

From Thick to Thin: Two Moral Reduction Plans¹

DANIEL Y. ELSTEIN
University of Leeds
Leeds, LS2 9JT
UK

and

THOMAS HURKA
University of Toronto
Toronto, ON M5R 2M8
CANADA

Many philosophers of the last century thought all moral judgments can be expressed using a few basic concepts — what are today called ‘thin’ moral concepts such as ‘good,’ ‘bad,’ ‘right,’ and ‘wrong.’ This was the view, first, of the non-naturalists whose work dominated the early part of the century, including Henry Sidgwick, G.E. Moore, W.D. Ross, and C.D. Broad. Some of them recognized only one basic concept, usually either ‘ought’ or ‘good’; others thought there were two. But they all assumed that other moral concepts, including such ‘thick’ ones as the virtue-concepts ‘courageous’ and ‘kindly,’ can be reductively analyzed using one or more thin concepts and some more or less determinate

1 We thank Simon Blackburn, Hallvard Lillehammer, Gopal Sreenivasan, Sergio Tenenbaum, and Ralph Wedgwood for helpful suggestions and are especially grateful to Simon Kirchin for first stimulating this paper and then improving it through a series of immensely helpful oral and written comments.

descriptive content. This was also the view of many non-cognitivists who wrote later in the century, including C.L. Stevenson and R.M. Hare. They thought judgments using thin terms express one or two basic moral attitudes, either pro or con and with distinctive formal features such as categoricity and universality, and that any thick terms can be reduced to thin ones plus some description.

In recent decades a contrary view has emerged that claims that thick concepts are irreducible. According to its proponents, terms like 'courageous' and 'kindly' have both morally evaluative and descriptive meaning, but the two interpenetrate each other in a way that makes the separation a reductive analysis requires impossible. Thick concepts are therefore not derivative from thin ones, which do not have the primacy the above-mentioned philosophers assumed. On the contrary, on some versions of this anti-reductive view it is the thick concepts that are primary, with the thin ones mere abstractions from them.²

The mark of a thin concept like 'right' is that it says nothing about what other properties an item falling under it has. If moral properties supervene on non-moral ones, as most philosophers accept, then any act that is right will have other, non-moral properties that make it right. In addition, if moral judgments are universalizable, as the non-naturalists and non-cognitivists believed, then any other act with the same non-moral properties will likewise be right. But while the claim 'x is right' says or implies that x has *some* right-making properties, it says nothing about what in particular they are. If we know the other evaluations that someone who asserts this claim has made, we may be able to guess what right-making properties he has in mind now; if we know the general evaluative practices of his culture, that knowledge may also help. But these speculations go beyond the semantic content of 'x is right' itself, which says only that some properties of x make it right without specifying what they are. The reductive view therefore analyzes 'x is courageous' into an evaluative component that does not say anything about x's non-moral properties and a descriptive component that does. The anti-reductive view says this separation is impossible.

In this paper we will defend the reductive view of thick concepts by answering the most common argument against it, and in so doing will defend a position held by more philosophers than the recent literature suggests. This literature has tended to associate the issue about thick concepts with that between cognitivist and non-cognitivist accounts of moral judgment, as if the reductive view were essentially non-cogni-

2 See, e.g., Christine Tappolet, 'Through Thick and Thin: *Good* and its Determinates,' *Dialectica* 58 (2004) 207-20.

tivist and therefore one all cognitivists must deny.³ Now, it is indeed essential to any non-cognitivist view to separate evaluation sharply from description and so to require some reduction of thick concepts; an attack on reductivism is therefore also an attack on non-cognitivism. But the historical record shows cognitivists like Sidgwick and Moore just as much as non-cognitivists like Stevenson taking the reductive line. Nor do we see the slightest inconsistency in combining the cognitivist view that moral judgments express beliefs with the reductive view that they all use a few thin concepts. So while non-cognitivism is committed to some version of the reductive view, cognitivism can either accept that view or reject it. Our defense of the view will therefore discuss the relation between thick and thin concepts in a way that is neutral between cognitivist and non-cognitivist accounts of either, and will propose analyses that can be used by non-cognitivists and reductive cognitivists alike.

The case against the reductive view is commonly credited to John McDowell, but there are two separate arguments detectable in his writing. A first, or 'uncodifiability,' argument claims, on the basis of Wittgensteinian ideas about rule-following, that moral judgments cannot be codified in general principles; part of this argument grants that the moral supervenes on the non-moral but denies that moral judgments are universalizable.⁴ Though we will return to this argument later, it will not be our prime focus, because we do not see how on its own it bears on the thick/thin issue. McDowell's uncodifiability claim presumably applies to all moral judgments, those using thin concepts as much as those using thick ones. If judgments about courage and kindness cannot be formulated in general principles, surely neither can ones about what is good, right, or 'the thing to do.'⁵ But if uncodifiability applies equally to thick and thin concepts, how can it support any conclusion about the relation between them? If judgments using both kinds of concept resist principles, it may be that there are no reductive relations between them, but it also may be that there are. Assume that

3 See, e.g., Jonathan Dancy, 'In Defense of Thick Concepts,' in Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein, eds., *Midwest Studies in Philosophy XX: Moral Concepts* (Notre Dame: University of Notre Dame Press 1996), 263-79.

4 John McDowell, 'Are Moral Requirements Hypothetical Imperatives?' *Proceedings of the Aristotelian Society, Supplementary Volume* 52 (1978) 13-29; 'Virtue and Reason,' *The Monist* 62 (1979) 331-50; and 'Non-Cognitivism and Rule-Following,' in S.H. Holtzman and C.M. Leich, eds., *Wittgenstein: To Follow a Rule* (London: Routledge & Kegan Paul 1981) 141-62. The acceptance of supervenience and denial of universalizability come in 'Non-Cognitivism and Rule-Following,' 144-5.

