4

A Theory of Hedged Moral Principles

Pekka Väyrynen

1 INTRODUCTION

Moral theories have explanatory aspirations. They purport not merely to tell us *which* things are right and wrong, good and bad, and just and unjust, but also to explain *why* those things have the moral features that they do. Many theorists think that explanations which help us understand or make sense of morality must in some way rely on general moral principles. Suppose, for instance, that Anna ought to help Adam who is very badly off. Views which are "generalist" in this sense hold that Anna ought to do whatever it is that she ought to do in some sense because of some moral principles with suitable content, such as that one ought to help the badly off. So, many moral theories are committed to the existence of moral principles which can contribute to explanations of the moral features of things. But it seems often to be granted that many moral generalizations might have exceptions. It might be that Anna has no duty to help Adam if he is badly off through his own fault, or doesn't deserve help, or would channel the help to some heinous end. It is far from obvious what kind of principles could reconcile these two ideas or how principles which

Material from this chapter received valuable feedback from audiences at the Fourth Annual Wisconsin Metaethics Workshop in Madison, the "Ethics without Principles" conference in Paris, and Universities of Oxford, Stockholm, and Turku. From conversations on these and other occasions, I recall particularly instructive objections or suggestions from Christian Coons, David Copp, Michael Fara, Eric Hiddleston, Brendan Jackson, Robert Johnson, Stephen Kearns, Mark Lance, Sean McKeever, Bernhard Nickel, Russ Shafer-Landau, Ian Spencer, Sarah Stroud, and Paul Teller. Michael Ridge, Connie Rosati, and a number of anonymous referees for this volume and other journals deserve special thanks for generously sacrificing their time to prepare extremely helpful comments on earlier drafts. I cannot blame these colleagues, or anyone else with whom I have discussed these issues, for the long time this paper stayed in preparation or any mistakes that still remain. I would like to express my deep gratitude to them all.

purport to do so are something that moral agents can grasp, and which can guide them.

The aim of this chapter is to show how there can be moral principles which can play the explanatory and epistemological roles required by moral theory and yet are capable of admitting exceptions. I'll argue that principles which take a certain kind of "hedged" form can be of this kind and that such principles can be used to capture not only various familiar principles which have been claimed to hold without exception but also less orthodox sorts of principles which are claimed to permit exceptions. Elsewhere I argue that these hedged principles can account for the kinds of exceptions to which moral particularists appeal in support of their view and explain how they can provide practical guidance.[1] Here I develop the account in more detail. Section 2 describes what kind of theory I seek and an important problem it must solve. Section 3 develops an account of what makes an exception permissible. Section 4 shows how this account can be used to hedge principles so as to make them tolerate exceptions and how such principles can be used to capture a wide range of principles. Section 5 gives some quick arguments to show how hedged principles can contribute to explanations of various kinds of moral facts even if they tolerate exceptions. Section 6 sketches how these principles are something that moral agents can grasp, and which can guide the moral judgments of those agents. Section 7 explains what explanatory and epistemological advantages this account of hedged principles enjoys over certain rivals.

One caveat that I cannot avoid from the start is that many of the issues raised below are extremely complex both in their own right and with respect to how various issues in metaphysics, epistemology, and the philosophy of language, science, and mind bear on them. Since I cannot address all of these issues fully here, my discussion should be understood as an outline of a theory of moral principles. But even the outline will, I think, warrant the conclusion that we should find nothing peculiarly odd or problematic about the idea of exception-tolerating and yet explanatory moral principles.

2 THE PROBLEM ABOUT PERMISSIBLE EXCEPTIONS

Moral theorists sometimes assume that substantive moral principles (henceforth 'principles') must be not only explanatory but also universally applicable and exceptionless. But such "classical" principles are hard to

[1] See Väyrynen (2006) and (2008). The present chapter amends my sketch of hedged principles in those articles.

find.[2] For instance, the universal generalization corresponding to 'Promises ought to be kept' is 'Any promise ought to be kept,' but counterexamples to the latter are easy to find. Sometimes you ought to break a promise because only that way can you save a life.

But often we don't seem to think that only exceptionless generalizations can be explanatory. Many think that 'Ravens are black,' 'Acids are corrosive,' and (in the sorry case of supply-side economists) 'Tax cuts create jobs' can express explanatory generalizations but that not just any counterinstance falsifies them as they would falsify the corresponding universal claims. Similarly, many think that 'Promises ought to be kept' and 'Lying is wrong' can express explanatory moral generalizations but that, again, not just any counterinstance falsifies them. Some who think that lying while playing a game of bluff isn't wrong at all, for example, don't thereby deny that 'Lying is wrong' can express a true principle. This departs from a classical picture of moral principles.

If we deny that explanatory moral principles must be exceptionless, then it becomes possible that moral principles can be the sort of generalizations which we know can be explanatory even if their truth is compatible with the existence of exceptional counterinstances. For instance, 'Lying is wrong' might express a true principle even if some lies aren't wrong to tell or there is no reason not to tell them. Another typical feature of generalizations of this kind is that there is no specific proportion of instances which must conform to them for them to be true. So, at least in principle, 'Lying is wrong' could be true even if only few lies happened in fact to be wrong.[3]

According to this alternative view, not all instances of a principle must be equally relevant to its truth. We treat albino ravens as merely apparent counterexamples to 'Ravens are black' because they are irrelevant to the truth of the kind of generic claim which this sentence is naturally read as expressing. Only *some* ravens matter to its truth, and its truth requires that *those* ravens be black. So the core idea is that a generalization may count as a principle even if its truth is properly assessed only relative to some restricted range of instances. The truth of the principle that lying is

[2] Classical laws of nature seem to be no less hard to find. On this general message works like *How the Laws of Physics Lie* (Cartwright 1983), *Science without Laws* (Giere 1999), and *Ethics without Principles* (Dancy 2004) agree.

[3] The features listed here are commonly attributed to "indefinite" generic sentences, such as 'Turtles are long-lived' and 'A potato contains thiamine and vitamin C,' which are often used to express non-accidental but exception-tolerating generalizations. While it is no accident that sentences like 'Lying is wrong' are naturally read as generic, my purpose here isn't to offer a syntactic or semantic analysis of any moral sentences. So, although my discussion may occasionally seem to assume a particular analysis of generics to make things concrete, my purposes here require no particular analysis. For various analyses of generics, see Carlson and Pelletier (1995), Koslicki (1999), Liebesman (Ms), and Nickel (Ms).

wrong, for instance, may require only that the relationship between lying and wrongness (or reasons not to lie, or the like) be exceptionless within some, possibly restricted, range of lies. The principle may imply only that *in all but the permissibly exceptional cases*, lying is wrong.[4]

This alternative view requires principles that can permit exceptions to have a certain kind of complex structure. But it can be applied even to ordinary principles which are often taken to be explanatory while yet admitting of exceptions. We already know that it is in general possible for sentences of a certain degree of complexity to express or convey propositions of a greater degree of complexity. Nothing seems to make the moral case exceptional. So it seems safe to assume that even utterances of such simple moral sentences as 'Lying is wrong' can, in suitable contexts, semantically or pragmatically (depending on one's views in the philosophy of language) convey moral propositions which are more complex than the simple surface form of these sentences might seem to suggest. The kinds of complexity which context can help moral utterances to convey include not only such restricting clauses as 'in all but permissibly exceptional cases' but also constituents which qualify a moral proposition as an explanatory principle. So, for instance, I'll assume that an utterance of 'Lying is wrong' can, in a suitable context, convey a principle to the effect that actions which involve lying are actions one has moral reason not to do *in virtue of* their involving lying. To keep things simple, I'll assume more specifically that principles can be expressed or conveyed by sentences of the form '*G*s are *M*s' when '*G*' picks out a feature that provides reasons and '*M*' picks out a moral property.[5] I'll understand 'feature' loosely enough that sentences of this form include 'Promises ought to be kept,' 'Lying is wrong,' 'An action's being a lie is a reason not to do it,' and 'That an action would kill a person is a reason against doing it.'

What makes a moral generalization count as a substantive principle is controversial.[6] But everyone can agree that a generalization can play the explanatory role required by moral theory only if its content satisfies some sort of relevance constraint.[7] Principles should only specify conditions

[4] The general idea here can be found, in different forms, in Silverberg (1996: 215), Morreau (1997: 195), Braun (2000: 214), Fara (2005: 66), and Nickel (Ms), among other places.

[5] This is a simplifying assumption because the class of sentences that can in ordinary discourse be used to express principles is heterogeneous. It includes bare plurals, various kinds of conditionals, and more.

[6] It remains controversial even once we agree that logical tautologies which employ moral terms, infinitely long moral generalizations, and blatantly analytic moral truths don't count as substantive principles.

[7] Here is a quick example of this problem of explanatory relevance. Suppose all killings are morally wrong and all and only killers happen to have some distinctive physical mark.

which are in some sufficiently direct way relevant to the instantiation of moral properties to count as explanatory of their instantiation.[8] Moreover, since explanation is asymmetric, principles should only specify conditions on which the instantiation of moral properties depends asymmetrically.

The main task for any account of moral principles which purports to be explanatory in the above sense and yet capable of tolerating exceptions isn't to establish a distinction between instances that are relevant and those that are irrelevant to a principle's truth. Wide agreement exists, for example, that killing a person out of curiosity about how difficult it would be, or because of a bad mood, are no exceptions at all to the wrongness of killing. Wide agreement likewise exists that killing a threat in necessary self-defense is a permissible exception. So sometimes we already judge that *these ones* but not *those ones* are among the relevant instances.

The main task for any such account is, rather, to explain how principles can be explanatory if they permit exceptions and how it is that when we judge certain cases to be permissibly exceptional we needn't be just guessing, but our judgments can result from some more or less reliable ability to detect permissible exceptions. In what follows I first develop an account of what makes an exception permissible and then show how the account can answer these demands, and more.[9]

In that case 'If you were to kill someone and had a certain physical mark, then your action would be morally wrong' would give a true sufficient condition for wrongness. But this seems as much a paradigm example of a true but explanatorily defective generalization as 'If you were male and took birth control pills, you wouldn't get pregnant.' Having a physical mark seems no more relevant to the wrongness of killing than taking birth control pills is relevant to failing to get pregnant if one is male. (I don't claim that all accidental generalizations are non-explanatory; cf. Lange 2000: 16–18.)

[8] The content of this relevance constraint is also controversial. For instance, even among those who think that principles identify moral reasons for (or against) actions, policies, etc., some think that what principles thereby identify are sufficient conditions for the presence of moral reasons, whereas others think that there are different forms of moral relevance which adequate principles must reflect (e.g. being a moral reason vs. being a background condition which determines whether some other feature of an action counts in its favor, and to what degree). My account of moral principles will be neutral on this issue of how moral reasons are individuated. It is one aspect of the debate between "holism" and "atomism" in the theory of reasons. See e.g. Dancy (2004: 38–43) and Hooker (2008: 23).

[9] Related remarks concerning generics in Nickel (Ms) have aided my thinking here. I bracket two other worries as unproblematic. (1) Some worry that exception-tolerating generalizations might express no complete propositions (cf. Schiffer 1991). (2) Some worry that such generalizations might express complete propositions only with exception clauses which make them uninformative, however unobvious that may be without further analysis (cf. Pietroski and Rey 1995: 87). In ethics, Feldman (1986: 142–4) and Dancy (1999: 27), respectively, raise these worries. But neither is compelling to the extent that we already draw the distinction between relevant and irrelevant instances.

