I suppose one could hold the default "birds fly" on an emu farm and explain the constant presence of evidence that birds don't fly by my unusual location. Most people don't live on emu farms; in that respect my whole existence is exceptional. But what about a moral agent who grew up on an island of only flightless birds? Then the default should surely shift. Flightless birds on the Island of Flightless Birds should not demand explanation. It would only be an irrational mind that didn't shift defaults to fit their context. The epistemic agent who has never seen a flying bird should react with awe at the sight of their first pigeon.

Default reasons must themselves be reasoned out from the epistemic positioning of the agent. A theory of defaults from John F. Horty in *Reasons as Defaults* (2012) captures this feature. Horty's broad view is that reasons, both practical and theoretical, operate on a default logic. He says:

> [G]iven a background theory $\Delta = \{W, D, <\}$ and a scenario S based on this theory, the proposition X should be refined as a *reason for Y in the context of S* just in case there is some default δ of the form $X \rightarrow Y$ from the underlying set D that is triggered in the context of S… In a situation like this, we will say that the default δ *provides* a reason for Y, and that the reason it provides is X.[56]

In this formulation, W is the set of propositions that the epistemic agent knows. It is the background information. D includes the default rules. Default rules are rules of the form $X \rightarrow Y$, where X is a premise that, once established as true, guarantees Y. The symbol "<" is used to represent a priority relation between the rules. This is where default logic differs most sharply from classical logic. Priority allows for a rule, $X \rightarrow Y_1$, to be "defeated" by another default rule, $X \rightarrow Y_2$, which takes priority. An agent could have two default rules, say δ_1 and δ_2, which are triggered by the same proposition X. There must be an ordering of the rules. In classical logic without defaults, this isn't possible, because both rules are simply triggered at once, even if this leads to an irrational result. With defaults, any rule is by nature defeasible, unless it gets prioritized over every other rule. Presumably our emu farmer does have a default that birds fly.

---

[55] Ibid., 113.
[56] John Horty, *Reasons as Defaults*, 41.

But when he encounters an emu, when he recognizes it as an emu a different default takes priority: that emus *don't* fly. A *scenario*, S, is a subset of the default rules, D, that (roughly) pertain to an agent's circumstances and support the agent's conclusions. What makes it into an epistemic agent's belief set are the conclusions drawn from the defaults belonging to the scenario, plus the background beliefs W. The defaults must be logically compatible with W.

There is much in this theory to delight the moral particularist. As Horty recognizes, default logic provides a framework for understanding the defeasibility of moral principles. If principles are defaults, then they can be defeated when a stronger principle wins out. The notion of *strength* is explained by priority, <. A man late to an important meeting, but an elderly lady has fallen in the street. Does he stop to help her? Priority allows for two default rules pressing on the moral agent, but only one can win out. There is supposed to be some logical priority to the two rules. (The rules are a. Help those in need; and b. keep promises.) Priority resolves this apparent moral conflict.

Horty recognizes his theory's general kinship with moral particularism. He presses default logic into the service of what he calls "moderate particularism," in contrast with Dancy's "extreme particularism."[57] Extreme particularism, according to Horty, entails the rejection of all moral principles insofar as they might be useful to moral judgment. Horty suggests that moral principles are default rules that, when taken together as a set, occasionally lead to failures of moral reasoning, but that rather than rejecting the set it should get repaired. There are three ways in which a principle can be defective: (1) when it, in conjunction with other principles, leads to conflicting obligations; (2) when, even without an explicit conflict between principles, the obligation that accords with our principles seems intuitively wrong; (3) when our principles fail to give a reason at all, but we feel strongly that a reason ought to be there.[58] Notice that in the first case the defect comes from principles conflicting with each other; in the latter two cases the principles conflict with our moral judgment. Already this picture commits Horty to a view of judgment not governed by principles. Hence it is not surprising when Horty claims that the revisionary process for moral principles happens by appeal to something outside the principled realm.

Appeal to an external authority is a familiar motif in ethics; we saw it with Lance and Little explaining "privileged conditions" via special human interests. Something must bridge the gap between the laws. Horty's search for an arbitrator of principles gets the relationship between principles and practical reality reversed. Horty and others appeal to practical reality to repair the defects of principles. But principles do not get revised in this way. Moral agents do not put an asterisk next to each moral principle listing all the exceptions. They pursue moral judgment using principles that they know full well are false. Moral theory must reckon with this reality.

---

[57] Ibid., 154.
[58] Ibid., 161.

Even granting that principles get revised, is such revision desirable? Lance and Little think not, because good moral judgment involves knowing when principles apply and when they don't:

> [E]pistemic generalizations, like most moral generalizations, are irreducibly porous. They are shot through with exceptions we cannot eliminate. These generalizations can nonetheless count as robustly explanatory and insightful. Adducing them has a power a list of instances does not, for it situates instances within a framework that maintains some as exceptions to others' rule.[59]

Purging moral principles of their exceptions is akin to purging them of their nuance. Principles without exceptions would call for a specification of narrow scope: "Honest acts are right, so long as the object of our honesty is deserving, what we are honest about causes no harm, honesty is expected rather than mendacity, one doesn't live in Nazi Germany… etc." On this view, as one learns more about honesty in different situations, one adds exceptions to the principle. But these principles don't seem very useful to a moral agent; they hardly apply anywhere and are unwieldy to deploy. And this isn't the sort of principle revision that I think Horty would defend. He speaks in terms of defaults. A default reason tells an epistemic agent how to act in a context, other things being equal. The presence of certain phenomena might dislodge the default, but these exceptions aren't *built into* the default, where the default δ is of the form X → Y. Where would the exceptions go? Insofar as moral principles receive revision, it is by adding new defaults with priority conditions based on new scenarios. When one moves to Nazi Germany, a new default assumes the position of priority.