5 McDowell, 'Virtue and Reason,' 332

thin judgments are uncodifiable and that thick ones are reducible to thin ones plus some description. Then the uncodifiability of the thick judgments simply follows: the uncodifiability of the thin concepts transfers to the thick concepts in whose analysis they figure, so the latter are uncodifiable just because and in so far as the former are. Assuming the uncodifiability of the thin, then, the uncodifiability of the thick is precisely what the reductive view predicts; uncodifiability alone does not tell against that view but is perfectly consistent with it.⁶

McDowell's other argument does bear on the thick/thin issue and is repeated without reference to uncodifiability by many later writers. This 'disentanglement' argument starts by assuming that any reductive analysis of a thick concept will have a particular form, which we will call a 'descriptively determinate two-part form.' This analysis contains, first, a descriptive component that fully determines the concept's extension, identifying descriptive properties that decide, as unambiguously as such properties ever do, what does and does not fall under it. It then adds an evaluative component that commends or condemns items for having those properties, so the general form of the analysis is 'x has descriptive properties A, B, and C (for specific A, B, and C), and is good/bad/right/wrong for doing so.' On this view a thick concept is an ordinary descriptive concept with an evaluation added on. Now, this pattern of analysis may be appropriate for a derogatory term like 'Kraut,' which has the fully determinate descriptive meaning 'is a German' and adds condemnation of any German for being one. But McDowell argues that it is inadequate to capture the thick concepts of moral interest, including virtue-concepts such as 'courageous' and 'kindly.' It implies that one can always identify the extension of a thick term using only its descriptive part and without any knowledge of the evaluations a speaker uses it to make. But this disentanglement, McDowell claims, is impossible for virtue-terms such as 'courageous,' whose extension depends crucially on evaluations. He writes:

Consider, for instance, a specific conception of some moral virtue: the conception current in a reasonably cohesive moral community. If the disentangling manoeuvre is always possible, that implies that the extension of the associated term, as it would be used by someone who belonged to the community, could be mastered independently of the special concerns which, in the community, would show themselves in admiration or emulation of actions seen as falling under the concept. That is: one could know which actions the term would be applied to, so that one would be able to predict applications and withholdings of it in new cases

6 We do not of course claim that adherents of the reductive view typically accept uncodifiability; most reject it. Our claim is just that if uncodifiability applies to thin as well as thick judgments, it cuts no ice against the reductive view.

— not merely without oneself sharing the community's admiration (there need be no difficulty about that), but without even embarking on an attempt to make sense of their admiration.⁷

Since that is impossible for terms such as 'courageous,' McDowell concludes that the reductive view is false.

This disentanglement argument is repeated by Bernard Williams in the discussion that introduced the terminology of 'thick' and 'thin.' He too assumes that any reductive analysis will have a descriptively determinate two-part form, so it treats a thick concept as 'a conjunction of a factual and an evaluative element' and analyzes any statement using it as saying 'something like "this act has such-and-such a character, and acts of that character one ought not to do."' He then says it follows that, for any concept so analyzed, 'you could produce another that picked out just the same features of the world but worked simply as a descriptive concept, lacking any prescriptive or evaluative force.' Against this he argues, citing McDowell, that 'critics have made the effective point that there is no reason to believe that a descriptive equivalent will necessarily be available,' so to know the extension of a thick concept one must also grasp 'its evaluative point.'⁸ And the same appeal to disentanglement without reference to uncodifiability is made by later critics of the reductive view such as Jonathan Dancy, Hilary Putnam, Charles Taylor, and Christine Tappolet.⁹

Whatever McDowell's initial intentions, then, the disentanglement argument apart from claims about uncodifiability has become the most common anti-reductive argument in the literature. We think this is no accident. However well it reflects aspects of McDowell's overall meta-ethical view, the uncodifiability argument has premises about rule-following and against universalizability that are controversial and would be questioned not only by reductivists but also by many with no settled opinion on the thick/thin issue. The argument is therefore unlikely to persuade many not already on the anti-reductive side. But the disentanglement argument does not have this flaw. Its key premise, that we cannot determine the extension of a thick concept without making some

7 McDowell, 'Non-Cognitivism and Rule-Following,' 144

8 Bernard Williams, *Ethics and the Limits of Philosophy* (Cambridge, MA: Harvard University Press 1985), 130, 141-2.

9 See, e.g., Dancy, 'In Defense of Thick Concepts,' 263-4; Hilary Putnam, 'The Entanglement of Fact and Value,' in *The Collapse of the Fact/Value Dichotomy and Other Essays* (Cambridge, MA: Harvard University Press 2002) 28-45, at 34-40; Charles Taylor, 'Ethics and Ontology,' *Journal of Philosophy* **100** (2003) 305-20, at 305-6; and Tappolet, 'Through Thick and Thin,' 213-17.

evaluations, does seem neutrally acceptable and even compelling, and in this paper we will grant it for the important thick concepts. So if it were true that the reductive view cannot accommodate this premise, that inability would be strong evidence against it. This has surely been a key part of the disentanglement argument's appeal: that it derives an anti-reductive conclusion from a premise that even reductivists will find it hard to deny.¹⁰

But it is not true that the reductive view cannot accommodate this premise. Like too many arguments in philosophy, the disentanglement argument considers only the most simple-minded version of an opposing view and takes a refutation of that to refute the view as a whole. McDowell, Williams, and those who follow them all explicitly assume that any reductive analysis of a thick concept will have what we call a descriptively determinate two-part form. We concede that the argument is persuasive against analyses with this form, and also concede that such analyses have been given by several reductivists, in particular the non-cognitivists Stevenson, Hare, and Simon Blackburn.¹¹ But the obvious question is whether the reductive view is restricted to this type of analysis, and we think the obvious answer is that it is not. There are at least two patterns of analysis the view can use, either separately or together, to capture McDowell's premise while still reducing thick concepts to thin ones plus description.