3 WHAT MAKES AN EXCEPTION PERMISSIBLE

An account of permissible exceptions can get going by observing that if something is a reason for (or against) something, then it is perfectly legitimate to ask why it is a reason, and a reason of that kind. To introduce terminology, when the fact that something would involve lying, for instance, is a reason not to do it, we can ask what is the "normative basis" of this fact's status as reason not to lie. By 'the normative basis,' I mean that factor (property, relation, condition) because of which the fact is a reason for (or against) performing the action, and which thereby explains why it is that kind of a reason in this instance. (Such analogous notions as the normative basis of a feature's status as right-making can be characterized in similar terms. I'll sometimes call reasons for doing something "positive reasons" and reasons against doing something "negative reasons.") I cannot here argue that every moral reason has a basis which makes it the kind of reason it is and explains its status as such.[10] But justification in ethics would threaten to be arbitrary unless at least generally there were an explanation of why something is a moral reason for (or against) something when it is, and why it isn't a moral reason for (or against) something when it isn't.

Thinking about exceptions in ethics in terms of this notion of a reason's normative basis turns out to have several advantages. It can be used to state a structural account of what makes an exception permissible. It can also be used to state a similarly structural account of how something gets to be a moral reason. Jointly these two will provide a satisfyingly unified account of why something is a reason for (or against) an action when it is, and why it isn't a reason for (or against) an action when it isn't. Finally, an advantage which is distinct from the previous two is that this notion of the normative basis can be used to articulate one particular form which genuinely explanatory and yet exception-tolerating moral principles could take. Or so I'll argue.

Let's begin with the account of permissible exceptions. The basic idea is that if something isn't a reason for (or against) doing something, this is a permissible exception to its status as such a reason when, and because, the normative basis of its status as such a reason is absent—when, and because, those factors fail to obtain in virtue of whose presence the feature would be a reason for (or against) doing what has it. The account is best developed through examples.

[10] I call this claim "the basis thesis," and give it some further support, in my (2006: 718–22).

Many who think that lying is wrong think that it can be permissible to tell a white lie about someone's appearance to bolster their self-confidence.[11] Their claim might be either that there is something wrong about such a lie but other considerations tell more strongly in favor of lying or that there is nothing at all wrong with such a lie. The former, weaker claim is no doubt preferable in many cases. But some think that the latter, stronger claim is preferable in others. Some think that there is no reason at all not to lie to a government death squad agent who is tracking down one's activist daughter.[12] Some think that there is nothing at all wrong about lying while playing the game Diplomacy, which is no fun unless the players lie rampantly. But if at least some of these cases really were permissible exceptions to the claim that something's being a lie is a reason not to do it, why might they be permissibly exceptional?

We can approach this issue by considering some toy theories about why we have reason not to lie. Let theory 1 say that an action's being a lie is a reason against it when, and because, lying contributes to undermining such beneficial social practices as trusting other people's word. And let theory 2 say that an action's being a lie is a reason against it when, and because, the addressee is owed the truth (or has a right to it, or lying violates her autonomy, or the like). Both agree that the status of this fact as a reason not to lie has a normative basis. And that is why both generate certain predictions about which exceptions to its status as a reason not to lie would be permissible. Theory 1 predicts that an action's being a lie is no reason at all against it when, and because, lying doesn't contribute to undermining a beneficial social practice (or else its contribution remains below some threshold). For example, there won't be any reason not to lie while playing Diplomacy if lying in that context has no (significant) bad spill-over effects on our trust in other people outside the context of the game. (Theory 1 also predicts that the more extensive or damaging these effects, the stronger the reason not to lie.) Theory 2 predicts that an action's being a lie is no reason at all against it when, and because, the addressee isn't owed the truth (or the like). One might hold that this is the case with Diplomacy insofar as the players have consented to playing knowing that it involves lying. And one might hold that government death squad agents have no right to information about activists' locations, given what they would do with it.

Theories 1 and 2 generate different predictions because they disagree about when and why an action's being a lie is a reason not to do it. But they agree on one thing: when an action's being a lie is a reason not to do it, its

[11] I owe this example to an anonymous referee.

[12] See Lance and Little (2007: 153–4). They attribute the Diplomacy example below to David McNaughton.

status as such a reason has *some* basis. According to theory 1, the proper basis for moral concern to avoid lying has to do with *sustaining beneficial social practices*. According to theory 2, it has to do with some such factor as *owing the truth to one's addressee*. They agree on another thing, too: the basis has a certain kind of structure. Each of the italicized phrases expresses a relational property which something may have if it involves lying and which may be morally significant in a way that can explain why an action's being a lie is a reason not to do it. The same structure appears in other examples. Accounts of why killing a person is wrong (or there is a moral reason not to do so, or the like) include that it *frustrates the victim's prudential interests*, that it *deprives the victim of future experiences that it would be valuable for her to have*, that it *manifests ill will*, and so on. On each of these views, the status of something's being a killing as a reason not to do it has a normative basis which explains this fact's contribution to what one has reasons to do, and each of them puts forward a candidate for what property fills that role.

These examples illustrate a general notion. When x is a lie, let "the designated normative basis" for the status of x's being a lie as a moral reason not to do x be that property P, whatever it is, such that x's being a lie is a moral reason not to do x when, and because, x instantiates P.[13] For instance, if something's being a lie is a moral reason not to do it, but not because it would *contribute to undermining a beneficial social practice*, then this letter property wouldn't qualify as the normative basis of a moral reason not to tell a particular lie even if that lie did instantiate it. If that were so, then theory 1 above would be incorrect. The designated normative basis of any other feature's contribution to some moral property can be characterized in the same way.

This is to define what sort of thing the designated normative basis is by its normative *role* in making something a reason for (or against) doing something. Schematically, an action's being G is a reason for (or against) doing it when, and because, the action, insofar as it is G, has some property which satisfies the above condition on P. A property satisfies this condition when it explains the status of the given fact as a moral reason. This definition doesn't stipulate properties into existence. It leaves open

[13] I take designation as a generic relation between a linguistic expression and what, if anything, it "stands for" or has as its "semantic value" (object, property, relation, function, etc.). I construe 'x instantiates P' loosely: 'x' can be satisfied by an act or its maxim or its agent, depending on P. I tried to capture the relational structure of normative bases by using the phrase 'the designated relation' in my (2006) and (2008). But this is unhelpful, and my talk of 'an action instantiating the designated relation' was sloppy. As I define 'the designated normative basis,' it typically picks out not a relation but a relational property: a role property whose realizers are the kinds of relational properties for which the italicized phrases in the text stand.

both whether something's being *G* in fact is a reason for (or against) doing it and whether that reason has a normative basis which makes it so, since 'the designated normative basis' is a definite description which may or may not be satisfied. The definition also leaves open just *which* property (if any) fills or realizes the normative basis role in the case of *being a lie* and *being a reason not to do an act*, and likewise for any other pair of features. Disputes about these issues belong to substantive moral theory. In short, then, the notion of the designated normative basis of a reason can be used to give a structural description of how something gets to be a moral reason, which can be common ground between different substantive views.

We can now state an analogous structural account of what makes an exception permissible. To ease comprehension, I state the account in terms of one of its specific instances:

(Perm) For any action *x* and any circumstances *C* such that *x* is a lie but this fact is no reason not to do *x* in *C*, *C* constitute a permissible exception to the status of something's being a lie as a moral reason not to do it when, and because, *x* fails to instantiate the designated normative basis of the status of something's being a lie as a moral reason not to do it.

If a lie fails to instantiate the relevant normative basis, this is because some feature of the circumstances operates as a "defeater" for the reason not to lie. But it counts as a defeater precisely when and because it makes the lie satisfy (Perm). So (Perm) specifies a condition because of which the circumstances are unsuitable, when they are, for the existence of a reason not to lie.

(Perm) gives the right results when plugged into theories 1 and 2 above. For example, if the relevant normative basis is the property of undermining autonomy, then (Perm) implies that a case where something's being a lie isn't a reason not to do it is a permissible exception when, and because, lying doesn't undermine autonomy. But it is important to note that we can derive specific conclusions about which cases, if any, are permissibly exceptional only once we conjoin (Perm) with some *substantive* view about what property fills the role of the designated normative basis.

(Perm) allows that a situational feature which in some other circumstances would prevent the designated normative basis from being instantiated may fail to do so in the presence of "defeaters for defeaters." These are, roughly, features which cancel the status of some other feature as a defeater that generates a permissible exception.[14] For example, if I threaten you with

[14] See esp. Horty (2007) for a valuable analysis of different kinds of defeaters and defeaters for defeaters.

force unless you promise not to do something that you are planning to do, this typically means that I am coercing you. Typically such cases seem to be permissible exceptions to the status of the fact that you promised not to do something as a reason for you not to do it. But coercion may still be just or permissible in some cases. Imagine someone who is planning to take another person's life or invade another country. It seems permissible to use threats of force to make them promise that they won't execute their plan.[15] In such cases, extracting a promise by threat of force needn't mean that the promise fails to instantiate the normative basis of the reason to keep promises.

It should be clear how to generalize (Perm) into a general account of permissible exceptions. All of the above points about (Perm) apply, *mutatis mutandis*, equally well to other moral reasons, to right- and wrong-making features and their normative bases, and further. They can also be accommodated by a wide range of theories of reasons, including theories which treat certain facts as "default" reasons. Moreover, nothing in this general account of permissible exceptions requires moral *principles*. Since moral particularists typically don't deny that moral facts have explanations, they can accept the above picture of moral reasons as entities which typically have a certain kind of basis and explanation. They should also be able to accept this picture of permissible exceptions as cases under which the designated normative basis of a reason fails to obtain. As we'll see, using this notion to develop an account of moral principles, as I do below, is a further move.

A caveat to (Perm) should make its compatibility with particularism clear. It seems logically possible that a property may be the normative basis of some fact's status as a reason for (or against) doing something even if the fact doesn't invariably function as a reason of that kind when the normative basis is instantiated. Suppose, for instance, that the normative basis of an action's being a lie as a reason not to do it is that being lied to undermines one's autonomy. Cases seem nonetheless possible where being lied to would undermine one's autonomy and yet the fact that the action would be a lie is no reason not to do it. One might sometimes deserve to be deceived in this way. I wish to allow that the status of a property as the designated normative basis may itself tolerate exceptions. In what follows, I'll understand this qualification to be implicit in (Perm).

This is no less a logical possibility if in many of these cases our preferred conclusion is that the property in question doesn't fill the normative basis

[15] See McNaughton and Rawling (2000: 270). One doesn't have to accept that in such cases threat of force would still count as coercion to accept the point that the example is meant to illustrate.

role after all. And it is a possibility we'll want to allow if we distinguish between non-derivative (ultimate, basic, primary) and derivative (subsidiary, secondary) reasons and principles.[16] For example, if the fact that you would waste another year there is a reason why you don't go back to Rockville, it would presumably be a derivative reason not to go back.[17] It would asymmetrically depend for its status as a reason on something like the longer-term harm to you from going back, which is a more basic reason not to go back. (Assume that wasting another year is something that would make you worse off.) If we draw this distinction, we should similarly distinguish derivative and non-derivative normative bases of reasons. It should be possible for factors because of which various facts count as moral reasons to be arranged in the kind of hierarchical relations in which explanations may in general be arranged. Furthermore, like definite descriptions in general, the expression 'the normative basis' is context-sensitive. So it may pick out the *most proximate* normative basis in some contexts, the *ultimate* normative basis in others, and perhaps something in between in yet others. At least the status of a property as a proximate normative basis could well be subject to permissible exceptions.[18]

The notion of the normative basis of a reason raises several yet further complications. Some of these are more usefully addressed later as concerns about my theory of moral principles. For now let me address three more immediate concerns about its role in the theory of reasons.