But this account of revising moral principles gives up what made Lance and Little's theory appealing: a clear-cut account of "trace," the lingering grip a principle exerts on the deviant circumstances where the principle doesn't obtain. Lance and Little argue that moral judgment depends on knowing privileged circumstances for a principle and grasping degrees of deviancy. Is this equivalent to knowing that one's default in a context hangs low on the chain of priority? I.e. does the honest person in Nazi Germany to whom we keep referring recognize the deviancy of their position?

I am far from saying that Horty's theory of default reasons has no merit for moral particularism. Just the opposite, I think his account is probably right in its analysis of the logic behind default epistemic positions. But I also think that moral principles are something else. They do leave a trace. The phenomenological elements of moral judgment—the feeling of recognizing deviant circumstances, of being shocked or regretful because a reason behaves in an unusual way—contradicts an account that holds the principles themselves to be defaults.

---

[59] Lance and Little, "Defeasibility," 309.

Moral Explanation in the Kingdom of Ends

I. Explanation in brief

In a widely cited paper on Kant, John Rawls employs a curious phrase. Rawls says that in considering the Kingdom of Ends, we explore the possibility of "ideal social cooperation and of the person, by formulating what I shall call 'model-conceptions.' We then reason within the framework of these conceptions…"[60] I come to the notion of "models" from a different angle, following Cartwright's scientific philosophy, but we mean similar things. My models lay out conditions in which maxims hold true, sets of circumstances with limited supplies of reasons. These, like Rawls's models, are idealizations. And, like Rawls's models, their purpose is explanatory. They empower a moral agent to understand the operation of maxim, despite its obvious exceptions. Models guard principled thinking from the unruliness of the real world.

Explanation shares a clear bond with reasons. Carl G. Hempel and Paul Oppenheim, in "Studies in the Logic of Explanation" (1948), a founding text in the philosophy of scientific explanation, write, "To explain the phenomena in the world of our experience, to answer the question "why?" rather than only the question "what?", is one of the foremost objectives of all rational inquiry."[61] Now, the distinction between "why" and "what" questions has been smudged by later philosophers. Many "why" questions can easily be transformed into "what" questions, and vice versa. For example, the question of "Why did he write a thesis?" could just as easily, with no change in meaning, be asked as "What drove him to write a thesis?" Responses to these questions often take the form of reasons. Explanations probe the "resultance base" of an action, the features that favored or disfavored it.

Hempel and Oppenheim defend the Deductive-Nomological theory of explanation (henceforth D-N). This theory has become the view philosophers of science argue through and against. It is the hegemon; other theories of explanation are understood in contrast to it. D-N holds that for any phenomenon in the physical sciences, it gets explained by conjoining some features of the phenomenon to general laws pertaining to it. Hempel and Oppenheim's best-known example is about a thermometer submerged in hot water.[62] What explains the fact that the mercury rises? Statements of two kinds work in tandem to do the explaining. "Antecedent conditions" describe the phenomenon in question, e.g. that it is a glass thermometer containing mercury, that it is submerged in water of a certain temperature. "General laws" state regularities in the natural world, e.g. laws of thermal expansion. The phenomenon—that the mercury rises—can be *deduced* from a set containing the two kinds of statements.

---

[60] Rawls, 520.
[61] Carl G. Hempel and Paul Oppenheim, "Studies in the Logic of Explanation," *Philosophy of Science* 15, no. 2 (April 1948): 135.
[62] Ibid.

D-N faces several substantial challenges, and few philosophers of science today would endorse the version of D-N proposed by Hempel in the 40's.[63] I will not go over all of D-N's deficiencies, but here is one example: it struggles with the direction of causality. Suppose that Hempel wants to explain the length of a shadow cast by a flagpole. Under D-N, Hempel explains it with statements about the elevation of the sun, the height of the flagpole, and the law of light propagation. So far so good. But suppose Hempel instead wants to explain the height of the flagpole. Reordering the explanation lets him explain the height of the flagpole by conjoining statements about the sun, light propagation, and the *length of the shadow*. The flagpole's height is explained by the length of the shadow it casts. Intuition protests this direction of explanation, though this type of objection is somewhat inconclusive.

My issue with repurposing D-N for moral explanation that it is incompatible with reason holism. The consequence of holism is that moral principles have exceptions. They cannot operate in formal deduction; one can't deduce from a "law" that only *sometimes* obtains. Hempel and Oppenheim write, "our definition of explanation will be so constructed as to guarantee the factual character of the totality of the laws."[64] Natural laws are (supposedly) facts expressing exceptionless regularities in the world (Cartwright disagrees). Moral principles may ostensibly describe regularities, but they are far from exceptionless.