Before we describe these patterns, we should emphasize the limited aim of our project. We are not giving a complete defense of the reductive view, nor a complete refutation of everything anti-reductivists have said. We are just answering one anti-reductive argument, though it has been a highly influential one. And since this argument denies that the reductive view can, even using its own resources, accommodate a compelling claim about thick concepts, we will feel free, in answering it, to help ourselves to those resources. Thus, if reductivists typically

10 Simon Blackburn has implicitly denied it, when he argues that most proposed thick concepts are purely descriptive, with no specific evaluations built into their content. 'Frugal,' for example, while often a term of praise, is the opposite when used of a host dispensing hospitality ('Through Thick and Thin,' *Proceedings of the Aristotelian Society, Supplementary Volume 66* (1992) 285-99, at 285-7). While this is a possible response to the disentanglement argument, we will follow a different tack and grant both that thick concepts have specific evaluations built into their contents and that their extensions depend on evaluations.

11 C.L. Stevenson, *Ethics and Language* (New Haven: Yale University Press 1944), 206-7; R.M. Hare, *The Language of Morals* (Oxford: Oxford University Press 1952), 121; Hare, *Freedom and Reason* (Oxford: Oxford University Press 1963), 121-9; and Simon Blackburn, *Spreading the Word* (Oxford: Clarendon Press 1984), 148-9.

assume universalizability, we will feel free to use universalizability in our analyses even though some on the other side reject it.

The first pattern of analysis is very straightforward. We have discussed two types of concept: at one extreme is a thin concept like 'good,' which says nothing about the good-making properties of items falling under it, at the other extreme is a descriptively determinate concept like 'Kraut,' which specifies those properties completely and therefore fully determines the concept's extension. Surely there is room between these extremes for a category of thick (or 'thick-ish') concepts whose descriptive component specifies good- or right-making properties to some degree but not completely, saying only that they must be of some specified general type but not selecting specific properties within that type — that is left to evaluation. Or, to put the point slightly differently, there can be concepts whose descriptive component defines an area in conceptual space within which admissible good- or right-making properties must be found, so any use of the concept associating it with properties outside that area is a misuse, but does not identify any specific point within the area as uniquely correct, as a concept like 'Kraut' does. The concept therefore has descriptive content, but this content is not completely determinate. The pattern of this analysis is something like 'x is good, and there are properties X, Y, and Z (not specified) of general type A (specified), such that x has X, Y, and Z, and X, Y, and Z make anything that has them good.' This pattern is reductive, because it uses only the thin concept 'good' and the descriptive concept 'A.' But it accommodates the key premise of the disentanglement argument because determining *which* properties of type A are the good-making ones, which we must do to determine the concept's extension, requires evaluative judgment.

To illustrate this pattern, consider the term 'just' in the specific sense used in discussions of distributive justice. And imagine two people who disagree about what makes for distributive justice, one saying that just distributions of, say, happiness are equal distributions, while another says they are distributions proportioned to people's merits, which can be unequal. If the reductive view had to use a fully determinate two-part analysis of 'distributively just,' it would have to say that these two people use different concepts of justice. For the egalitarian, 'x is just' means 'x is an equal distribution, and good for being so,' while for the desert-theorist it means 'x is a distribution proportioned to merit, and good for being so.' It would not follow that the two cannot disagree about distributive justice, since the claims their respective concepts imply about the basis of goodness in distribution contradict each other. But in debating the justice of a particular distribution they could not be disagreeing about whether a single concept applies to it, since they use different concepts. But does it not look as if they are disagreeing about

the application of a single concept? And is an analysis not preferable that allows this?

We can construct such an analysis if we use the first pattern of analysis and give 'distributively just' an only partly determinate descriptive component. Then 'x is distributively just' will mean something like 'x is good, and there are properties X, Y, and Z (not specified) that distributions have as distributions, or in virtue of their distributive shape, such that x has X, Y, and Z, and X, Y, and Z make any distribution that has them good.' This analysis places significant restrictions on the extension of 'distributively just.' If someone says an act of generously helping a stranger is distributively just, she is misusing the concept, because only distributions can be just in this sense. And if she calls a distribution just because it was brought about by generous actions, she is likewise misusing the concept, because only a distribution's properties as a distribution can bear on its justice. But these restrictions do not completely determine the concept's extension, and in particular do not determine whether it contains equal distributions or ones based on desert. That depends on which properties of distributions are in fact good-making, so the analysis captures McDowell's premise. To know the intended extension of 'distributively just' as used by some person or community, it is not enough to know the descriptive part of that term's meaning; we must also know what evaluations they use it to make, that is, which properties they take to make distributions good. And to know the term's actual extension, we must know which properties in fact make distributions good.¹²

For another illustration of this pattern, consider the concept 'selfish' in a sense that concerns acts apart from their motivations and indicates a specific ground of wrongness. (We analyze a different and perhaps more common sense of 'selfish' that does concern motivations using our second pattern below.) Here 'x is selfish' will be analyzed as something like 'x is wrong, and there are properties X, Y, and Z (not specified) that acts have in virtue of somehow bringing about the agent's happiness rather than other people's, such that x has X, Y, and Z, and X, Y, and Z make any acts that have them wrong.' There again can be competing views about what the relevant X, Y, and Z are. A strict

12 McDowell and Williams sometimes blur the distinction between the intended and actual extensions of a thick concept, which can make their positions seem relativist. Our formulations are meant to avoid any hint of relativism by clearly separating the claim that a speaker's intended extension for a thick term depends on his evaluative beliefs from the claim that its actual extension depends on evaluative truths. Our defense of the reductive view could be mounted with respect to either claim, but for clarity's sake we keep them separate.