The first concern is that the notion of the normative basis is superfluous. Perhaps the theoretical work I assign to this notion can be done by distinctions we already have between different kinds of reasons. This might be a legitimate concern if something could serve as the normative basis of a reason only if it were itself a more basic reason from which the reason that is being explained is derived. But not all theories of reasons accept that all explanations of reasons must themselves be more basic reasons by another name. In some cases the normative basis which explains why something is a reason is better treated as just a condition for other things to be reasons.

One example is Kant's Categorical Imperative. Instead of thinking that the fact that a maxim violates the Categorical Imperative is itself a reason

[16] See e.g. McNaughton and Rawling (2000). McKeever and Ridge (2006: 130–4) argue that the distinction is a poor one. I intend my general account to be neutral on these family disputes in the theory of reasons.

[17] R.E.M., '(Don't Go Back To) Rockville' (IRS Records, 1984). For an example concerning principles, consider the discussion of "the duties of self-improvement" in Ross (1930: 21, 25–6).

[18] Such exceptions would be instances of (Perm) when the normative basis property itself provides reasons.

against acting on it, it may be more attractive to think of this as a condition for other features to count as reasons against acting on it. This move would enable Kantians to avoid the objection that they direct us to act on the wrong kind of reasons because they direct us to respond to an abstract rule instead of responding directly to the weal and woe of others. It would also allow Kantians to say that what *makes* wrong actions wrong is not their violating the Categorical Imperative but simply their being lies, killings, etc.

Another example is contractualism. It says that an act is wrong just in case any principle which permitted the act could, for that reason, reasonably be rejected (Scanlon 1998: 195). Instead of thinking that the fact that a principle permitting the act could reasonably be rejected is itself a reason against doing it, it may be more attractive to think of it as a condition for other features to count as reasons against doing it. This move would enable one to claim that the contractualist principle isn't redundant while agreeing that whenever a principle is reasonably rejectable because it permits actions which have feature F, those actions are wrong because they have F and not because their having F makes principles permitting them reasonably rejectable.[19]

I conclude that in developing a general structural account of reasons we shouldn't assume that the normative basis of something's status as a reason for (or against) doing something must itself be a more basic reason for (or against) doing it. This is a substantive assumption which is rejected by some theories which a general structural picture should accommodate. What is more, analogous assumptions in other domains are questionable. For instance, it seems better to say that the laws which relate things together as cause and effect are something in virtue of which causes have their causal powers, rather than that they must themselves be causes or parts of causes.

The second concern is that my definition of 'the designated normative basis' *trivially* entails that every moral reason has a normative basis. Suppose something's being a lie is a moral reason not to do it. Then *being an instance of lying which provides a moral reason against lying* might seem to count trivially as the normative basis of the status of something's being a lie as a moral reason not to do it. If something has this property, its being a lie is a moral reason not to do it.

It is far from clear that my definition has such trivial instances. Consider the property *being an instance of lying which provides a moral reason against lying*. Even if something's having this property entails that its being a lie is a moral reason not to do it, it is far from clear that this latter fact holds *because*

[19] For a related discussion of this "redundancy objection" to contractualism, see Stratton-Lake (2003).

of the former fact in any sense in which the former *explains* the latter.[20] But trivial instances might pose no real problem anyway. Substituting a property like *being an instance of killing which provides a moral reason in favor of doing it* into my definition would, as desired, typically still result in a falsehood.[21]

The third concern is that my picture of moral reasons and permissible exceptions applies generally only if every moral reason has a normative basis (non-trivially), but that this assumption leads to an infinite regress and disallows that there are any "brute" moral reasons whose status as reasons has no deeper metaphysical basis and can be given no further explanation.

No infinite regress follows, however. The normative basis of a feature's status as a reason needn't be in various senses distinct from that feature. It needn't be conceptually distinct. Some think, for instance, that it is part of the concept of cruelty that something's being cruel is a reason not to do it. In that case it would seem natural to say that the factor which explains this is some particular feature of the concept of cruelty. Nor need the normative basis of a feature's status as a reason be metaphysically distinct from it. Suppose, for instance, that the fact that something promotes well-being is a moral reason to do it. The normative basis of this reason won't be distinct from the fact in question if the property of being a moral reason is reducible to that of promoting well-being. If being a moral reason is nothing over and above promoting well-being, then one could explain why the fact that something promotes well-being is a moral reason to do it by pointing to this reduction. (The same might hold if these were one and the same property.) On this view, facts which provide moral reasons have *promoting well-being* as the normative basis of their status as moral reasons. In those cases where the fact which provides a reason to do something is, specifically, that it promotes well-being, it would seem natural to say that what grounds and explains the status of this fact as a moral reason is built into it as a matter of necessary but synthetic metaphysical truth.

It follows that a moral reason can be "basic" in one recognizable sense even if it has a normative basis: a moral reason can have an explanation even if it has no "deeper" or "further" explanation. Fundamental prima facie duties à la W. D. Ross (1930) could perhaps be analyzed as basic moral reasons in this sense. Ross appears to think that what is a fundamental prima facie duty is a feature which is intrinsically a moral reason for (or

[20] Even if the relevant notion of explanation allows that in some cases (cases of synthetic a posteriori property identity, perhaps) something's having one property can be explanatory of its having a differently described property that is identical with the first, these putatively trivial instances wouldn't seem to be cases of that kind.

[21] Thanks to David Copp for suggesting this point as well as the concern outlined in the previous paragraph.

against) doing what has it. For something to be intrinsically *F* is for it to be *F* solely in virtue of its intrinsic features. If, further, the relation expressed by 'in virtue of' is an explanatory one, then an action type's status as a fundamental prima facie duty can perhaps have an explanation in terms of its intrinsic features. Features which are moral reasons intrinsically may even be another case where the normative basis of a feature's status as a reason isn't distinct from that feature. In any case, it doesn't follow directly from the notion of a fundamental prima facie duty that such duties have no normative basis.

Some theorists may still insist that some moral reasons are genuinely brute in some stronger sense which does preclude their having normative bases.[22] Whether any moral reasons of this kind exist is a substantive issue not to be settled by definitional fiat. My definition of 'the designated normative basis' doesn't entail in any non-trivial way that every moral reason has a normative basis, and I don't take this to be an analytic truth on any other ground either. So here is where things stand with such theorists. On the one hand, if there are reasons which have no normative bases, then the above picture of reasons and permissible exceptions may be unable to accommodate them. But, on the other hand, those who claim that there are such reasons are to that extent unable to exploit the explanatory and epistemological advantages of this picture of reasons and permissible exceptions and the theory of moral principles which can be developed out of it.

I have argued that the notion of the normative basis of a reason can be used to state structural accounts of how something gets to be a moral reason and what makes an exception to its status as a reason permissible. It is worth noting that this account is unified in a particularly satisfying way: what explains why something is a reason for (or against) an action, when it is, is the presence of precisely that factor whose absence explains why the circumstances are permissibly exceptional, when they are, and so explains why it doesn't function as such a reason, when it doesn't.

4 AN ACCOUNT OF EXCEPTION-TOLERATING MORAL PRINCIPLES

I'll now argue that the notion of the normative basis of a reason can also be used to state an account of moral principles as a kind of "hedged" principles. Substantive principles concerning moral reasons can be captured

[22] Thanks to Sarah Stroud for pushing me on my response to theorists who take this kind of line.

by principles which are hedged by reference to the normative bases of those reasons.[23] For instance, if something's being a lie is a reason not to do it when it instantiates the designated normative basis for the reason not to lie, then we cannot render the implications of the generalization 'Lying is wrong' for when something's being a lie is a reason not to do it any *less* accurate if we hedge it by reference to the designated normative basis, like so:

(Lie) Something's being a lie is always a reason not to do it, provided that it instantiates the designated normative basis for this fact's status as moral reason not to lie.

(Lie) takes no stand on which property, if any, fills the designated normative basis role. It requires just that there be a property whose instantiation by a lie explains why its being a lie is a reason not to do it.[24] If there is no such property, then 'the designated normative basis' fails to refer, in which case (Lie) either is false or lacks truth value. So, as with (Perm), we can use (Lie) to derive conclusions about which instances of lying provide reasons not to lie, and which if any are permissibly exceptional, only once we conjoin (Lie) with some substantive view of what the relevant normative basis is. For example, if lying sometimes fails to treat someone with respect without contributing to undermining a beneficial social practice, then our toy theories from Section 3 disagree over which lies we have moral reason not to tell. But both could still accept (Lie). So (Lie) is acceptable to a variety of moral theories which make injunctions against lying.

Accordingly, the account of principles which we get by generalizing (Lie) implies no particular view about which features provide reasons for (or

[23] I'll focus mainly on principles which identify moral reasons to keep the discussion manageable and focused on explanatory principles. If something is a moral reason for (or against) an action, then it has got something to do with what explains the moral status of that action. But the account I'll develop will also be able to capture principles which don't directly concern moral reasons, at least on further plausible assumptions concerning the relationship between reasons and other normative properties. For example, it is often assumed that if a feature of an action is a reason for (or against) doing it, then it is right (or wrong) *pro tanto*, so far as its having that feature goes. If that is right, then principles which identify moral reasons can be used to capture principles concerning the rightness and wrongness of actions and principles which identify "right-making" and "wrong-making" factors. For, even though the reason relation and the right-making relation seem distinct (Dancy 2004: 79), it seems plausible that if, for example, something's being a lie makes it wrong, then its being a lie is a reason (for suitably situated agents) not to do it. (The converse fails: not all reasons against an action are features which make it wrong.)

[24] My earlier caveat that the status of a property as the normative basis may itself be subject to permissible exceptions applies to (Lie) as well. So I'll continue to implicitly allow the possibility that something is a lie and instantiates the relevant normative basis and yet its being a lie *permissibly* fails to be a moral reason not to do it.

against) performing actions that have them, and why. Far from aiming to supplant familiar consequentialist, deontological, and other substantive views, the account is a schema for a form that various substantive principles could take:

(HP) Any x that is G is M [e.g., x's being a lie is always a moral reason not to do x], provided that x instantiates the designated normative basis of G's contribution to M.[25]

As a mere schema, (HP) doesn't entail that there *are* true moral principles which are explanatory but can admit of exceptions. The purpose of articulating one form which such moral principles could take isn't to help us determine the truth value of particular principles of that form—at least not on its own. But, as I'll argue, the schema can be used to show how there *can be* such principles, and also to locate competing substantive moral theories in a common structural framework.

So how are principles of the form (HP) capable of tolerating exceptions? What determines whether such a principle tolerates exceptions is the property which actually fills the relevant normative basis role. But the principle doesn't say what property, if any, does so—at least not on its own. For instance, (Lie) in conjunction with (Perm) implies only that, in all but permissibly exceptional cases, something's being a lie is a moral reason not to do it. It remains possible that the set of permissibly exceptional lies is empty. If the relevant normative basis were such that every lie instantiates it, then (Lie) would tolerate no exceptions. For example, in the context of Kant's moral theory, (Lie) implies that something's being a lie is a reason not to do it when lying to a person fails to treat her as an end in itself. According to a rigorist version of Kant's theory, lying always constitutes this kind of assault on rational agency. In the context of this rigorist version of Kant's theory, (Lie) implies that its injunction against lying has no permissible exceptions. Thus (Lie) *doesn't require* the existence of permissible exceptions.

[25] Principles needn't ordinarily be stated using the proviso in (HP) or other hedging expressions (e.g. 'other things being equal' or 'normally') to make restrictions on their scope. Our earlier assumption that the proposition expressed may be more complex than the sentence expressing it (see Section 2) secures the possibility that propositions of the form (HP) can, in suitable contexts, be semantically or pragmatically conveyed simply by sentences of the form 'Gs are Ms.' I'll remain neutral on the precise logical form of these propositions. It affects nothing of substance whether the proviso in (HP) is to be treated as a propositional operator, part of the antecedent of a conditional, or some other kind of clause. I also cannot examine whether such non-moral generic claims as 'Turtles are long-lived' or 'Dogs are smaller than horses' can be captured by propositions of the form (HP). But my account doesn't require this.