Fortunately, as Uri Leibowitz points out, moral particularism need not meet the demands of D-N; there are other theories of explanation available.[65] Many theories of explanation that philosophers of science have defended with will not fit morality. Explanations relying on statistics and probabilities seem obviously wrong for moral explanation. When we think about explaining morals, it isn't about what occurs with regularity, but about what makes it so. Other theories of explanation show more promise. The theory of explanation I will defend is based on Cartwright's view of modern physics. Her view has two main components: (1) the laws of physics lie by overgeneralizing about phenomena. (2) Even laws with exceptions can play an explanatory role, expressing what happens under constraints and describing the capacities of certain features.

Many scientists assume that every fact in the universe can be explained. A theory that leaves much of the universe unexplainable is judged harshly. But the importance of this criterion—how much a unified theory can explain—belies the usefulness of many theories of limited scope. Newtonian physics faces an explanation deficit. It does not explain what happens at very large nor very small scales. Neither does any other single theory; relativity and quantum mechanics only successfully explain the phenomena within their domain. They do not help an engineer to explain why a bridge stays up. The search for a "Theory of Everything" in physics aims at a single theory that explains every fact in the universe. Perhaps this search is quixotic.

---

[63] Uri D. Leibowitz, "Scientific Explanation and Moral Explanation," *Nous* 45, no. 3 (September 2011): 473.
[64] Hempel and Oppenheim, "Studies," 153.
[65] Leibowitz, "Scientific Explanation and Moral Explanation," 483-488.

Cartwright defends a "patchwork" view of the world, laws applicable only to their limited scopes, Newton's laws to bridges, quantum mechanical laws to particles.[66]

Is there a similar theory in ethics? The friendliest view to Cartwright's is moral particularism.[67] A particularist holds that moral situations are *sui generis* insofar as they are not governed by exceptionless moral principles, which would naturally link them to other moral situations. Not all situations where one tells a truth are the same; in some of them a consideration will defeat the obligation to be honest, and in some of them honesty will be wrong-making. This reveals an important difference from Cartwright's project. Cartwright emphasizes that laws hold only for the physical models they explain, placing less emphasis on explaining phenomena that appear to bridge several models, since they crop up less often in physics. But moral agents must do just that—work *between* maxims.

In such cases, it's worth considering precisely what we aim at when we aim to explain moral phenomena. Philosophers of ethics have rarely, if ever, considered "moral explanation" as a topic of independent interest. Moral explanation is the tool that philosophers wield all the time, without ever glancing down at what they hold. Non-natural moral philosophers argue that there are non-natural facts that can't be completely explained by natural facts. There are foundationalists in moral philosophy who takes some moral concepts as brute, including Derek Parfit's "to-be-doneness." In Kant's *Groundwork* III, he comes up against an intractable explanatory barrier, which will clarify the explanatory project:

> But any human reason is entirely unable to explain *how* pure reason, without other incentives that might be taken from somewhere else, can by itself be practical, i.e. how the mere *principle of the universal validity of all its maxims as laws* (which of course would be the form of a pure practical reason) without any matter (object) of the will, in which one could take some interest in advance, can by itself yield an incentive, and produce an interest that would be called purely *moral*…all the effort and labor of seeking an explanation for it are lost.[68]

Kant's moral system requires the presupposition that humans have a free will. From what we know about ourselves and moral obligation, we know, synthetic a priori, that free will is a necessary truth. But the metaphysics behind free will is unknowable. There occurs a "boundary of human reason."[69] Things-in-themselves are the hook on which we hang our metaphysical hat, but their nature remains mostly inscrutable. Kant thinks demanding explanation of the things-in-

---

[66] Nancy Cartwright, *The Dappled World: A Study of the Boundaries of Science* (Cambridge: Cambridge University Press, 1999), 33.
[67] Constantine Sandis makes a similar point in "Dancy Cartwright." His conclusion, though, is that Cartwright should adopt Dancy's particularism, whereas I borrow Cartwright's notion of explanatory-yet-false laws to improve moral particularism's explanatory power.
[68] Immanuel Kant, *Groundwork for a Metaphysics of Morals*, 4:461.
[69] Ibid., 4:463.

themselves is unreasonable. The only way for a practical law to have the character of necessity is if it is partly inscrutable, balanced on the tip of the unknowable world of pure rationality.

This passage opens a discussion about moral explanation more broadly. The strength of Kant's system is that almost everything can be explained. Few facts must be accepted as brute. Specifically, as I will continue to argue in the final section, Kant opens a window for insight into the connection between normativity and reasons. One sees that a worthy reason for acting is one featured in a well-formulated maxim integral to one's practical identity. That is, good reasons are just the reasons which the Categorical Imperative takes to produce binding obligations. The CI, a formal law on practical reasoning, gets explained by the absolute worth of pure rationality. The system is relatively closed; few loose ends. But I am less interested in performing a close examination of Kant's explanatory structure, than in seeing how Kant can help us explain moral phenomena.