impartialist will say that any act is wrong that prefers the agent's own happiness to the greater happiness of others, so it is selfish to choose 10 units of happiness for oneself over 11 units for others. But those who grant an 'agent-centred prerogative' permitting each person to prefer his own somewhat lesser happiness will be more sparing in their use of the term: preferring 10 units of one's own happiness to 11 units of another's will not be selfish, because it is permitted by the prerogative, while preferring 10 to 1000 will.¹³ The descriptive part of this analysis again places some restrictions on the term's extension: one cannot call an act that causes gratuitous harm or shows a lack of self-respect 'selfish.' But these restrictions do not fully settle the term's extension, and in particular do not decide between the impartialist and agent-centred views. That depends on which specific property of preferring one's happiness is wrong-making, which is again an evaluative matter.¹⁴

We do not claim that the analyses of concepts like these have a uniquely correct content. Return to the case of 'distributively just' and imagine a utilitarian who says that just distributions are those that result in the most happiness. Is this a legitimate use of the concept? The answer depends on whether the descriptive content of 'distributively just' allows as good-making properties of distributions only their intrinsic properties — in which case this use is a misuse — or also their causal ones. There seems to us no settled truth about this; here the concept's boundaries become vague. And this seems to be mirrored in the term's usage. Faced with the above utilitarian, some (mostly other

13 See Samuel Scheffler, *The Rejection of Consequentialism* (Oxford: Clarendon Press 1982).

14 Stephan L. Burton has proposed an analysis of thick concepts that also involves partial indeterminacy but locates it in a different place. He reads 'x is F' as 'x is good for having some instance of properties X, Y, and Z (for specific X, Y, and Z),' where other instances of the same properties need not make what has them good ('"Thick" Concepts Revisited,' *Analysis* 52 (1992) 28-32). For many terms the specified properties will have to be disjunctive, for example, 'is an equal distribution, or one proportioned to merit, or ...' But Burton denies that one can specify, even in principle, any narrower property such that all and only the instances of X, Y, and Z that do make things good have that property. Whatever property one selects, even one as specific as equality, some instances of it will be good-making while others are not. Apparently influenced by McDowell's argument about uncodifiability, Burton here denies universalizability; his analyses may therefore appeal to those who likewise reject it. But we wish to retain universalizability, as Sidgwick, Hare, and others did, and therefore take a different line. We do not hold, as Burton does, that the sense of a thick term fully specifies a set of properties; it leaves them partly indeterminate. But there is always some property that fully determines the term's extension, in the sense that all and only instances of it are in the relevant way good-making.

utilitarians?) will say he is offering a legitimate rival view of distributive justice, others that the best description of the view is that it dispenses with justice. We think there are similar vaguenesses, even at multiple points, in many other thick concepts. But there is also always some determinateness: to call a distribution 'distributively just' because it resulted from generous actions is clearly to misuse the concept, while to call it 'just' because it is equal is clearly permitted.

This first pattern of analysis places concepts like 'distributively just' and 'selfish' in a middle position between purely thin ones and fully thick ones such as 'Kraut.' In so doing it explains why anti-reductivists like McDowell have to assume that any reductive analysis of these concepts will have a descriptively determinate two-part form. It is only if the descriptive part of the analysis fully determines a concept's extension that there is no room for evaluations to help do so. Given an only partly determinate descriptive part, however, that room is clearly available.

That there are concepts between the purely thin and fully thick, so instead of a sharp division there is a smooth continuum, is also noted by Samuel Scheffler, though he does not question these concepts' irreducibility.¹⁵ But something essentially equivalent to our first pattern of analysis has been proposed by Allan Gibbard in his answer to the disentanglement argument. Gibbard's candidate thick judgment is 'x is lewd,' which he analyzes as follows: 'L-censoriousness [a special feeling of shock and censure that goes with finding something lewd] toward the agent is warranted, for passing beyond those limits on sexual display such that (i) in general, passing beyond those limits warrants feelings of L-censoriousness toward the person doing so, and (ii) this holds either on no further grounds or on grounds that apply specially to sexual displays as sexual displays.'¹⁶ The reference to limits on sexual display in this analysis gives the descriptive meaning of 'lewd,' and Gibbard emphasizes that this meaning is only partly determinate. It rules out applying 'lewd' to an act of beating one's wife or acting intellectually arrogant, but it does not specify what the relevant limits are and therefore cannot by itself determine the term's extension. That depends also on the term's evaluative meaning, but Gibbard's specification of this incorporates three aspects of his more general meta-ethical theory that may make his central point harder to see. One is his

15 Scheffler, 'Morality Through Thick and Thin: A Critical Notice of *Ethics and the Limits of Philosophy*, *Philosophical Review* 96 (1987) 411-34, at 417-18.

16 Allan Gibbard, 'Thick Concepts and Warrant for Feelings,' *Proceedings of the Aristotelian Society, Supplementary Volume* 66 (1992) 267-83, at 280-1

view that all moral concepts, including thin ones such as 'right' and 'wrong,' can be analyzed in terms of the warrant or appropriateness of certain feelings. The second is that there is an irreducible plurality of such feelings, so what is appropriate to lewd acts is not some generic con-attitude but a specific attitude of L-censoriousness. (Here Gibbard departs from the classical non-cognitivists, who thought there was only one basic pro- and one basic con-attitude.) And the last is his non-cognitivist view that judgments about warrant do not state truths but express one's acceptance of certain norms. In our discussion, we are abstracting from issues about cognitivism vs. non-cognitivism and treating the thin concepts as unanalyzable. And if we remove the elements of Gibbard's analysis that depend on his specific metaethics, we get something like the following: 'x is lewd' means 'x is wrong, and there are properties X, Y, and Z (not specified) that involve somehow passing beyond limits on sexual display, such that x has X, Y, and Z, and X, Y, and Z make any act that has them wrong.' That is essentially our analysis.¹⁷