But (Lie) *allows* the existence of permissible exceptions, since it allows the possibility that not all lies instantiate the relevant normative basis. If (Perm) is right that a permissible exception arises when, and because, a lie doesn't instantiate the relevant normative basis, and if some lies don't instantiate it, then (Lie) permits those lies as exceptions. They would be just the lies that violate the proviso in (Lie). Thus (Lie) also tells us what the exceptional lies would have in common.

To summarize: (Lie) hedges the claim that lying is wrong by reference to the normative basis of the status of something's being a lie as a moral reason not to do it. Whether (Lie) permits any exceptions depends on whether the property which actually fills the normative basis role is such that all lies instantiate it. It is because (Lie) alone takes no stand on what that property is that it allows, but doesn't require, the existence of permissible exceptions. Since (Lie) isn't special, we can generalize that principles of the form (HP) can be used to model both exceptionless and exception-tolerating principles. We can plug into such principles my account of permissible exceptions, whose particular instances will play the same role in principles of the form (HP) as (Perm) plays in (Lie). Each states a condition under which a feature's failure to provide a reason for (or against) doing what has it would be a permissible exception to the corresponding principle.

This account is the more broadly applicable, the greater the range of principles that can be construed as instances of (HP). We saw how it can model both exceptionless and exception-tolerating principles. It can also capture many other distinctions between different kinds of principles. Here I'll focus on two in particular, and address various objections to the account in the process.

One dimension of difference between principles which my account can capture concerns whether or not they are derived from other, more basic principles, and so whether or not they have independent normative weight. One can accept (Lie) irrespective of which kind of principle one takes 'Lying is wrong' to express. To illustrate, Kantians would presumably accept (Lie) only if they thought that lying to people is a way of failing to treat them as ends in themselves. Suppose they think this "Kantian property" is the normative basis of the status of something's being a lie as a moral reason not to do it. If they also thought that something's having the Kantian property is itself a reason not to do it, they would presumably think that (Lie) is derived from (End):

(End) An action's failing to treat someone as an end in itself is a reason not to do it, provided that it instantiates the designated normative basis for the status of something's having this Kantian property as a moral reason not to do it.

(End) is a hedged principle which captures the old, familiar view that the "real" reason not to lie is that lying to people is a way of failing to treat them as ends in themselves. Other possible views can also be captured on my account. One is the view that the status of something's failing to treat someone as an end in itself (or, indeed, its being a lie) as a moral reason not to do it has some such further explanation as that you cannot consistently will the maxim of such an action as a universal law. Another is the view that (End) is a non-derivative principle. This option is secured by the possibility that the Kantian property provides basic moral reasons whose normative basis isn't distinct from that property. Essentially the same menu of theoretical options will be available if we think that the wrongness of lying has something other than a Kantian explanation.

One might object that, in fact, my account has trouble capturing derivative moral principles. Letting '*DNB*' stand for the property, whatever it is, which fills the designated normative basis role, a principle of the form (HP) says that *G*s are *M*s, provided they are *DNB*s. Given what I have said about hedged principles and normative bases, such a principle seems to commit one to the following two counterfactuals concerning any action *x* that has *G*, *M*, and *DNB*:

(C1) If *x* had been *G* but not *DNB*, *x* wouldn't have been *M*.

(C2) If *x* hadn't been *G* but had nonetheless been *DNB*, *x* would still have been *M*.

The objection is that if (C1)–(C2) are true of *x*, then *x*'s being *G* is epiphenomenal with respect to its being *M*. To illustrate, suppose *G* is the property of being a lie, *M* is the property of being wrong, and *DNB* is the property of betraying trust. If so, then it would seem that what really is wrong with lying is that it betrays trust. But shouldn't we in that case think that the corresponding principle (Lie) is false?[26] And if (Lie) is false, it cannot be rescued by treating it as a derivative principle.

This objection shows that certain hedged principles are false. Often when (C1)–(C2) are true of something, it is plausible that what makes it *M* isn't its being *G* but its being *DNB*. If so, then any hedged principle which purports to identify *G* as right or wrong making, and so implies otherwise, is false. But these are substantive claims which pose no problem for my structural account. And the objection fails to generalize in ways which would pose a real problem for my account.

[26] I am grateful to Robert Johnson for this objection. I should note that the objection had greater force against an earlier version of this chapter, which focused more heavily on principles that identify right- and wrong-making features.

One sort of hedged principles which escape the objection are those in whose case *DNB* isn't distinct from *G*. In that case (C1)–(C2) have impossible antecedents and hence count only as vacuously true. Another sort of hedged principles which escape the objection are those which do purport to identify *G* as a right- or wrong-making feature but in whose case *DNB* is best treated as a condition for other features to be right or wrong making. Examples include principles which would face a Euthyphro-type problem if *DNB* were treated as a feature that is itself right or wrong making. This is the sort of reason why contractualists, for example, typically deny that what *makes* wrong actions wrong is that they are permitted by a principle that could reasonably be rejected. My account does commit such a contractualist to holding that (C2) is true of an action which is *G* and wrong and has this contractualist candidate for *DNB* only if in those nearby possible worlds where the action has *DNB* and is wrong it also has some feature other than *G* which satisfies the condition set by *DNB*. But this seems not unreasonable. Nothing can have the contractualist basis property brutely, without having some other feature.

A wide range of derivative hedged principles also escape this objection. I have in mind principles which purport to identify reasons provided by features that aren't right or wrong making. What makes going back to Rockville a bad idea isn't so much that you would waste another year as that this would be a way of doing something that is bad for you. But this doesn't mean that the fact that you would waste another year doesn't count as a (derivative) reason not to go back. Something's being *G* can be a derivative reason not to do it even if (C1)–(C2) are true of it, if doing something that is *G* is a way of doing something that one has a non-derivative reason to do.[27]

[27] Stephen Kearns worried that, in fact, my account delivers *too many* derivative principles. Suppose I promise not to go skiing, but go anyway. One might claim that the property *breaking a promise not to go skiing* satisfies my definition of 'the designated normative basis': that something involves going skiing is a reason for me not to do it because it would break a promise not to go skiing. So my account delivers a true principle concerning each thing that one might promise to do. But such principles seem at worst false (what gives the reason isn't that I go skiing but that I break a promise) and at best wholly unnecessary. But this objection fails. If we want an account that can model both derivative and non-derivative principles and *if* some features which seem morally trivial nonetheless function as derivative moral reasons in some type of circumstances, then delivering a derivative principle which identifies those features as moral reasons just in those circumstances is hardly objectionable. I emphasize the second 'if' because the objection generalizes only under certain substantive views on reasons. It seems to require that if something's being *F* is a reason to do it and doing something that is *G* is a way of doing what is *F*, then its being *G* is a derivative reason to do it. Whether this holds generally is controversial. For instance, in certain trolley cases killing one is causally or constitutively close enough to saving five that it may count as a way of saving five. But it is controversial that the fact that a certain action would involve killing someone is even in such cases a derivative reason to do it. It is similarly controversial that such seemingly trivial features as shoelace color come to function as derivative reasons when they are suitably connected to features which are agreed to give reasons. My account doesn't require views at this level of specificity concerning the individuation or derivation of reasons.

Another dimension of difference between principles which my account can capture concerns their strength. For instance, 'Lying is wrong' may express either an "overall" principle that lying is wrong all things considered or a merely "contributory" principle that lying is wrong *pro tanto*, that is, so far as its being a lie goes. Something's being a lie can be a reason not to do it without determining that, overall, one ought not to do it. What one ought to do overall is some function of those factors which make some moral contribution plus the strengths of their contributions.[28]

Hedged principles capture the difference between contributory and overall principles in the strength of the normative basis role which they describe. Read as an overall principle, (Lie) entails that the designated normative basis is such that whenever lying instantiates it, one has decisive or most moral reason not to lie. Read as a contributory principle, (Lie) entails only that the designated normative basis is such that whenever lying instantiates it, one has some moral reason not to lie. These two conditions on the normative basis are distinct. The difference between them also determines whether the principle in question can permit as exceptions only lies which one ought, overall, to tell, or also lies which one has no reason at all not to tell.[29]

The distinction between overall and contributory principles is sometimes used to reconcile the appearance that principles permit exceptions with the view that genuine principles must be exceptionless. Even if sometimes lying isn't wrong overall, it might still always and invariably make some contribution to wrongness. In this way, one could claim that, although

[28] I take no stand on the nature of this function here. But see e.g. the two deontic logics for modeling the calculation of "all things considered oughts" from the relevant "*prima facie* oughts" in Horty (2003). I cannot here go into the complications which arise from the possibility that some features which don't themselves count as reasons may yet intensify or diminish the strength of the reasons given by other features. See e.g. Dancy (2004: 42).

[29] An intriguing question which I cannot pursue here is whether the two types of permissible exception could be analyzed by adapting from epistemology the distinction between "rebutting" and "undermining" (or "undercutting") defeaters for evidence or reasons. See e.g. Pollock and Cruz (1999). Jonathan Vogel (in conversation) and Horty (2007: 15) suggest, to my mind plausibly, that undercutting defeat can be analyzed as a special case of rebutting defeat.

all true overall principles are false because they have exceptions, there are exceptionless contributory principles.[30]

But even if all true principles turn out, at the end of the day, to be exceptionless, a general account of principles should accommodate the possibility of contributory principles which can tolerate exceptions. Nothing seems to rule out the idea of such principles as incoherent.

This point has implications for the dialectic between moral generalists and particularists. Many particularists are moved by the thought that even putative contributory principles have exceptions. Clearly this supports particularism only on the assumption that genuine principles must be exceptionless. If this assumption is dubious, then what particularists should deny is not the possibility of true exception-tolerating principles but rather the existence of any comprehensive set of such principles or else the dependence of moral reasons on their existence (cf. Dancy 2004: 7–8).

My account of hedged principles helps us see what substantive issues are at stake in these claims. I discuss these issues in detail elsewhere (Väyrynen 2006). The only point I wish to note here is that the standard semantics for definite descriptions like 'the designated normative basis' assigns to any principle of the form (HP) the substantive commitment that it is true only if there is a *unique* property which fills the designated normative basis role.

Particularists are likely to think that what explains why something's being a lie is a reason not to do it can be one factor in some cases, another factor in other cases, and some yet different factor in yet other cases. This view is consistent with the claim that moral reasons have normative bases but it seems inconsistent with the uniqueness condition. Its availability confirms that using the notion of a reason's normative basis to state an account of moral principles is indeed a further move from the picture of reasons and permissible exceptions introduced in Section 3.

The general objection here to my account of hedged principles is that it will systematically generate false principles because it will systematically fail the uniqueness condition. So we should either reject my account of principles in favor of a better one or accept particularism.

[30] Ross (1930: ch. 2) is usually read as holding this view. In a later work Ross argues from cases of vicious pleasure that pleasure isn't intrinsically good, and seems to conclude that we have no prima facie duty to promote our own pleasure (Ross 1939: 272–5; cf. Stratton-Lake 2002: 130–4). Note that this seems to follow only on the assumption that a prima facie duty to promote our own pleasure couldn't permit vicious pleasures as exceptions.