II. Preliminaries to explaining moral phenomena

Defining the explanandum

When a moral philosopher claims that A explains B, the meaning of their claim depends on the thing to-be-explained, the *explanandum*. The thing that does the explaining, the *explanans*, takes its cues from there. To understand moral explanation, we therefore need a clearer idea of what constitutes a moral phenomenon.

Many moral facts need explaining. I am curious to know exactly why abortion is wrong, when it is wrong. The explanation, I suggest, would explain a particular moral phenomenon. Let a moral phenomenon be a state of affairs expressed by a statement of the form "χ is F," where F is some moral property and χ an action or action-type. It is roughly equivalent to an *event* in physics, e.g. the trajectory of a single electron. Traditional moral principles have the same syntactic structure as moral phenomena, because they express the moral valence of action-types. The particularist, though, is more interested in particular actions.

I am also curious to hear explained the book *The Sources of Normativity*, by Christine Korsgaard: a notoriously difficult tome. This sort of explanation would explain a *theory* rather than a phenomenon. Although theories explain, they, too, need to be explained. The physics parallel might be a broad explanation of quantum mechanics. As I said, a different thesis would look closely at Kant's *Groundwork* and pull out strings of theoretical explanations. But since my goal is to integrate Kant with particularism, I take most of this theoretical data as given.

To further clarify: interior explanations are about particular moral phenomena. Exterior explanations are about moral theories. An example of an interior explanation might come in response to the question, posed to a utilitarian, "Why is it wrong to divert the trolley?" The Kantian is tasked with explaining a moral phenomenon—the wrongness of diverting the trolley—but they are not, at least not immediately, tasked with explaining Kant's moral philosophy as such. As they are generalists, they might respond, "It is wrong to divert the trolley

because it violates the maxim prohibiting murder," and this would serve as an interior explanation. Exterior explanations illuminate theoretical difference. Their *explananda* are philosophical tenets and arguments. If someone asks their professor to explain utilitarianism, they are likely asking for exterior explanation. The professor's answer had better not assume knowledge of utilitarianism as background information.

Under D-N theories of explanation, phenomena are explained by reference to principles; principles themselves usually depend on some background theoretical commitments. Principles organize individual moral phenomena according to a shared feature, ϕ, and a shared outcome, F. With moral principles, there is no guarantee phenomena are *correctly* described by the principles that apparently subsume them. But for finite, well-ordered beings like ourselves, the move to organizational clarity is necessary. Insofar as the will is rational, it is consistent, which is to say the same reasons lead to the same actions. Each case of this gets captured by a principle of the will, a maxim. But we know that many moral situations do not neatly fit under a single well-formulated maxim.

Formal and informal explanations

I want to better clarify the standards to which philosophers should hold themselves when giving explanations. It depends on the aim. When a philosopher says fact A explains fact B, they may refer to a specific relation that holds between two facts. The relation could be that A grounds B, B in virtue of A, or B because A. Other philosophers (and sometimes the very same philosophers) use "explains" more loosely. They will say things like "And now I will explain…" without intending to say that they themselves ground what they will then proceed to explain. They take an "explanation" to express to a thing one says to clarify an issue or to engender comprehension in a listener. The former notion of "explanation" I call *formal*, because it entails that there is a form to the "explaining" relation, a law underwriting every instance of it. The latter understanding I call *informal*, not just because it pops up when philosophers are writing more casually but because it does not require a single form of explanation.

The requirement that explanations take just one form goes against our intuition. It may be that grounding represents the paradigmatic case of explaining, but is it the *only* case? Moral phenomena often refuse clear grounding or "in virtue of" relations. For any action χ, the generalist who believes in exceptionless moral rules may insist that χ's moral valence, F, is explained by a rule, e.g. the rightness of telling the truth is explained by the principle, "honest acts are right." But this explanation seems superficial at best. It hardly answers the question of what makes χ right. If someone offered me this explanation for the rightness of their action, I would be nonplussed. At worst, this explanation is just *wrong*. I have argued that most moral principles are false statements because most have exceptions.

Yet it is often what we see. Many moral philosophers assume a Deductive-Nomological model of explanation (D-N), where moral facts get explained by exceptionless rules. D-N is a theory of formal explanation that holds that all explanations subsume the object of explanation

under exceptionless rules. The explanation consists in features of the *explananda* conjoined by the rules "triggered" by them. This model has undergone a theoretical siege in the philosophy of science the past 60 years, but in ethics it lingers quietly, casting a long shadow. Moral philosophy contented itself with D-N because it likewise contented with moral principles. But if the holism of reasons is true, then D-N must go. One can't subsume moral facts under rules full of exceptions.

One motivation for sticking to D-N is clarity. With formal explanations, we can clearly tell what is and isn't an explanation, and whether attempts at explanation are good or bad. But even as far as clarity goes, context ultimately rules. Even a formal explanation is unclear in an inappropriate context. Scriven writes:

> [T]he completeness or correctness of an explanation is a notion without meaning except in a given context from which the type [of explanation] can be inferred and in which the required facts are known. This is greatly obscured by the supposition that in science there is always something known as *the* explanation of a particular phenomena regardless of context. On the contrary, there are many non-competing types of explanation for scientific phenomena, just as for historical.[70]

According to D-N, if an explanation doesn't rely on exceptionless laws, then it isn't an explanation. This is far too limiting. There are ways to classify and evaluate explanations without purging explanation of its variety. In fact, rigid adherence to formality belies the context, nuance, and complexity of explanations. There is usually more than one explanation possible for a phenomenon. The question posed to the researcher is which explanation explains best. The next section looks at ways to evaluate informal moral explanations.