The central point of Gibbard's analysis has been recognized by one anti-reductivist, Jonathan Dancy, but he dismisses it. To Gibbard's claim that a thick term does not have enough descriptive meaning to determine its extension, Dancy replies:

I just don't understand the notion of 'not enough descriptive meaning.' Any amount of descriptive meaning is enough descriptive meaning to be the whole meaning of some term, even if it does not make for a very interesting or useful one. If a term has descriptive meaning, its descriptive 'aspect' is capable of standing alone as a neutrally descriptive concept.¹⁸

But consider the descriptive meaning of 'x is distributively just': 'there are some properties X, Y, and Z (not specified) that distributions have as distributions, or in virtue of their distributive shape, and x has X, Y, and Z.' It implies that x is a distribution and can therefore be said to determine an extension, namely one including all and only distributions. But this is not the extension of the term 'distributively just.' To specify *that* extension we have to add an evaluative judgment saying which X,

17 Though Gibbard's analysis of 'lewd' makes it involve evaluations of the agent, it does not refer to her motivations but seems to concern only her acts and whether, motivation apart, they exceed some limits. For us this makes Gibbard's 'lewd' not a virtue-concept but one suited to the first pattern of analysis and explicable using 'wrong.'

18 Dancy, 'In Defense of Thick Concepts,' 275

Y, and Z are good-making and therefore which distributions are just. Dancy simply repeats the basic error behind the disentanglement argument, assuming that if the descriptive part of a thick concept does any work in determining its extension it has to do all the work. But the point of the analysis that we, like Gibbard, have proposed is that even if the descriptive part of a thick concept determines *some* extension, it does not contain enough to determine *that concept's* extension.

This first pattern of analysis fits many thick concepts, but we do not believe it is adequate to capture virtue-concepts such as 'courageous' and 'kindly.' This is important, because these concepts have been central to the anti-reductive case. They call for a second pattern of analysis, which likewise reduces thick concepts to thin ones plus description but does so in a different way.

This second pattern involves a three-part analysis, because it supplements the global thin evaluation that governs the whole concept (the 'x is good ...' or 'x is right ...' of the first pattern) with a further thin evaluation that is embedded within the descriptive content. Its presence means that we cannot determine the extension of the thick concept without determining the extension of the embedded thin one, that is, without making some evaluations.

To illustrate this second pattern, consider the virtue of integrity. It is a morally good trait that involves, roughly, sticking to one's ideals and projects despite temptations or distractions. But not any fidelity to a project counts as integrity. Someone who persists in building his beer-mat collection despite the rise of Nazism around him and the temptation to fight against it would hardly be described as acting with integrity. The reason is that only fidelity to good or important goals counts as integrity. An initial analysis of 'x is an act of integrity' therefore runs something like 'x is good, and x involves an agent's sticking to a significantly good goal despite distractions and temptations, where this property makes any act that has it good,' and where the second 'good' indicates an embedded evaluation. Given this analysis, we can only know what counts as integrity if we know which goals are independently good, and there can be disputes about this. Consider a pope who remains faithful to the views on homosexuality and the ordination of women that he and his church have long held, despite calls from modernizers within and outside the church for change. Supporters of traditional Catholic teaching may say his resistance to reform shows integrity; those who reject that teaching will not.

Or consider courage, often cited by non-reductivists as a paradigm thick concept. It involves accepting harm or the risk of harm to oneself, but not all accepting of harm is courageous. The stand of the Spartan soldiers at Thermopylae was courageous, but resisting a robber's demand for 'A nickel or your life' is foolhardy. And the difference between the

two seems obvious. Accepting harm is courageous when the reasons, and especially the moral reasons, for doing so outweigh the reasons against, and it is foolhardy when they do not. In many though not all cases these reasons concern the goods that can be achieved by risking the harm. The stand of the Spartan soldiers was courageous because the good of preserving Greek civilization against Persian invasion was far greater than the evil of the soldier's deaths, while resisting the robber is foolhardy because retaining one's nickel is trivial compared to one's death. So for these cases 'act x is courageous' can be analyzed as something like 'x is good, and x involves an agent's accepting harm or the risk of harm for himself for the sake of goods greater than the evil of that harm, where this property makes any act that has it good,' and where, again, the second 'good' is an embedded evaluation. Given this analysis, it is impossible to determine the extension of 'courageous' without knowing what count as good goals, a topic about which there can again be disagreements. Imagine someone who sacrifices her career to make a purely symbolic protest against a wrong her company has committed, a protest that will lead to no further benefits. Those who think symbolic acts have significant value in themselves may describe her act as courageous; those who recognize only instrumental reasons for acting will call it foolhardy.

We think this three-part analysis applies very widely, since many and even all virtues and vices involve a relation to some independently given moral consideration, often (though not always) an independent good or evil. Thus, benevolence involves a desire for another's good and malice a desire for his evil, with the extension of these concepts depending on what in particular is good or evil. Since most people regard happiness as good, most count a desire for another's happiness as benevolent. But what about a desire that she know important truths even when doing so will not affect her happiness? If knowledge is good apart from its effects, then this desire may count as benevolent; if not, it cannot. Something similar holds for the virtue of distributive justice. It is the trait of desiring and trying to bring about just distributions, and we cannot tell which people have it unless we know which distributions are just, for example, whether they are equal or based on desert. Or consider the vice of cruelty, which Hilary Putnam says the reductive view cannot accommodate. Like McDowell and Williams, he assumes that a reductive analysis of cruelty will have a descriptively determinate two-part form, and considers the suggestion that its descriptive part will read 'causes deep suffering.' Against this he writes:

Before the introduction of anaesthesia at the end of the nineteenth century, any operation caused great pain, but the surgeons were not normally being *cruel*, and behaviour that does not cause obvious pain at all may be extremely cruel. Imagine

that a person debauches a young person with the deliberate aim of keeping him or her from fulfilling some great talent! Even if the victim never feels obvious pain, this may be extremely cruel.¹⁹

Putnam's first objection in this passage requires just an obvious refinement to the descriptive part of the analysis. As a vice-concept, cruelty turns on the motives behind an act rather than its causal consequences; otherwise a falling rock could be cruel. And if this part reads 'causing great pain either from desire for the pain for its own sake or with indifference to it,'²⁰ the nineteenth-century doctors were not being cruel. More importantly, his second objection is met by replacing the specific reference to pain in the analysis with an embedded evaluation. If cruelty involves causing some *evil* from desire or with indifference, and a young person's failing to develop her talents is either an evil or the deprivation of a great good, then the debaucher's act is indeed cruel. What Putnam says cannot be captured is perfectly easily explained.

Many virtues involve not just one attitude but a combination of attitudes that are properly proportioned to their objects' values, as their opposing vices are not. Thus courage involves caring more about a greater good such as preserving Greek civilization than about the lesser evil of one's death, while cowardice involves the opposite. Similarly, unselfishness involves properly balanced desires for one's own and others' goods, which on the most common view are desires that weigh the two equally. The contrary vice of selfishness, which involves caring too much about one's own good, differs from the selfishness analyzed earlier using the first pattern, since it concerns a person's motivations and not the rightness or wrongness of his acts. Imagine that someone prefers 10 units of his own happiness to 11 units for another person. His motivation is on the view we are currently considering somewhat selfish, and therefore if not positively vicious at least not ideally good. But if there is an agent-centred prerogative, his acting on this motivation may not be at all wrong: his act will be selfish in the sense concerned with vice but not in the one connoting wrongness. For all these virtues of proportion, however, the embedded evaluation concerns not a single value but a comparison among values, so the extension of the relevant concept depends on comparative judgments about which people can differ. Consider nepotism, which we take to be a vice involving a stronger preference for one's children's happiness over strangers' than the

19 Putnam, 'The Entanglement of Fact and Value,' 38

20 The *Oxford English Dictionary* defines 'cruelty' as involving 'delight in or indifference to another's pain'; our refinement builds those psychological elements into the term's analysis.

value relative to oneself of their happiness makes appropriate. Someone who thinks that from a parent's point of view her child's happiness has 10 times as much value as a stranger's will say that caring 20 times as much about the child's happiness is nepotistic; someone who thinks the relevant ratio is 30:1 will not.

These three-part analyses again capture McDowell's key premise, but in a way he does not envisage. When he says that to determine a thick concept's extension we must grasp evaluations of 'actions seen as falling under the concept',²¹ he assumes that the only relevant evaluations govern the concept as a whole and apply to whatever falls under it, as the first evaluative component of our analyses does. Williams does the same when he talks of grasping a thick concept's 'evaluative point' in the singular.²² But the second evaluative component does not function in this way. If we analyze compassion as sorrow at another's evil, the term 'evil' does not apply to the sorrow; it applies to its intentional object. Similarly, if judging the Spartans' sacrifice to be courageous requires judging the preservation of Greek civilization to be good, that second evaluation concerns not the Spartans' act but its goal. It therefore, and importantly, has implications beyond the context of courage. If there are properties that made Greek civilization significantly good, they were worth promoting even when there was no threat of war and no call for courage, so the embedded evaluation is relevant to many more than just courage-judgments. Similarly, the embedded evaluation for 'integrity' concerns not a person's sticking to his goal but the quality of that goal itself and so also has further implications, for example, for how much others should support his pursuit of it with money.

The idea that the virtue-concepts involve embedded evaluations is a distinctive one that we cannot defend fully here.²³ But it was accepted and applied by all the early 20th-century reductivists. Responding to the suggestion that ethical theory should use irreducible virtue-concepts, Sidgwick says 'our notions of special virtues do not really become more independent by becoming more indefinite; they still contain, though perhaps more latently, the same reference to 'Good' or 'Wellbeing' as an ultimate standard.' This is especially clear, he says, when we contrast a virtue with its associated vice, for example, courage with foolhardiness.²⁴ Sidgwick takes the relation to 'good' that defines the virtues to

21 McDowell, 'Non-Cognitivism and Rule-Following,' 144

22 Williams, *Ethics and the Limits of Philosophy*, 142

23 See Thomas Hurka, *Virtue, Vice, and Value* (New York: Oxford University Press 2001) for such a defense.

24 Henry Sidgwick, *The Methods of Ethics*, 7th ed. (London: Macmillan 1907), 392-3

be causal, so they are traits that result in something good. This is also Moore's view in Ch. 5 of *Principia Ethica*, but elsewhere he applies virtue-terms to states with an intentional relation to the good, as in loving or admiringly contemplating something good. And that intentional relation is central to the accounts of virtue of Rashdall, Ross, and Broad. Rashdall grounds the goodness of virtue in the 'idea of the intrinsic worth of promoting what has worth,'²⁵ where the second 'worth' is an embedded evaluation, while Ross recognizes three forms of virtue: the desire to do what is right, the desire to bring into being something good, and the desire to produce some pleasure for another being, where that pleasure is good and the basis of a virtue only because it is good.²⁶ Each of Ross's three desires involves an intentional relation to something that falls under a thin concept, either rightness or goodness, and we cannot identify the resulting concepts' extensions without first identifying the thin ones' extensions.