To see one form of this objection, suppose there are two kinds of promises.[31] Imagine that one kind of promise gives a reason to keep it because it would be irrational to break, for no good reason, a voluntarily undertaken commitment. And imagine that the other kind of promise gives a reason to keep it because the promisee has some such right as to determine that one do what one promised or to receive the fruits of the promise. Presumably what would explain why I ought to keep my promise would be different in the two cases: perhaps something about rationality and autonomy in the first but something about the promisee's rights in the second. But then the principle that one ought to keep one's promises would seem not to designate a unique normative basis. Particularists might take this as conformation that no such true principle is to be had. Generalists might think instead that surely *this* wouldn't show that it isn't true that promises ought to be kept.

My own response is that the example involves two distinct principles of promissory obligation. The sentence 'Promises ought to be kept' can be used to express one proposition in the context of one kind of promises but another, different proposition in the context of the other kind of promises. First, if there were two different kinds of promises, then the semantic value of 'promise' would vary with context of utterance, which seems to mean that so would the principle expressed. Secondly, the two kinds of promises would also differ in their implications for when one ought to keep a promise. Two propositions are distinct if they have different implications. So if there were two different kinds of promises, then 'Promises ought to be kept' would in different contexts express different principles. Nothing in the example shows that those two principles don't each designate a unique normative basis. (*Exercising rational autonomy* might fill that role in one principle, *satisfying the promisee's right to determine what the promisor does* in the other.)

This response exploits the fact that the uniqueness implications of definite descriptions are notoriously subtle and context-sensitive to suggest that the uniqueness implications of hedged principles are going to be no different. What it offers is not an advance proof that hedged principles won't systematically fail the uniqueness condition, but rather a conceptual tool for substantive moral inquiry to use in assessing whether particular principles of the form (HP) fail it. Elsewhere I apply this idea to show how my account could approach the objection that the normative bases which hedged principles designate will be disjoint in the sense that these role properties will systematically be realized by different properties in different contexts (Väyrynen 2006: 733–4).

[31] Thanks to Ralph Wedgwood for pressing this example. I have changed some inessential details.

I conclude that we can capture a wide range of substantive moral principles in the kind of hedged principles introduced in this section. I also hope to have conveyed some sense of the resources this account of hedged principles can muster up for explaining whatever further distinctions we might want to draw among various kinds of moral principles. Next I'll argue that hedged principles can play the explanatory and epistemological roles required by moral theory: they can contribute to explanations of particular moral facts and moral agents can grasp and be guided by them.

5 HEDGED PRINCIPLES IN EXPLANATION

Moral theories aspire to explain various kinds of moral facts.[32] The less a moral theory helps make sense of such facts, the more epistemically imperfect state it leaves us in regarding important aspects of morality. Any account of moral principles can be expected to indicate the role which the kinds of principles that it proposes can play in explaining particular moral facts. I'll now argue that moral principles can be genuinely explanatory even if they permit exceptions.

There is a general objection to the possibility of such principles. Principles that permit exceptions tend to be compatible with the possibility that even some large proportion of their instances are permissibly exceptional. Hedged principles, for instance, don't determine what proportion of their instances instantiate the normative bases they designate because they don't themselves determine such contingent facts as how frequently lies aren't owed to the addressee, how often killings are done in necessary self-defense or just for fun instead, and so on. The objection is that principles which allow this possibility cannot be genuinely explanatory because principles can provide reliable explanatory applications only if they have some large proportion of conforming instances.[33]

[32] By 'explanation' I mean the content of an answer to a why-question which makes a claim that something is the case *because* something else is the case—that this something else is at least part of *why* it is the case. (I don't mean the activity of giving such an answer. What it takes to convey the content to an audience is a topic for the pragmatics of why-questions.) We may need to add that something counts as an explanation only if it also satisfies certain epistemic conditions. For example, it may be that the content of an answer to a why-question counts as an explanation only if it is (or represents) a body of information which is structured in such a way that grasping that body of information would constitute a certain kind of epistemic gain regarding what is being explained.

[33] Earman and Roberts (1999: 463) endorse effectively just this when they claim that *ceteris paribus* generalizations provide reliable applications only if the *ceteris paribus* are *paribus* in "sufficiently many" applications of the corresponding unqualified universal generalization.

I'll argue in reply that hedged principles can be used to explain both why their conforming instances have the moral properties they do and why certain other instances are permissibly exceptional, irrespective of their proportions.[34] I wish the argument to work under reasonably neutral assumptions about explanation. In seeking an explanation we are seeking understanding and trying to make sense of things. We can agree that an explanation of a fact F contributes to these aims insofar as it is stable or robust in the sense that according to this explanation, F couldn't easily have failed to obtain. In other words, we can agree that it is a virtue in an explanation of a fact F if, according to this explanation, F is stable under some range of hypothetical changes in the circumstances and if, therefore, the explanation exhibits a pattern of counterfactual dependence.[35] We can take a generalization to inherit this explanatory virtue insofar as it plays some important role in explanations which have that virtue. And we can agree that if factor F counts as (part of) an explanation of a fact G only given some further factor H, then H plays an important kind of explanatory role. Let's say that if F explains G, then H "contributes to" this explanation if H is either part of F or some important condition for F to explain G. What I'll argue is that hedged principles can in this sense contribute to explanations of particular moral facts. The general idea will be that these moral facts stand in systematic patterns of dependence on other factors, so that the latter explain the former, only given some moral principles of the form (HP).

Hedged principles contribute to two kinds of explanations concerning their conforming instances. First, (Lie) can allow that when something is a lie, this can explain why there is a moral reason not to do it.[36] But it explains

[34] Sean McKeever and Michael Ridge propose that a moral principle "articulates true application conditions for a given moral concept by referring to those features of the world which explain why the concept applies when it does" but continue that "to count as a moral principle on this criterion a generalization need not explain why those considerations which are of direct relevance themselves count as having such relevance" (2006: 6). If we limit ourselves to standards for the correct application conditions for moral concepts, then our theoretical purposes don't require that principles contribute to explanations of moral reasons and permissible exceptions. But this is compatible with thinking that an account of moral principles is better if its principles also contribute to explanations of such facts.

[35] On explanatory stability, see White (2005). On explanations which exhibit systematic patterns of counterfactual dependence, see Lange (2000) and Woodward (2003). These ideas can be accommodated by the view that explanations track what makes things happen or makes something the case, insofar as these metaphysically more robust relationships imply corresponding patterns of counterfactual dependence. See Ruben (1990) and Kim (1994).

[36] My account can accommodate the claim that hedged principles, and facts to the effect that the relevant normative basis is instantiated, may enter into explanations of particular moral facts as background conditions on which those explanations rely without

this only when the instance of lying at hand instantiates the designated normative basis. Given the way (Lie) incorporates a reference to that basis, something's being a lie explains the existence of a moral reason not to lie only given a principle like (Lie). So, generalizing, hedged principles contribute to explanations of why there are moral reasons to do certain things but not others. Secondly, (Lie) implies that the fact that something's being a lie is a moral reason not to do it holds only when it instantiates the designated normative basis. But, given the way (Lie) incorporates a reference to that basis, the latter explains the former only given a principle like (Lie). So, generalizing, hedged principles contribute to explanations of why certain facts but not others have the status of moral reasons for (or against) doing various things.

The natural objection is that (Lie) is superfluous in these explanations because what in fact makes the contribution which these arguments attribute to (Lie) is the designated normative basis. Yet this is at most half correct. Facts about whether lies instantiate the relevant normative basis do contribute to these explanations. But this doesn't mean that hedged principles make no further contribution to them. The designated normative basis is just a property which particular lies instantiate or not. It exhibits systematic patterns of counterfactual dependence between whether something is a lie and whether there is a moral reason not to do it only when embedded in a generalization like (Lie). What (Lie) asserts is precisely a complex but systematic relationship of dependence between these two factors and the designated normative basis. It asserts a connection between something's being a lie and there being a moral reason not to do it which is stable under any hypothetical changes under which it still instantiates the designated normative basis, but which might not hold outside this range of conditions.[37] So the designated normative basis explains moral reasons in a systematic way only given a principle like (Lie). Hence this objection doesn't show that hedged principles fail to contribute to explanations of moral reasons.

Hedged principles also contribute to explanations of permissible exceptions. (Lie) implies that instances of lying which are permissible exceptions

constituting parts of those explanations (cf. Dancy 2000: 152). Since I can, therefore, deny that explanation requires that the explanans be sufficient for the explanandum, I can also avoid the potential objection that if a principle like 'Lying is wrong' permits exceptions, then we cannot appeal to it and the fact that you lied to explain why your action was wrong (cf. Pietroski and Rey 1995: 87).

[37] That is to say, changes in other conditions matter only to the extent that they are relevant to whether a lie instantiates the designated normative basis, and if there were a moral reason not to lie outside this range of conditions, this would be due to some other factor. Note that this sort of invariance doesn't imply exceptionlessness.

to (Lie) are permissibly exceptional because they fail to instantiate the designated normative basis. To illustrate, suppose the status of something's being a lie as a reason not to do it is based on the way in which lying contributes to undermining a beneficial social practice. If, as some writers claim, there may sometimes be no reason not to lie to a person who is going to harm innocent people, (Lie) can contribute to explaining why. If you had such a reason, then in some circumstances one could generate for you a duty not to lie simply by aiming to harm innocent people and coming to you for information one needs to achieve that aim. But a social practice which involves such a mechanism could hardly be said to be a beneficial one. Thus in lying to such a person you wouldn't be undermining a beneficial social practice. But, given the way in which (Lie) incorporates the designated normative basis, the failure of that basis to be instantiated explains why circumstances are permissibly exceptional only given a principle like (Lie). So, generalizing, hedged principles contribute to explanations of permissible exceptions.[38]

The natural objection to this argument is that the explanatory contribution it attributes to (Lie) in fact belongs to (Perm). Again, this is at most half correct. Both (Perm) and (Lie) imply that when something is a lie, its being a permissible exception to the status of its being a lie as a moral reason not to do it is invariant under any hypothetical changes under which the lie still fails to instantiate the designated normative basis. But there can be this kind of systematic pattern of counterfactual dependence for (Perm) to exhibit only if there is a systematic pattern between something's being a lie and there being a moral reason not to do it. As we just saw, (Lie) exhibits just such a pattern. So, the failure of a particular lie to instantiate the designated normative basis explains why it is a permissible exception only given a principle like (Lie). Hence this objection doesn't show that hedged principles fail to contribute to explanations of permissible exceptions.

We can now see why hedged principles contribute to explanations of permissible exceptions irrespective of the proportion of the permissibly exceptional instances. Suppose killing in necessary self-defense is a permissible exception to the wrongness of killing. According to the corresponding hedged principle, this is so when, and because, killing in necessary self-defense fails to instantiate the normative basis of killing's contribution to wrongness. But it is similarly when, and because, killing instantiates *that very same property* that it is wrong. So, the principle can be used reliably to

[38] Baldwin (2002: 104–6) argues from examples that principles are typically qualified in order to explain exceptions. Irwin (2000: 121) argues that Aristotle has a notion of the basis of a principle which explains exceptions. But neither provides a general account of principles to develop these claims in detail.

explain why killing is wrong, when it is. This is so even in a Mad Max world, where most killings are done in necessary self-defense, because the principle can also be used to explain why most killings are permissible exceptions in the Mad Max world.[39] Given the facts on the ground, killings in the Mad Max world couldn't easily have instantiated the normative basis of killing's contribution of wrongness.

I have argued that hedged principles make a genuine contribution to explanations of particular moral facts. Their contribution is also satisfyingly unified in character: whether we are explaining moral reasons or permissible exceptions, hedged principles exhibit stable patterns of dependence between these moral facts and the normative bases which they designate. Much more could and needs to be said about the contribution of hedged principles to explanations of particular moral facts if their contribution were assessed against different theories of explanation. But I hope that already these quick arguments are, for the present purposes, sufficient to show that hedged principles can be genuinely explanatory even if they permit exceptions.