Evaluating moral explanations

Scriven gives three ways a poor explanation can fail; hence three correlative ways a good explanation *must* succeed.[71] Explanations can be inaccurate, including extreme doubtfulness; they can be inadequate, not meeting a threshold for explaining a phenomenon; or they can be irrelevant, belonging to a type not called for by the context. Accordingly, good explanations must have truth-justifying grounds, role-justifying grounds, and type-justifying grounds. That explanations must have these grounds does not mean they must *include* these grounds. It would be odd for an engineer's explanation of the sturdiness of a bridge to include Newton's laws *and* a justification for their presence, e.g. "this is a classical physics problem on a human scale, calling for such laws." It would be even stranger to include in the explanation, as truth-justifying grounds, a derivation of Newton's laws or statistical proof of their general reliability. Scriven

---

[70] Michael Scriven, "Truisms as the Grounds for Historical Explanations," in *Theories of History,* ed. Patrick Gardiner (New York: The Free Press, 1959): 449.
[71] Ibid., 446.

rejects any single criterion for including or excluding from an explanation the three types of grounds; it all depends on context.

The three types of grounds in Scriven's theory of historical explanation conveniently mirror the three members of a holistic reason's "resultance base." As I stated in the Introduction, a resultance base is a set of features that necessitate the rightness or wrongness of an action. It has three sorts of members: contributory reasons, modifiers, and conditions. Contributory reasons favor or disfavor an action. Modifiers intensify or attenuate a contributory reason, while conditions enable or disable the reason. A defense of the robust distinctions between these three sorts of features can be found in "Conditions, Modifiers and Holism" by Ralf Bader. Tellingly, Bader sometimes refers to the "resultance base" as the "full explanation."[72]

The table shows these concepts in parallel:

Failures of explanation	Grounds of explanation	Resultance Base
Inaccuracy	Truth-justifying grounds	Contributory reason/source of reason
Inadequacy	Role-justifying grounds	Modifiers (intensifiers and attenuators)
Irrelevance	Type-justifying grounds	Conditions (enablers and disablers)

I have disputed the supremacy of formal explanations and instead supplied criteria for evaluating informal explanations. There is substantially more to say about why D-N fails for science and history, but it shouldn't take much to convince a particularist to look elsewhere. Of course, there are formal theories other than D-N. But as particularism celebrates the unruliness of the moral plane, it would be odd for there to be only one form of moral explanation. Explanation gives moral agents a compass to navigate the shapelessness, but the tool must fit the occasion. We should recognize a diversity of explanations as a positive attribute, not a defect.

III. Modeling moral principles

One possible objection to Scriven's criteria for explanations, and to Dancy's "resultance base," is that an explanation or base could expand indefinitely, proliferating to facts without limit. For any enabler in the base, there is a state of affairs in the world that *makes* it an enabler—a truth-ground for the enabler. The same goes for modifiers; they can have enablers

[72] Ralf Bader, "Conditions, Enablers, and Modifiers," *Weighing Reasons*, ed. Errol Lord and Barry Maguire (New York: Oxford University Press, 2016): 24.

and truth conditions, too.[73] Dancy resolves this apparent difficulty by marking a difference between explaining the rightness an action and explaining the *explanation*. Nesting explanations effectively severs the possibility of an explanation entering infinite regress.[74] Scriven's solution is similar. Scriven says that because explanation already relies on context, there is no cost in insisting that a *good* explanation need not be "complete" in the sense of "infinite." He notes that the very idea of a complete explanation is incoherent. Explanations link what we understand to what we do not. If we attempt to explain everything, this just implies that we understand nothing and can explain nothing: infinite explanation is "like the sound of one hand clapping,... a logical echo, a thing of no substance whose loss is no loss."[75]

Cartwright models represent one way to limit the scope of an explanation and avoid infinite regress. Let's look closely at the process of explanation that Carwright defends, here summarizing Duhem:

> There is nothing more than the rough facts of nature that sometimes some things behave like others, and what happens to one is a clue to what the others will do. Explanations provide a scheme that allows us to make use of these clues. Light and electricity behave in similar ways, but the procedures for drawing the analogies are intricate and difficult. It is easier for us to postulate the electromagnetic field and Maxwell's four laws, to see both light and electricity as a manifestation of one single underlying feature. There is no such feature, but if we are careful we are better off to work with these fictional unifiers than to try to comprehend the vast array of analogies and disanalogies directly.[76]

Models are theoretical boxes that constrain the universe of facts. They pare down a situation to its simplest elements. In doing so, they explain patterned phenomena. Consider the unruliness of modern physics. We can't run our finger along the inscrutable path of an electron. There is no simple, real, Platonic form for this object. Instead, we model electrons with the "electron cloud." Clouds enable us to comprehend essential facts about an electron's behavior: it's quantum linkage to another electron with opposite spin; its untestable trajectory. The cloud model provides a way for finite beings to comprehend electrons. More than that, it provides mathematical structure for the concept of an electron. Schrodinger's equation, $H\psi = E\psi$, gives the probability of finding an electron within a general region around an atom's nucleus.