This second way of capturing the key premise of the disentanglement argument differs from the first, which involved partly determinate descriptive content, but the two can and often should be combined. We have said that many virtues and vices involve a relation to some independently given good or evil, but there can be disputes about what this relation is. Compare Sidgwick's view that the virtues involve a causal relation to something good with Rashdall's and Ross's view that they involve an intentional relation. We do not think the everyday concept of virtue mandates either one of these views in preference to the other; each gives a possible specification of a concept of virtue that is neutral between them, as the concept of distributive justice is neutral between egalitarian and desert-based specifications. An analysis of, say, 'x is benevolent' will therefore say 'x is good, and there is some relation R (not specified) that is a positive or favouring relation, such that x stands in R to something good, and standing in R to something good makes anything that does so good.' The exact extension of 'x is benevolent' will then depend on which positive relation is good-making, and here the two views can come apart. Consider a world where the desire to cause others pleasure somehow regularly causes them pain. An intentional view like Rashdall's will call this desire benevolent and virtuous; a causal view like Sidgwick's will not. While each view uses the same partly determinate concept of virtue, each specifies it in a different way and therefore generates a different extension for 'benevolent.'

25 Hastings Rashdall, *The Theory of Good and Evil*, 2 vols. (London: Oxford University Press 1907), vol. 1, 59

26 W.D. Ross, *The Right and the Good* (Oxford: Clarendon Press 1930), 161

There can be further such indeterminacies that call for the first pattern. For a person's pursuit of a goal to show integrity or be courageous, must her goal be in fact good or is it enough if she believes it to be good, assuming her belief is (note the further embedded evaluation) non-culpable? Again the concepts of integrity and courage seem not to settle this issue, but require us to decide evaluative questions. Moreover, many of the concepts discussed in connection with the first pattern of analysis also require the second. Justice, for example, is a good of distribution, but not all things are such that their distribution gives rise to issues of justice. If someone said it was unjust that some people have more hair on their forearms than others, he would, we take it, be misusing the concept. This is because only distributions of things that have value, either intrinsic or instrumental, raise questions of justice, and there can be disagreements about what these are. Someone with a purely spiritualist theory of value may deny that distributions of money can be unjust, because money has no significant value, but say that inequalities in access to religious teachers can; an atheist with materialist values may say the opposite. And this point holds even more strongly for a desert-theorist. Desert involves awarding goods to people in proportion to their merit, so to know what a particular desert-theorist takes to be distributively just we must know what specific goods he is thinking of distributing (happiness? money?) and what for him constitutes merit (virtue? economic contribution?) In fact, the analysis of many thick concepts will contain multiple points where the concept's extension depends on evaluations, some because its descriptive part is only partly determinate (the first pattern) and some because it contains embedded evaluations (the second). But all these evaluations will involve only thin concepts: they will concern what pattern makes a distribution good, what degree of preference for one's own happiness makes an act wrong, or which goals are good or better than others.

We are not committed to all the details of the above analyses, which are only sketches intended to illustrate the general resources the reductive view has. Nor, to repeat, do we take ourselves to have provided a complete defense of that view or a conclusive refutation of everything anti-reductivists have said. We have tried to answer just one anti-reductive argument, namely the disentanglement argument, but it is commonly presented in the literature as if it were on its own decisive. We believe we have shown that this is not so. The argument starts from the premise that one cannot determine the extension of an interesting thick concept without making evaluative judgments, but if the reductive view is not restricted to descriptively determinate two-part analyses, it can happily agree.

This concludes our main discussion, but we will end by considering two possible anti-reductive responses to it. We will take these responses

to concern the virtue-concepts, which have been central to anti-reductivism, and therefore mainly to address our second pattern of analysis. And we will assume that they grant this pattern's general form, that is, grant that virtue-concepts can be analyzed using embedded thin concepts.

Both responses appeal to the idea of uncodifiability, which they use to supplement claims about disentanglement. We have argued that on its own uncodifiability does not tell against the reductive view: if both thick and thin concepts are uncodifiable, the former can be reducible to the latter and uncodifiable precisely because of that reduction. But the responses we will now consider combine uncodifiability claims with ones about disentanglement to yield what they claim is a compound vindication of the anti-reductive view.

The first of these responses grants our second-pattern analyses but says their appearance is misleading. The analyses suggest that to determine whether, say, someone's sticking to a goal shows integrity we must first, and independently, determine whether his goal is simply and generally good. But this, the response says, gets things backward. To know whether his goal is good in the way relevant to integrity, or what the extension of 'good' in this context is, we must appeal to independent and prior judgments about integrity. It is these thick judgments that determine the extension of the thin one about goodness rather than vice versa, so the thin one is not on its own codifiable. Something similar holds, the response continues, for other virtues. To know whether the goal a person accepts harm or risk to achieve is good, we must appeal to independent thick judgments about courage, and similarly for benevolence, unselfishness, and more. Our analyses suggest that thin judgments have priority over thick ones, but the opposite is true: it is thick judgments about the virtues that determine the truth of thin ones about what is good or right.

The reductive view can certainly grant that thick judgments are sometimes evidence for thin ones. If it is in general benevolent to desire other people's good, and also benevolent to desire their pleasure, then their pleasure must be good.²⁷ But the view denies that this is the only way to arrive at thin judgments; they can also be made independently. And whatever comes first in the order of discovery, it is the embedded thin judgments that are prior in explanation. It is the fact that a goal is good that explains why a person's sticking to it shows integrity, not his action's showing integrity that explains why his goal is good; and the fact that Greece was worth preserving that explains why the Spartans' stand was courageous, not vice versa.