These arguments require an obvious caveat, however. All by themselves, without further substantive assumptions about what properties fill the normative basis roles, hedged principles omit a whole lot of information concerning the actual factors because of which certain of their instances provide moral reasons and others are permissibly exceptional. Those factors are given only a relatively formal kind of role description. So, all by themselves, hedged principles make only a thin and limited contribution to explanations of moral facts. But this caveat raises no deep problem.

Since my account aims to articulate only a particular form which explanatory and yet exception-tolerating principles could take, it should be no surprise if particular principles of this form turn out to be explanatory only in the context of further substantive moral assumptions. But this is the kind of context in which we typically operate when we assess competing moral theories or when it is for some other reason important to know whether, for instance, we have a moral reason not to lie because lying betrays trust, or because it undermines a beneficial social practice, or because it fails to give the addressee something owed to them, or because of something else.

Hedged principles may also be able to contribute to explanations of particular moral facts in virtue of their general form. For instance, perhaps all that is required in some contexts to explain why something's being a lie is a moral reason not to do it is that the relationship between these two factors is stable within some range of circumstances and that we have good reason to believe that the present circumstances fall within that range. The first

[39] I owe the Mad Max world to Sean McKeever.

claim follows from the general form of (Lie). The second claim can at least in some contexts be supported by fairly neutral and minimal substantive assumptions which require no particular view about what property fills the normative basis role. These could concern, for instance, some implications or other identifying characteristics which we have good reason to believe to be possessed by whatever property fills the normative basis role. This sort of information is the most we often have available anyway, since we often are uncertain, ignorant, or agnostic about just what is wrong about lying, killing, and so on.

The kind of limited contribution which hedged principles can make to explanations of particular moral facts in virtue of their form can be a genuine contribution at least if hedged principles themselves can be exploited to improve our sense of the implications and other identifying characteristics of the properties which fill the relevant normative basis roles. This would improve our sense of the range of conditions under which particular moral facts hold and how those facts might have been different had the conditions been different. If hedged principles played such a role, they could help exhibit concrete patterns of dependence between particular moral facts and other factors. Hence I now turn to discuss the epistemological role of hedged principles in moral inquiry.

6 HEDGED PRINCIPLES IN MORAL INQUIRY

Moral theory would require moral principles not only to contribute to explanations of particular moral facts but also to play certain epistemological roles: they should be something that moral agents can grasp, and which can guide those agents' moral thinking. But if moral principles have exceptions, then avoiding systematic moral errors requires a reliable ability to detect both the presence of moral reasons and the presence of permissible exceptions, including in some range of novel sets of circumstances. Otherwise our moral judgments will all too easily be mistaken. So how can hedged principles guide our judgments as to whether circumstances are permissibly exceptional?[40]

My account of what makes an exception permissible suggests an account of the content of the ability to judge cases as permissibly exceptional. The idea I'll develop is that one's judgment of (say) an instance of lying as a permissible exception to the status of something's being a lie as a reason

[40] I address further issues about how hedged principles can provide adequate moral guidance in my (2008).

not to do it can be guided by one's *conception of* the normative basis of this reason.

The moral principles we accept symbolize our commitment to the moral ideals we care about. Thinking that lying is wrong embodies some kind of ideal of not deceiving people. Some may interpret the principle as one expression of some more fundamental ideal, such as respecting people or promoting practices which benefit them. But more typically our conception of what moral concerns or ideals underlie the principles we accept is inchoate or incomplete. This isn't changed by the plausible idea that acquiring some initial understanding of such moral concerns and ideals is part of acquiring moral concepts. Even assuming that moral knowledge is possible, it is doubtful that most of us fully grasp all the principles we accept. Actual moral outlooks are works in progress.

We have seen that hedged principles generate substantive moral conclusions only in conjunction with further assumptions about what properties realize the normative bases they designate. So they do little to guide our judgments unless we have some grasp of those properties. But the extent to which most of us probably grasp them will often leave it indeterminate just what properties they are. Jonathan Dancy, who reports that he considers freedom of expression important but has no determinate sense why, thinks this would be a problem for principles which claim to incorporate an explanation of the moral reasons they identify (Dancy 2004: 153). I disagree.

Typically our acceptance of a principle like 'Curtailing freedom of expression is [*pro tanto*] wrong' isn't brute. I would be a defective moral agent if I thought, for instance, that it is wrong for the government to censor the press or ban protests at speeches by its officials, but didn't think that there was any basis for judging such government actions to be bad. So long as I think there is some such normative basis, a hedged form of the principle that curtailing freedom of expression is wrong is available to me. And I can perfectly well accept that principle even if I have no clear sense of just which kinds of expression are those to which freedom is important (academic freedom, pornography, assertion of the Armenian genocide, denial of the Holocaust …), or whether some restrictions on freedom to them are appropriate, or why it is important for them to be free.

The point I wish to make here is twofold. First, even if my grasp of the normative basis designated by a hedged principle to the effect that curtailing freedom of expression is wrong is incomplete, this isn't a kind of incompleteness which would leave the proposition expressed by the principle incomplete or indeterminate. The incompleteness lies mainly in one's grasp of what property realizes a normative role which is itself

reasonably determinate. Secondly, even if my grasp of what property realizes the relevant normative role is inchoate or incomplete, it may still have enough content reliably to guide my judgments, at least within a certain range of cases.

This second point has several strands. Even if I am unsure about why freedom of expression is important, I may know that the designated normative basis is such that it is wrong to curtail open discussion of public policy and academic freedom. Even a grasp this limited of the implications or other identifying characteristics of the designated normative basis may be enough reliably to guide my judgment, at least within a certain range. (This can be so even if I am unsure or altogether mistaken about its implications in some other cases.) Moreover, even if I suspect that the importance of freedom of expression is connected to some further factors, such as considerations of harm or preconditions of a healthy democracy, I might not know exactly what harm is or what such preconditions are. Still, if I think that they have got something to do with such properties as the flourishing or rational autonomy of persons, my judgments may still be guided by a pretty good proxy, at least within a certain range.[41] Finally, even if I have some particular property in mind as the normative basis—such as helping people flourish or protecting their autonomy—but don't think I fully grasp it, I may still grasp it and its implications and other identifying characteristics well enough for my judgments to be reliable, again at least within a certain range.

These are some of the ways in which our judgments may be guided, at least within a certain range, by acceptance of hedged principles and a grasp of the implications and other identifying characteristics of the normative bases they designate, even if we have no particular properties in mind as those which fill the relevant normative basis roles or have only a limited understanding of those properties.[42] What explains these possibilities, according to my

[41] Similarly, I might not fully understand what welfare is. Still, if I think that it has got something to do with happiness, my judgments concerning permissible exceptions to various welfarist, person-affecting, or retributivist moral principles would seem to be guided by a pretty good proxy. Thanks to Connie Rosati for this example.

[42] These possibilities are general to definite descriptions. I needn't have any particular number in mind when making a *de re* utterance of 'The number of Supreme Court justices is odd.' Similarly, I needn't have any particular property in mind when making a *de re* utterance to the effect that the basis for lying's status as a reason is such-and-such. I may be speaking truly in making an attributive utterance of 'The man in the corner drinking a martini is a spy' and having James Bond in mind, even if I am uncertain or mistaken about whether he is James Bond. Similarly, I may be speaking truly in saying that the normative basis of some reason is such-and-such, even if I am uncertain or mistaken about whether whatever property I have in mind in fact does fill the relevant normative basis role.

account, is that our reliability at judging whether certain facts are moral reasons or whether the circumstances are permissibly exceptional is a function of how well we are tracking the bases of those facts' status as reasons. Since this is something that comes in degrees, hedged principles can be used to describe both the degree and the scope of the reliability of our judgments in terms of how accurate and complete a conception we have of the properties which fill the relevant normative basis roles. Even a limited grasp of those properties can help us see what the permissible exceptions have in common and so increase the reliability of our judgments, perhaps also in some novel sets of circumstances. Similarly, the better we grasp these properties, the more robustly reliable our judgments are going to be.

We can similarly explain why such interfering factors as uncertainty, ignorance, and error concerning what properties fill the normative basis roles will tend to make our judgments less reliable, or reliable only in some limited range of cases. Since what properties fill these roles is a substantive moral issue, my account correctly classifies uncertainty, ignorance, and error about moral principles and their implications as concerning substantive moral issues, such as what explains our moral reasons. Parallel points apply to moral disagreement. Disagreements about whether killing, lying, or curtailing freedom of expression are things we have moral reason not to do, and whether killing in self-defense, telling white lies for a well-meaning end, or curtailing expressions of Holocaust denial also are things we have moral reason not to do, can be explained as disagreements about what factors ground and explain whatever reasons we have in these cases. So can disagreements about whether these reasons tolerate any exceptions. For example, suppose you think that freedom of expression matters because it is crucial for healthy government whereas I think it matters because it is crucial for exercising rational autonomy. This will lead to predictable sorts of disagreement about the implications of the principle that curtailing freedom of expression is wrong.

So how to resolve uncertainty, ignorance, error, and disagreement concerning hedged principles and their implications? How can we improve our grasp of the principles we accept and our ability to detect the presence of moral reasons and permissible exceptions? Such progress will require substantive moral inquiry. To improve our grasp of when killing is wrong, for example, we need to think hard about self-defense, abortion, euthanasia, capital punishment, war, and so on. There are various ways of doing this, at least so long as conditions are generally favorable for judgment. Just how we should proceed depends on what the proper method of moral inquiry is.

To illustrate the general idea, consider what we should do if the proper method is to seek a wide reflective equilibrium among our non-moral

background theories and the moral judgments and principles of various levels of generality which we accept provisionally. We should consult our moral experience and hypothetical cases to determine what our considered judgments are concerning the permissibility of killing in various contexts. We should figure out whether we accept other principles which permit killing in some circumstances. We should figure out what sorts of interests are at stake in these contexts and what biology, medicine, psychology, sociology, and other such sources would tell us about how killing someone would affect those interests. We should determine whether these sources suggest that things other than killing might have the same sort of significance for those interests. And we should organize and revise this information so that it all hangs together well.

Reasoning of this kind can no doubt improve our grasp of the principles we accept and our reliability in applying them. According to my account of hedged principles, this is because it can improve our grasp of at least the implications and other identifying characteristics of the normative bases of moral reasons. Even if such information doesn't directly identify the properties which fill these roles, it may help us determine what properties best satisfy their implications and identifying characteristics or at least rule out certain candidates. It can thereby support conclusions about what properties fill these roles via an inference to the best explanation of these facts. This kind of reasoning characterizes one sort of inquiry into what is wrong about killing, lying, and so on. The moral progress that we could make via such inquiry, even when slow and piecemeal, would still be progress. And insofar as such progress is possible, hedged moral principles are something we can grasp, and which can reliably guide our moral judgments.

One might object that hedged principles cannot adequately guide our judgments unless the restrictions on their scope can be captured in purely non-moral terms. Unless we can state their application conditions in purely non-moral terms, principles can be replaced by reasoning by cases and will fail to provide a rational basis for resolving moral disagreements (Goldman 2002: 13–16).