Models create stability that the real world lacks. A model acts in tandem with a "nomological machine," which is a fixed arrangement of features—fixed to produce regularities.

---

[73] Bader disagrees. His position is that enablers are either present or not, and if they are not present, then they do not qualify as "enablers." He says: "only enabled grounds of enablers classify as enablers" (14). But this is really just a convenient classificatory move, not a substantive argument about the infinite extension of the resultance base. He admits that fully specifying moral principles would require stating "every non-normative ground of a normative condition… this procedure will generate infinitely complex principles" (14).

[74] Dancy, *Ethics without Principles*, 46-47.

[75] Scriven, "Truisms," 447.

[76] Nancy Cartwright, *How the Laws of Physics Lie*, 95.

We fix the features, not nature. Nature hurls up unruly electrons, and we find mathematical models for them. In the passage, the scientist seeks to explain the similar behavior patterns of light and electricity. We know about their behavior by looking *at* them, loosely speaking, but not looking *inside* them. So the scientist postulates a model—the electromagnetic field and Maxwell's four laws—that explains their behavior. Without the model, we could not explain these phenomena, because they don't present a gift-wrapped natural law that determines their similarity. Outside of models the world is highly chaotic and true. Inside the models, the world is organized and false. Cartwright says, "Where there is a nomological machine, there is law-like behavior."[77] And the law-like behavior is *only* there.

The same is true for contributory reasons. Take a feature that some actions have: they tell the truth. The principle "truth-telling actions are right" is false. We cannot generalize to all actions that honesty makes them right; this has been established. But imagine a model that makes the principle true: a toy world where honesty's right-making capacity were always enabled. For each moral principle that we have, there is such a model. An idealized universe designed *for* the principle. I have been saying that moral principles are false. Let me refine this claim: they are globally false, false in the real world without reference to models. But they are *locally* true to their corresponding model.[78]

The trolley problem is a paradigm moral model. A professor asks her class to imagine a trolley racing down the tracks towards three helpless victims. The class has the option to switch a lever and redirect the trolley to just one victim. Should they? In an introductory philosophy class, the students would ask about variables the professor excluded. How old are the victims? Is there any chance of them escaping the train or surviving the crash? Does the conductor have time to intervene? The professor pushes back, insisting that the students work within the limited model. The model simplifies a complex hypothetical situation to isolate key questions about moral duties and the value of human life. The model explains the Principle of Utility by conjoining it to an idealized situation. The messiness of real life is not supposed to interfere.

Models and their governing principles do not exist for moral agents to encounter; their usefulness is predicated on the world being too complicated for principles to capture. What models do is *impose shape on the moral plane*. Critics of particularism worry about a shapeless relation between the non-moral and moral sets of facts. I argued in the introduction that this worry has legs. Shapelessness would strike against the possibility of fully explaining moral phenomena. In the particularist framework, we scratch the bottom of explanation rather quickly.

In the next section, I will clarify how models improve the chances for moral explanation under particularism's constraints. I will argue that the particularist would be wise to adopt

---

[77] Nancy Cartwright, *The Dappled World*, 57.
[78] Don't let truth get tangled up here. Although a model has its own facts, these are only imagined states of affairs. Models are a convenient fiction out of which one can easily step. So to say that something is "true" in a model is to speak pragmatically.

Kantian maxims as the explanatory-yet-false principles that do the work of explaining. Maxims, as I have shown, can be understood as defeasible principles, meeting the particularist's one condition. In the case of maxims, each maxim is a model unto itself, demonstrating the perfect rationality found in the kingdom of ends.

IV. The shape of ethics

Kant offers a third formulation of the Categorical Imperative, the Formula of the Kingdom of Ends: "Act according to maxims of a universally legislating member of a merely possible kingdom of ends."[79] The kingdom is a toy-world of perfectly rational beings who, in virtue of their perfect rationality, regard themselves and each other as ends, never merely as means. The maxims that one being wills are supposed bind universally. This notion has been, in my view, distorted by political theorists. On one sketch of universal legislation, the maxims exceptionless across all beings with regards to actions. If a maxim prohibits lying, none shall lie. On the other sketch, the maxims are universal insofar as they *hold for each being's will*. If a maxim links a set of reasons to an action, then when faced with a certain set of reasons, each rational being will take the same action. The maxims, though, can still be defeasible in respect to actions; different reasons produce different obligations, so that one set of reasons can actually *obligate* any rational being in that particular circumstance to lie.

For obvious reasons, I prefer the second sketch. The first sketch, which is more generally accepted, including by Rawls, presupposes that the "laws" each rational being legislates are literal legal doctrinaire. But maxims have always been principles of willing. They have always connected reasons to actions, without necessarily presenting an ideal legal code for states. Kant himself is somewhat ambivalent on this point, sometimes dragging maxims into his political philosophy of rights and duties. But keeping them separate makes more sense, at least in the *Groundwork*, in the light of the nature of maxims and the argument for their universality.