27 For an argument of this type, see Ross, *The Right and the Good*, 135.

The anti-reductive view this response proposes is certainly logically possible, but we find its claims about the thin concepts deeply implausible. The view implies that we can have determined that a goal is good in the way that bears on courage, and determined that it is good in the way that bears on benevolence, yet not know whether it is good in the way that bears on integrity. For that we need independent thick judgments about integrity. We can even have determined that the goal is good in a way that is independent of thick concepts and relevant to such thin questions as whether we ought to help someone else pursue it, say, by offering him money.²⁸ But we still cannot know whether it is good in the way relevant to integrity without separate judgments about integrity. ('I know this goal is worth dying for, but do not yet know whether it is worth sticking to despite distractions.') What reason is there to accept this radical pluralism about 'good'? If the same thin concept appears in the analyses of all these virtues, should its extension not be determined in the same way in them all? And how does the view explain the obvious connections between the virtues? Surely if the Spartans' stand against the Persians was courageous, their refusal to abandon it in the face of offers of money would show integrity. The reductive view explains this connection easily: if the preservation of Greece was sufficiently good to make fighting for it courageous, it was sufficiently good to make sticking to it an instance of integrity. But how can the connection exist if the good relevant to integrity is different from that relevant to courage? Or, if it does exist, how can it be anything other than a coincidence? Something like the pluralism this response proposes might make sense given the original anti-reductive view that thick concepts blend descriptive and evaluative meaning in an indissoluble whole. Then integrity, benevolence, and courage would be irreducibly distinct concepts and it would be natural for their extensions to be determined differently. But if the analysis of each concept uses the same concept of goodness, should that concept not function in the same way in each?

This may for some suggest a return to the original anti-reductive view with its descriptive-evaluative blends and rejection of our patterns of analysis. But the question then is what the argument for that original view is. It cannot be just uncodifiability, for we have shown that that is perfectly consistent with the reduction of thick to thin. Nor can it be just the disentanglement argument, since there are two answers to that.

28 Some versions of the anti-reductive view may deny that such independent judgments of goodness are possible, since all thin judgments derive ultimately from thick ones. But these are more radical versions of the anti-reductive view than are commonly defended.

And without some positive anti-reductive argument, we see no reason not to accept the three-part analyses, and given those analyses, no reason not to think the extensions of their thin components are purely thinly determined.

The second response to our argument grants that the extensions of the thin concepts in our analyses are thinly determined, but holds that the virtue-concepts so analyzed are uncodifiable not at one but at two distinct points. One is in their embedded thin concepts, as we have discussed. But a second arises in the move from those concepts to the embedding thick ones. Even if we know the extension of 'good' as that bears on integrity, this response says, we cannot determine the extension of 'integrity' without making further evaluative judgments that themselves cannot be codified. And this second uncodifiability blocks the reduction of thick to thin.²⁹

This is another logically possible anti-reductive view, but we again see no reason to accept it. It is more complex than any defended in the literature, since it involves two separate uncodifiabilities in a single concept. (There is no hint of this in McDowell, for example, since he does not consider embedded evaluations.) And the question as always is what the argument for the second uncodifiability is. It could again be the disentanglement argument: knowing the extension of 'good' and the descriptive meaning of 'integrity' is not sufficient to determine the extension of 'integrity,' which requires further evaluative judgments. But the reductive view can again accept the premise of this argument if it combines the second pattern of analysis with the first, as we have argued it often should. Then the descriptive content of 'integrity' will be only partly determinate, saying, for example, that integrity involves sticking to a goal in a way somehow related to goodness while leaving it undetermined whether that relation is causal (as for Sidgwick) or intentional (as for Rashdall and Ross), and whether the goal must be in fact good or merely one a person non-culpably believes is good (with further specification then needed of 'non-culpably'). Given a mixed analysis of this sort, there will already be a significant gap between the descriptive and thin evaluative content of a virtue-concept on the one hand and the concept as a whole on the other, due to the partial indeterminacy of the descriptive content. And evaluative judgments will be required to fill that gap, so the key premise of the disentanglement argument is again accommodated. A defender of the second response may say that not all the second uncodifiability can be captured in this way; there is more, and that more blocks any thick-thin reduction. But

29 We owe this second response to Simon Kirchin.

what is the argument for this claim? What reason is there to believe that more evaluative judgments are needed to determine the extension of 'integrity' than those allowed for by the mixed analysis, and what exactly are those judgments? At this point the anti-reductivist seems, to quote Aristotle, like someone 'maintaining a thesis at all costs.'³⁰

While there may be many versions of the anti-reductive view, our question has been what the argument is for adopting that view in the first place, and we have therefore considered the most common argument in the literature. It starts by assuming that any reductive analysis of a thick concept will have a particular, descriptively determinate two-part form. We grant that the argument succeeds against analyses of this form and also grant that some reductive theorists have given such analyses. But a serious critique of a philosophical view cannot consider only what its proponents have actually said; it must also ask whether there are better formulations the view could be given. We believe the anti-reductivists have not done this work. Just a little reflection shows there are at least two forms a reductive analysis can take other than the descriptively determinate two-part one, and given these forms it is indeed true, on at least two possible grounds, that the extension of any interesting thick concept depends on evaluations. The key premise of the disentanglement argument does not tell against the reductive view but is perfectly compatible with it. Of course rebutting a common argument against a philosophical view does not count as making a positive case for that view. But we believe that if the reductive view's resources were more widely appreciated, it would be more widely embraced.

The issue about thick and thin concepts is philosophically important. If the anti-reductive argument we have discussed were sound, then all those metaethicists, cognitivist and non-cognitivist alike, who have taken their principal task to be that of understanding judgments like 'x is good' and 'x is right' would be fundamentally misguided. Likewise, all those normative ethicists who have formulated moral theories using 'good' and 'right' — all the utilitarians, perfectionists, deontologists, and more — would have no clue how to think about ethics. These are very strong conclusions to draw from a brief argument about the extensions of moral concepts. In our view this argument rests on an uncharitable assumption about the view being opposed and, once that assumption is exposed, has not a shred of force.

Received: July 2008

30 Aristotle, *Nicomachean Ethics*, trans. D. Ross (Oxford: Oxford University Press 1980), 1096a2