But I see no reason to think that hedged principles can guide our judgments only if their application conditions can be stated in purely non-moral terms. Our ordinary moral, prudential, and legal reasoning, where we commonly rely on principles employing normative terms, seem not to depend on the contingency of whether our language offers us purely non-moral vocabulary adequate for expressing moral properties. No such reduction is also required by my account of what makes an exception permissible or how we can resolve uncertainty and disagreements about principles. Whether principles provide moral insight or a rational basis for resolving

moral disagreements seems not to turn on whether they offer purely non-moral starting points for moral reasoning. Descriptions which appear non-normative can be controversial; consider the famous example of 'no vehicles in the park' (Hart 1958). And claims which employ normative terms needn't pose any serious issues of disagreement; consider injunctions against torturing the innocent for no gain. I see no good reason why hedged principles could guide our judgments only if they gave us an entry to moral facts and distinctions through purely non-moral descriptions, so long as our judgments can rely on an improvable grasp of the moral concerns and ideals which underlie our acceptance of moral principles.

7 HEDGED PRINCIPLES AND RIVAL ACCOUNTS

We can usefully round out the picture of what we gain by thinking about principles and exceptions in ethics along the lines of my account of hedged principles by considering some advantages my account enjoys over some of its rivals. I'll argue that these rivals encounter difficulties which my account avoids in explaining why circumstances are permissibly exceptional when they are (or why not when not) or how our judgments about permissible exceptions can be reliably guided.

One way to make principles tolerate exceptions would be to build into them some clause simply to the effect that there are no exceptional conditions present. Call this the "quantified account":

(QA) Something's being a lie is always a moral reason not to do it, unless something occurs to prevent its being a lie from being a reason not to do it.

(QA) requires the absence of features that make a situation permissibly exceptional. So when such features are present, (QA) correctly implies that something's being a lie isn't a reason not to do it.

(QA) has an obvious problem. We know in advance of (QA) that principles imply such qualified conditionals as 'If something is a lie and nothing in the circumstances prevents its being a lie from being a reason not to do it, then its being a lie is a reason not to do it.'[43] But the quantifier

[43] Here I draw on the criticism of Morreau's (1997: 192–200) "fainthearted conditional" account of disposition ascriptions in Fara (2005: 56–9). The same problem plagues Hausman's (1992: 136–7) proposal that '*Ceteris paribus*, all *F*s are *G*s' is true in context *X* iff *X* picks out a property *C* such that 'Everything that is both *F* and *C* is a *G*' is true. It also plagues Braun's (2000: 215) truth conditions for "*ceteris paribus* conditionals," which can be stated in a simplified form as follows: 'If *A* then *ceteris paribus B*' is true at world *w*, with respect to context *c*, iff '$(A \,\&\, N_c) \,\Box\!\!\rightarrow\, B$' is true at *w*—where N_c are the conditions that would be determined by *c* to be non-exceptional with respect to the connection between *A* and *B*, and '$\Box\!\!\rightarrow$' is the standard subjunctive conditional.

clause says nothing more about what can prevent something from providing a reason for (or against) something. Thus no progress is made with respect to specifying which circumstances are permissibly exceptional, or under what condition they are so, by saying that the fact that an action is a lie is a reason not to do it unless something prevents it from being one. Although (QA) ensures *that* the circumstances are permissibly exceptional when they in fact are, it cannot be used to explain *why* they are. Nor can (QA) guide our judgments about permissible exceptions. So (QA) cannot suit principles to the explanatory and epistemological roles required by moral theory.

One alternative, the "list account," eliminates the quantifier clause in favor of its satisfiers:

(LA) Something's being a lie is always a reason not to do it, unless ____.

The blank in (LA) is meant to stand for a complete list of permissible exceptions. What might be used to motivate (LA) is the assumption, often found in the literature on *ceteris paribus* generalizations, that hedge clauses avoid becoming catch-alls that render a generalization vacuously true only if they are shorthand for an explicit list of background provisos.[44] There are two different ways of trying to fill out such a list, corresponding to two different versions of the list account.[45] On the "merely extensional" version, the relevant list is simply a list of the conditions under which something's being a lie isn't a reason not to do it. On the "constitutive" version, the conditions on the relevant list must be ones which make it the case that something's being a lie isn't a reason not to do it. Either version of (LA) can play the explanatory and epistemological roles required by that moral theory only if the list of permissible exceptions is finite and, indeed, manageably short. We can grasp only finite principles, and they can guide our judgments only if they are cognitively manageable. The two versions of (LA) fare differently with respect to this constraint.

It seems highly unlikely that we can enumerate all the permissible exceptions to generate the kind of list which the merely extensional version of (LA) describes. Such efforts have proved unpromising outside ethics. The literature on *ceteris paribus* generalizations displays some consensus that exceptions to such generalizations rarely are finitely specifiable; it may

[44] See e.g. Hempel (1988), Schiffer (1991), and Earman and Roberts (1999).

[45] I am grateful to Sean McKeever for pressing this distinction.

even be that hedge clauses of one or another sort are needed precisely when no explicit list of background provisos is available.[46] In the absence of any a priori guarantee that morality is special, it is especially compelling that no complete list of permissible exceptions would be manageably short. But in that case any manageably short list which we might be able to generate would merely exemplify, not exhaust, the class of permissible exceptions.[47] Which conditions our incomplete list would include would be highly contingent. Since such a list could easily be unrepresentative or highly heterogeneous, it could easily fail to project appropriately to the other cases. So it seems unlikely that the merely extensional version of (LA) would reliably guide our judgments.

Even if the merely extensional version of (LA) could supply a complete but manageably short list of the permissible exceptions, it would still be both explanatorily deficient and unnecessary. It would be explanatorily deficient because to give such a list isn't yet to explain why the conditions it mentions are the ones to make the circumstances permissibly exceptional, let alone to explain why certain facts would be reasons when those conditions don't obtain.[48] A satisfactory account should do this. If the status of something as a reason permits exceptions, surely it is no accident which of its instances are permissibly exceptional and which aren't. But if we can explain these facts, the merely extensional version is unnecessary. For then we can bypass the list in favor of a condition which states when an exception is permissible and is explanatorily prior to any particular list we might happen to grasp. If our judgments concerning permissible exceptions can be guided by such a condition, then grasping any particular list of exceptions will likewise be unnecessary for guiding our judgments. So the merely extensional version of (LA) neither can nor is needed to suit

[46] See e.g. Fodor (1991) and Pietroski and Rey (1995) for the first point, and Smith (2007) for the second.

[47] Cf. Pettit (1999: 24) and Lange (2000: 170–4). The argument in Donagan (1977: 92–3) that the principle 'It is impermissible for anybody to break a freely made promise to do something in itself morally permissible' tolerates exceptions presupposes an extensional version of (LA). See also Shafer-Landau (1997: 593–4).

[48] A related point is that saying that something would be a reason if it occurred in some counterfactual situation isn't a satisfying account of what it is for something to be a reason in the circumstances at hand. This tells us at most how what is supposed to give the reason here would operate in a very different kind of situation (Dancy 1993: 97–8; 1999: 27; 2004: 19). Such views include: Ross's official analysis of prima facie duty (Ross 1930: 19–20); Montague's proposal that moral principles are guaranteed to hold only in "morally simple" situations where only one reason is present, whereas in morally complex situations their force is merely epistemic (Montague 1986: 646–7); and those deontic logics which assign truth conditions to principles in terms of "good-and-simple" possible worlds, in which things go as they morally ought to in the morally simple way (e.g. Asher and Bonevac 1997: 165).

moral principles to the explanatory and epistemological roles required by moral theory.

The constitutive version of (LA) is meant to avoid these problems. The list of conditions which make it the case that something isn't a reason for (or against) something is also more likely to be manageably short than a merely extensional list. Whether any particular constitutive version of (LA) in fact achieves these things depends on just what conditions it throws on the list. But my account has advantages over certain constitutive versions of (LA) independently of this issue.

In arguing this point, we should note that my account of hedged principles could be construed as one constitutive version of (LA). It specifies a condition on circumstances—namely, the presence of the relevant normative basis—whose failure to hold makes circumstances permissibly exceptional. What makes something not instantiate the relevant normative basis is some situational feature. So my account implies a condition in virtue of satisfying which a situational feature makes it the case that some feature of an action permissibly fails to provide a reason for (or against) doing it.

But now it follows that my account can subsume any version of (LA) whose list contains just the situational features which satisfy my condition. For whether something fails to instantiate the normative basis designated by a hedged principle is explanatorily prior to any such list. My account has the advantage of requiring no particular list of such conditions or that it be manageably small in number. (I am not, of course, recommending that we ignore examples of features which generate permissible exceptions. Consideration of examples is usually helpful.) The account can also unify any list given by these versions of (LA). It articulates a deeper factor which the situational features on the list have in common and in virtue of which each makes certain circumstances permissibly exceptional. The proviso requiring the instantiation of the designated normative basis presupposes no particular list of the possible ways of failing to instantiate it. So my account provides greater explanatory depth, unity, and economy than these versions of (LA).

This result has broader significance. Sean McKeever and Michael Ridge, who defend the kind of constitutive version of (LA) to which I just compared my account, claim that "the best explanation of the possibility of practical wisdom . . . entails that practical wisdom involves the internalization of a finite and manageable set of non-hedged principles" (2006: 139). (Practical wisdom is, roughly, a reliable ability to make correct moral judgments, also at least in some novel sets of circumstances.) I cannot address their extended argument for this claim. But my account suggests an explanation of the possibility of practical wisdom which is at least no worse than theirs. So, even if we can construct a finite and manageable set of true non-hedged

principles to be internalized, the best explanation of the possibility of practical wisdom doesn't require this.

Both accounts explain how acceptance of moral principles can guide one's judgments concerning moral reasons and permissible exceptions and how the reliability of these judgments can be improved. But my explanation requires fewer and weaker assumptions concerning the list of permissible exceptions while providing greater explanatory depth and unity. It turns on whether, and how well, we track a condition (the presence of the relevant normative basis) whose satisfaction explains why a feature of an action is a reason for (or against) doing it and whose failure explains why certain situational features make the circumstances permissibly exceptional. McKeever and Ridge's explanation equally requires us to track these facts by grasping certain lists of features of actions and situations. But the crucial condition in my explanation gives a deeper unifying specification of what the situational features whose presence would generate permissible exceptions have in common, whereas McKeever and Ridge's doesn't. Insofar as we can track this condition, we can detect the presence of these situational features, at least within some range, without needing to grasp any particular list of features *ex ante*. On my explanation, we can also track this condition even if the situational features which generate permissible exceptions don't come in a manageably short list. Hence my explanation avoids making the possibility of practical wisdom and the reliability of our judgments hostage to what the number of these situational features happens to be.

Another rival to my account comes from Mark Lance and Margaret Little's account of defeasible moral generalizations (2007: 165). They state this "privileging account" (as I'll call it) as follows:

(PA) $P(\forall x)(Gx \mathrel{\Box\!\!\rightarrow} Mx)$, where 'P' is a modal operator 'in privileged conditions' and '$\Box\!\!\rightarrow$' is the standard subjunctive conditional.

Principles of the form (PA) and the form (HP) share the implication that the connections they describe between factors such as lying and wrongness are stable or invariant under a certain range of hypothetical changes. In privileged conditions, that something is a lie will always count towards its wrongness. What explains why this contribution fails to hold, when it does, is the way the context deviates from these privileged conditions. For example, lying while playing Diplomacy isn't wrong at all because in such a context lying takes place against consensual agreements made in circumstances in which lying would have its "default" status of counting towards wrongness (Lance and Little 2006: 313–14). We can understand why circumstances are permissibly exceptional, when they are, by grasping what conditions are relevantly privileged and what implications the different kinds of deviations from those conditions would have for

the connection between lying and wrongness.[49] Such an understanding should also be able to guide our judgments as to whether circumstances are permissibly exceptional. So this account agrees with mine that being reliable in these judgments requires no *ex ante* grasp of any particular list of exceptions.