Because maxims are principles of willing, they exhibit the rationality of the being who self-prescribes them. Insofar as the being is rational, their maxims accord with the Categorical Imperative. Each other perfectly rational being must therefore hold the exact same principles of willing. Perfectly rational beings would, under the same circumstances, act the same. A single set of *rationes obligandi* should always and only necessitate a single duty, no matter the identity of the moral agent facing them. Otherwise two perfectly rational beings could respond differently to the same set of reasons, and this simply shows inconsistency in rationality itself, which is intolerable. The universality of rationality is precisely the universality belonging to the principles of willing that a perfectly rational being would take for itself.

Each maxim collaborates with the model that makes the maxim true, and together they explain the morality of rational action. The kingdom of ends represents an "overall" model in

---

[79] *Groundwork,*

which all perfectly rational maxims hold true. As Rawls suggests, the kingdom is a "model-conception," stipulating aspirations like "ideal social cooperation" and treating each other as ends rather than means. At least, this is how I understand the kingdom of ends.

But the kingdom isn't here yet, so for now we must learn to navigate the topsy-turvy alleys of real-world moral agency. I now return to the problem of shape.

Practical reason holism demonstrates that the non-moral plane is shapeless with respect to the moral. It would be unwise to attempt to list features of the non-moral world as conclusively right-making or wrong-making. Any feature could count as a reason for or against action, depending on other morally relevant features. This isn't to say that an infinite being couldn't codify the non-moral, using infinitely long chains of description to precisely refer to all sets of right-making reasons and all sets of wrong-making reasons. Dancy allows this.[80] But such a being is pie in the sky to us mortals. This being would be perfectly rational, always doing precisely as obligation demands. Humans have a sensible component, with sometimes faulty rationality, and so we could only dream of such unblemished principles of the will. This perfectly rational being might be considered a citizen in the kingdom of ends.

Because infinite principles resolve it, the problem of shape is also one of finitude. Humans must comprehend the moral plane; they need moral agency. I have argued that maxims, as principles of the will, organize and classify the will in relation to actions. They codify reasons for acting and, with models, explain why certain features of actions match certain moral valences. One should understand maxims in conversation with the problem of shape.

Shoelaces and charities: it befuddles philosophers that the two could have equal status in the non-moral plane. The befuddlement comes from the fact that charity seems to feature much more heavily in explanations of rightness, whereas shoelaces would feature only in unusual circumstances. The ubiquity of moral explanation explains why we sense an apparent difference between the two; we are responding to patterns we notice in how we explain the moral plane to *ourselves*.

But why should it seem that charity features more heavily? The primary reason is that maxims must assume the kingdom of ends even before it has come: "every rational being must so act as if through its maxims it were at all times a legislating member of the universal kingdom of ends."[81] I have already shown how beneficence falls out of the Categorical Imperative as a wide duty; giving to charity would be commonplace in a society that treated others as ends.

That said, giving to charity would not *always* be for the better, let alone obligatory, in the kingdom of ends. Members of the kingdom of ends are perfectly rational. They have even less need for exceptionless rules than us finite beings do. Their maxims, like ours, would display a non-monotonic logic. They would be defeasible. So although we sense a "shape" to morality,

---

[80] *Ethics without Principles*, 82.
[81] *Groundwork*, 4:438.

one need not exist, at least not one with exceptionless patterns. Our perception of a shape comes from our patterns of explanation, which include defeasible principles.

The union of particularism and Kantianism does not fully satisfy Audi's criterion of epistemic completeness, but I don't think that it should. The theory would be epistemically complete if a moral agent could use the theory to always know what they were obligated to do. But then there would be no space for the agent to *choose*, which is just as essential to the Kantian system. Wallace says:

> Some theorists hope to reformulate moral rules to give them the required degree of precision, comprehensiveness, and consistency. Such programs, however, encounter intractable difficulties. Individuals who proceed reasonably and intelligently in rule-governed activity do not simply passively do what the rules direct; they actively apply the rules.[82]

Wallace shares the particularist's distaste for exceptionless moral principles. Principles instill passivity in moral agents. Good moral agents do not merely follow an algorithm. They continuously strive to be more rational. They accumulate knowledge about the behavior of features in the world, like honesty and shoe-laces: knowing when they act as a contributory reason, what enables or disables them, and what strengthens and weakens them. Maxims get selected accordingly, partly based on the agent's a posteriori sensible experience and partly based on formal demands of the CI.

There is shape to the moral plane insofar as it reflects the structure of pure rationality. However, this does not imply that its patterns must be recognizable. Our finite codifications will always need exceptions, just as the particularist claims. But principles are still important. They are the *principles of willing*, maxims, that we represent to ourselves, organizing moral phenomena. Moral reality is shaped because we shape it. Creating order from disorder represents one of our primary functions as rational beings. Though moral principles may be false, they make the dappled world more knowable.

---

[82] Wallace, *Moral Relevance and Moral Conflict*, 56.

Bibliography

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

Bader, Ralf. “Conditions, Enablers, and Modifiers.” In *Weighing Reasons*, edited by Errol Lord and Barry Maguire, 27-55. New York: Oxford University Press, 2016.

Bonevac, Daniel. “Ethical Impressionism: Response to Braybrooke.” *Social Theory and Practice* 17, No. 2 (Summer 1991): 157-173.

Cartwright, Nancy. *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press, 1999.

———. *How the Laws of Physics Lie.* Oxford: Clarendon Press, 1983.

Coeckelbergh, Mark. *Imagination and Principles: An Essay on the Role of Imagination in Moral Reasoning*. New York: Palgrave Macmillan, 2007.

Dancy, Jonathan. *Ethics Without Principles*. New York: Clarendon Press, 2004.

———. *Practical Reality*. New York: Oxford University Press, 2002.

———. “In Defense of Thick Concepts.” Midwest Studies in Philosophy 20 (1995): 263-79.

———. “On Moral Properties.” *Mind* 90, no. 359 (July 1981): 367-385.

Frigg, Roman. “Models and Fiction.” *Synthese* 172, no. 251. Netherlands: Springer, 2010: 251-268. https://doi.org/10.1007/s11229-009-9505-0.

Hempel, Carl G., and Paul Oppenheim. “Studies in the Logic of Explanation.” *Philosophy of Science* 15, no. 2 (April 1948): 135-175.

Hernandez, Jill Graper. *The New Intuitionism*. New York: Continuum International Publishing Group, 2011.

Hill, Thomas E., Jr. *Virtue, Rules, & Justice: Kantian Aspirations*. Oxford: Oxford University Press, 2012.

Horty, John F. *Reasons as Defaults*. Oxford: Oxford University Press, 2012.

Joyce, Richard. “Moral Fictionalism.” In *Fictionalism in Metaphysics*, edited by Mark Eli Calderon, 287-313. Oxford: Clarendon Press, 2005.

Jackson, Frank, Philip Pettit, and Michael Smith. “Ethical Particularism and Patterns.” In *Moral Particularism*, edited by Brad Hooker and Margaret Olivia Little, 79-99. Oxford: Clarendon Press, 2000.

James, Aaron. “Constructivism about Practical Reasons.” *Philosophy and Phenomenological Research* LXXIV, no. 2 (March 2007): 302-325.

Kant, Immanuel. *Groundwork of the Metaphysics of Morals*, edited by Mary Gregor and Jens Timmerman. Cambridge: Cambridge University Press, 2012.

Korsgaard, Christine. “Kant’s Formula of Universal Law.” *Pacific Philosophical Quarterly* 66, no. 1-2 (1985): 24-47.

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

Lance, Mark, and Margaret Little. “Defeasibility and the Normative Grasp of Context.” In *Contextualisms in Epistemology*, edited by Elke Brendel and Christoph Jäger, 293-313. Springer Netherlands (2005). https://doi.org/10.1007/1-4020-3835-6_18.

———. “Where the Laws Are.” In *Oxford Studies in Metaethics* 12, ed. Ross Shafer-Landau, 149-171. Oxford: Oxford University Press, 2017.

Leibowitz, Uri D. “Scientific Explanation and Moral Explanation.” *Nous* 45, no. 3 (September 2011): 472-503. https://doi.org/10.1111/j.1468-0068.2010.00775.x.

McKeever, Sean, and Michael Ridge. *Principled Ethics: Generalism as a Regulative Ideal*. New York: Oxford University Press, 2006.

———. “What Does Holism Have to Do with Moral Particualrism?” *Ratio* 18, no. 1 (March 2005): 6-34.

Millgram, Elijah. *Ethics Done Right: Practical Reasoning as a Foundation for Moral Theory*. New York: Cambridge University Press, 2005.

Prior, Elizabeth. *Dispositions*. New Jersey: Aberdeen University Press, 1985.

Rawls, John. “Kantian Constructivism in oral Theory.” *The Journal of Philosophy* LXXVII, no. 9 (September 1980): 515-572.

Rilke, Rainer Maria. *Duino Elegies*. Translated by Edward Snow. New York: North Point Press, 2001.

Rosen, Gideon. “What is a Moral Law?” In *Oxford Studies in Metaethics* 2, ed. Ross Shafer-Landau, 135-159. Oxford: Oxford University Press, 2007.

Ross, W. D. *The Right and the Good*, ed. Philip Stratton-Lake. New York: Clarendon Press, 2002.

Sandis, Constantine. “Dancy Cartwright: Particularism in the Philosophy of Science.” *Acta Analytica* 21, no. 2 (June 2006), 30-40.

Scriven, Michael. “Truisms as the Grounds for Historical Explanations.” In *Theories of History*, edited by Patrick Gardiner, 443-475. New York: The Free Press, 1959.

Swanton, Christine. *Virtue Ethics: A Pluralistic View*. New York: Oxford University Press, 2003.

Timmermann, Jens. “Kantian Dilemmas? Moral Conflict in Kant’s Ethical Theory.” *De Gruyter* 95, no. 1 (2013): 36-64. DOI: 10.1515/aghp-2013-0002.

Varden, Helga. “Kant and Lying to the Murderer at the Door… One More Time: Kant’s Legal Philosophy and Lies to Murderers and Nazis.” *Journal of Social Philosophy* 14, no. 4 (Winter 2010): 403-421.

Wallace, James D. *Moral Relevance and Moral Conflict*. Ithaca: Cornell University Press, 1988.