The privileging account can use deviation from privileged conditions to explain why certain cases but not others are permissibly exceptional. But why are certain conditions privileged in the first place? According to Lance and Little, we are supposed to see which conditions are privileged by understanding the "defeasibly wrong-making nature" of lying, causing pain, and so on—and there the explanation stops. It is supposed to be a basic fact about what pain is, for instance, that in privileged conditions causing pain is defeasibly bad making.[50] This is a significant feature. If the nature of pain doesn't involve such explanatorily basic facts about the relevant privileged conditions, or if we don't grasp these facts, then the privileging account cannot explain how we can be reliable at judging whether circumstances are permissibly exceptional or not.

But the privileging account is unnecessary anyway, because the profile of the privileged conditions associated with a given feature needn't be treated as an explanatorily basic fact about it. The privileged status of those conditions can be accounted for by reference to the normative basis of the feature's status as a reason for (or against) doing what has it. For instance, the privileged conditions in which something's being a lie is a moral reason not to do it will be those in which it instantiates the normative basis of such a fact's status as a reason not to lie. The status of certain other conditions as non-privileged can similarly be accounted for by reference to the failure of a lie to instantiate this normative basis. Together all this makes the privileging account unnecessary at least in every case in which the normative basis of something's status as a moral reason is distinct from it (see the end of Section 3). But I suspect that in many cases what explains why something provides moral reasons isn't a part of that something or otherwise intrinsic to it. If that is right, then my account of

[49] Lance and Little's discussion of the different ways in which the contributions of features in non-privileged conditions may depend on their contributions in privileged conditions is subtle and insightful. But what I say here doesn't turn on the details. I detect similar ideas, but developed in a more specific Kantian context, in Schapiro (2006).

[50] Lance and Little think that true instances of (PA) involve a strong enough necessity to entitle us to say that the relevant features are "constitutively" such that in privileged conditions their instances have the specified moral character; they are "moral kinds" whose "essence" this connection characterizes. See Lance and Little (2006: 316–17) and (2007: 165–6). This seems too strong. I suspect I can fully well understand the nature of pain and many other bad-making features without understanding what the privileged conditions are in which they are bad making.

hedged principles seems to generalize better. Moreover, since the normative bases can be used to explain both why certain conditions are privileged and others are non-privileged, the explanations of moral reasons and permissible exceptions which my account provides have a satisfying unity which the privileging account seems to lack.

I conclude that my account of hedged principles compares favorably with the rivals I have considered.[51] The account shows how moral principles are something that we can grasp, and which can reliably guide our moral judgments, even if they can permit exceptions. This includes a unified account of the abilities to judge whether certain features provide moral reasons and whether the circumstances are permissibly exceptional. What supplies this account is an explanation of moral reasons and permissible exceptions which derives explanatory depth and unity from its appeal in both cases to whether or not something instantiates that factor—the designated normative basis of a moral reason—in virtue of whose presence or absence, respectively, a feature provides moral reasons or the circumstances are permissibly exceptional. While the account allows for variation in the degree and scope of the reliability of our moral judgments, it also shows how their reliability can be improved by improving our grasp of the moral ideals which explain moral reasons and permissible exceptions. The account can also explain how our judgments can be reliable irrespective of such contingencies as how frequent the permissible exceptions are and whatever particular lists of exceptions we happen to grasp *ex ante*. I know of no other account of moral principles which gives us as much if we think about exceptions in ethics along the lines that it recommends.

8 CONCLUSION

In other work, I have argued that my account of hedged principles captures many insights which are supposed to motivate moral particularism but nonetheless supports moral generalism. In this paper, I have developed this account in more detail and highlighted some of its explanatory and epistemological advantages. Although some parts of the theory remain no more than a sketch here, I think I may conclude that hedging moral principles in the way I propose shows how principles can permit exceptions while still playing the explanatory and epistemological roles required by moral theory. We should therefore find nothing peculiarly odd or problematic about the possibility of exception-tolerating and yet genuinely explanatory generalizations in morality.

My parting observation is that thinking about moral principles along the lines I have recommended directs us to such questions as: What is it about censoring the press that makes it wrong? Why is it that well-being matters to what our moral obligations are? What is so bad about killing a person as to make it wrong? Why might it be all right to tell someone a white lie to bolster their confidence? Since these are just the sorts of questions which exercise moral theorists, perhaps something like the hedged principles I have articulated already are an implicit part of their kit.

[51] Unfortunately, space prevents me from comparing my account with the sophisticated but somewhat complicated "dispositionalist" account of exception-tolerating principles developed in Robinson (2006). But see Robinson (2008).

REFERENCES

Asher, Nicholas, and Bonevac, Daniel (1997) 'Common Sense Obligation,' in D. Nute (ed.), *Defeasible Deontic Logic* (Dordrecht: Kluwer).

Baldwin, Thomas (2002) 'The Three Phases of Intuitionism,' in P. Stratton-Lake (ed.), *Ethical Intuitionism: Re-evaluations* (Oxford: Clarendon Press).

Braun, David (2000) 'Russellianism and Psychological Generalizations' *Nous* 34: 203–36.

Carlson, Gregory N., and Pelletier, Francis Jeffry (eds.) (1995) *The Generic Book* (Chicago: University of Chicago Press).

Cartwright, Nancy (1983) *How the Laws of Physics Lie* (Oxford: Clarendon Press).

Dancy, Jonathan (1993) *Moral Reasons* (Oxford: Basil Blackwell).

—— (1999) 'Defending Particularism' *Metaphilosophy* 30: 25–32.

—— (2000) 'The Particularist's Progress,' in B. Hooker and M. O. Little (eds.), *Moral Particularism* (Oxford: Clarendon Press).

—— (2004) *Ethics without Principles* (Oxford: Clarendon Press).

Donagan, Alan (1977) *The Theory of Morality* (Chicago: University of Chicago Press).

Earman, John, and Roberts, John (1999) '*Ceteris Paribus*, There Is No Problem of Provisos' *Synthese* 118: 439–78.

Fara, Michael (2005) 'Dispositions and Habituals' *Nous* 39: 43–82.

Feldman, Fred (1986) *Doing the Best We Can* (Dordrecht: D. Reidel).

Fodor, Jerry A. (1991) 'You Can Fool Some of The People All of The Time, Everything Else Being Equal; Hedged Laws and Psychological Explanations' *Mind* 100: 19–34.

Giere, Ronald (1999) *Science without Laws* (Chicago: University of Chicago Press).

Goldman, Alan H. (2002) *Practical Rules: When We Need Them and When We Don't* (Cambridge: Cambridge University Press).

Hart, H. L. A. (1958) 'Positivism and the Separation of Law and Morals' *Harvard Law Review* 71: 593–629.

Hausman, Daniel M. (1992) *The Inexact and Separate Science of Economics* (Cambridge: Cambridge University Press).

Hempel, Carl G. (1988) 'Provisos: A Problem Concerning the Inferential Function of Scientific Theories' *Erkenntnis* 28: 147–64.

Hooker, Brad (2008) 'Particularism and the Real World,' in M. Potrc, V. Strahovnik, and M. Lance (eds.), *Challenging Moral Particularism* (London: Routledge).
Horty, John F. (2003) 'Reasoning with Moral Conflicts,' *Nous* 37: 557–605.
——(2007) 'Reasons as Defaults' *Philosophers' Imprint* 7(3), http://www.philosophersimprint.org/007003/.
Irwin, T. H. (2000) 'Ethics as an Inexact Science: Aristotle's Ambitions for Moral Theory,' in B. Hooker and M. O. Little (eds.), *Moral Particularism* (Oxford: Clarendon Press).
Kim, Jaegwon (1994) 'Explanatory Knowledge and Metaphysical Dependence' *Philosophical Issues* 5: 51–69.
Koslicki, Kathrin (1999) 'Genericity and Logical Form' *Mind and Language* 14: 441–67.
Lance Mark, and Little, Margaret (2006) 'Defending Moral Particularism,' in J. Dreier (ed.), *Contemporary Debates in Moral Theory* (Oxford: Blackwell Publishers).
——(2007) 'Where the Laws Are,' in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, ii (Oxford: Oxford University Press).
Lange, Marc (2000) *Natural Laws in Scientific Practice* (Oxford: Oxford University Press).
Liebesman, David (Ms) 'Simple Generics.' Unpublished.
Little, Margaret O. (2000) 'Moral Generalities Revisited,' in B. Hooker and M. O. Little (eds.), *Moral Particularism* (Oxford: Clarendon Press).
McKeever, Sean, and Ridge, Michael (2006) *Principled Ethics: Generalism as a Regulative Ideal* (Oxford: Clarendon Press).
McNaughton, David, and Rawling, Piers (2000) 'Unprincipled Ethics,' in B. Hooker and M. O. Little (eds.), *Moral Particularism* (Oxford: Clarendon Press).
Montague, Phillip (1986) 'In Defense of Moral Principles' *Philosophy and Phenomenological Research* 46: 643–54.
Morreau, Michael (1997) 'Fainthearted Conditionals' *Journal of Philosophy* 94: 187–211.
Nickel, Bernhard (Ms) 'Processes in the Interpretation of Generics and CP-Laws.' Unpublished.
Pettit, Philip (1999) 'A Theory of Normal and Ideal Conditions' *Philosophical Studies* 96: 21–44.
Pietroski, Paul, and Rey, Georges (1995) 'When Other Things Aren't Equal: Saving Ceteris Paribus Laws from Vacuity' *British Journal for the Philosophy of Science* 46: 81–110.
Pollock, John L., and Cruz, Joseph (1999) *Contemporary Theories of Knowledge*, 2nd edn (Lanham, MD: Rowman & Littlefield).
Robinson, Luke (2006) 'Moral Holism, Moral Generalism, and Moral Dispositionalism' *Mind* 115: 331–60.
——(2008) 'Moral Principles Are Not Moral Laws' *Journal of Ethics and Social Philosophy* 2(3), http://www.jesp.org.
Ross, W. D. (1930) *The Right and the Good* (Oxford: Clarendon Press).
——(1939) *The Foundations of Ethics* (Oxford: Clarendon Press).
Ruben, David-Hillel (1990) *Explaining Explanation* (London: Routledge).

Scanlon, T. M. (1998) *What We Owe to Each Other* (Cambridge, MA.: Harvard University Press).
Schapiro, Tamar (2006) 'Kantian Rigorism and Mitigating Circumstances' *Ethics* 117: 32–57.
Schiffer, Stephen (1991) 'Ceteris Paribus Laws' *Mind* 100: 1–17.
Shafer-Landau, Russ (1997) 'Moral Rules' *Ethics* 107: 584–611.
Silverberg, Arnold (1996) 'Psychological Laws and Non-Monotonic Logic' *Erkenntnis* 44: 199–224.
Smith, Martin (2007) 'Ceteris Paribus Conditionals and Comparative Normalcy' *Journal of Philosophical Logic* 36: 97–121.
Stratton-Lake, Philip (2002) 'Pleasure and Reflection in Ross's Intuitionism,' in P. Stratton-Lake (ed.), *Ethical Intuitionism: Re-evaluations* (Oxford: Clarendon Press).
——(2003) 'Scanlon's Contractualism and the Redundancy Objection' *Analysis* 63: 70–6.
Väyrynen, Pekka (2006) 'Moral Generalism: Enjoy in Moderation' *Ethics* 116: 707–41.
——(2008) 'Usable Moral Principles,' in M. Potrc, V. Strahovnik, and M. Lance (eds.), *Challenging Moral Particularism* (London: Routledge).
White, Roger (2005) 'Explanation as a Guide to Induction' *Philosophers' Imprint* 5 (2), www.philosophersimprint.org/005002/.
Woodward, James (2003) *Making Things Happen* (Oxford: Oxford University Press).