IS MORAL REASONING CONCEPTUAL INTERPRETATION?

RICHARD H. FALLON, JR.[*]

INTRODUCTION

In *Justice for Hedgehogs*, Ronald Dworkin advances not only the general thesis that "interpretation" is the ubiquitous methodology of reasoning, arguing, and knowing in the domain of value, but also the more specific proposition that moral and ethical reasoning are species of "*conceptual* interpretation."[1] In this short Essay, I shall explore the latter claim.

Justice for Hedgehogs is a book of stunning ambition and brilliance. It deals systematically with many of the largest issues in moral philosophy, ethics, political theory, metaethics, epistemology, and the philosophy of mind. Because of the interconnections among the topics that Dworkin covers and the mutual support that his various arguments afford one another, it is nearly impossible to assess any of his claims without struggling with, and either accepting or rejecting, the overarching philosophical position that he, as a self-advertised hedgehog, very deliberately presents as "one big thing."[2]

[1] *See* RONALD DWORKIN, JUSTICE FOR HEDGEHOGS (forthcoming 2010) (Apr. 17, 2009 manuscript at 101, on file with the Boston University Law Review) (emphasis added).

[2] *Id.* (manuscript at 7).

To the greatest possible extent, however, I shall try to elide the large challenge of overall appraisal of *Justice for Hedgehogs*'s central arguments. Focusing as narrowly as possible on Dworkin's claim that moral reasoning is conceptual interpretation, I shall raise three sets of questions. First, what does it mean to say that moral and ethical reasoning are species of conceptual interpretation? Second, if moral and ethical reasoning are a form of conceptual reasoning, is all moral reasoning conceptual in the same sense? More pointedly, should we view ordinary people who need to make non-deliberative moral judgments as engaging in the same kind of conceptual reasoning as moral philosophers writing in journal articles? Third, what is the relationship between Dworkin's thesis that moral reasoning is conceptual interpretation and his thesis asserting "the unity of value"?[3] What does he mean when he says, in explication of the unity-of-value thesis, that "all true values form an interlocking network"?[4] Does the methodological premise that moral reasoning is conceptual interpretation entail or presuppose the conclusion that all genuine values cohere harmoniously? This third set of questions may shed light on the nature of the distinction between "hedgehogs" and "foxes"[5] that Dworkin employs as an organizing motif – and, what is more, may suggest some reasons for skepticism about how far that distinction can usefully be pressed.

I.   INTERPRETATION (GENERALLY) AND CONCEPTUAL INTERPRETATION (SPECIFICALLY)

At the heart of *Justice for Hedgehogs* lies a distinction between two grand domains of human thought and inquiry.[6] First, there is the domain of science, in which scientists pursue knowledge on the assumption that true beliefs are somehow caused by elements of the physical universe.[7] Second, there is the domain of interpretation, which roughly, but not precisely, corresponds to the realm of value.[8]

In both science and interpretation, we form beliefs and make assertions. In the realm of science, I may say, "Heat rises." The question then may come up, "Is it true that heat rises?" In answering this question, or making sense of what it would mean for it to be true that heat rises, we might say it is true that heat rises if, but only if, certain states of affairs, sometimes called truth conditions, obtain. The claim that "It is true that heat rises" is a second-order claim that asserts something about the first-order claim that heat rises – namely, that it is true.

---

[3] *Id.*

[4] *Id.* (manuscript at 76).

[5] *See id.* (manuscript at 7) ("[T]he fox knows many things, but the hedgehog knows one big thing.").

[6] *See, e.g.*, *id.* (manuscript at 79, 99).

[7] *See id.* (manuscript at 98).

[8] *See id.* (manuscript at 99-100).

By contrast, with respect to claims of value, Dworkin maintains that "there are no sensible independent, second-order, metaphysical questions" to be asked or answered.[9] If I assert that torturing puppies is wrong, and someone asks whether it is true that torturing puppies is wrong, or what in the world makes torturing puppies wrong, my interlocutor has made a mistake. Dworkin contends that a precept he terms "Hume's principle" – which holds that it is impossible to derive a proposition of ultimate value from a proposition of fact – shows the foolhardiness of second-order inquiries.[10] But that insight does not, says Dworkin (contrary to what many philosophers have thought), embarrass claims of worth and value, or of right, duty, and obligation.[11] From Hume's principle, he instead draws the conclusion that truth in the domain of value is a matter of conviction and argument.[12]

To support conviction and argument in the domain of value, and ultimately to underwrite the idea of truth, Dworkin introduces the concept of "interpretation" as the distinctive methodology of moral reasoning and persuasion (as it is also the distinctive methodology of reasoning and persuasion in a variety of other non-scientific practices[13]). His general theory of interpretation is complex and multi-dimensional. Here, I shall mention only a relative few of its aspects.

First, interpretation is a social, practice-based phenomenon,[14] which proceeds through the ascription of value. Someone engaging in interpretation begins by identifying a pertinent practice – such as law or literary criticism or, if I understand Dworkin correctly, morality or ethics – and then ascribes to it whatever point or value would reveal the practice in the best moral light. The ascription of value to ongoing social practices endows interpretation with a dialectical aspect. On the one hand, practices are inherently social and collective. On the other hand, the interpretation of practices is often "argumentative," with judgments about what would count as the best interpretation rooted in personal values and convictions.[15]

Second, it is crucial to Dworkin's theory that many important practices – which he characterizes as "interpretive" – require, or are structured by, or even

---

[9] *Id.* (manuscript at 9).

[10] *See id.* (manuscript at 19).

[11] *See id.*

[12] *See id.* ("Our moral convictions can finally be sustained or challenged only by other convictions and arguments drawn on that dimension.").

[13] *See, e.g.*, *id.* (manuscript at 79) ("Historians interpret events and epochs, psychoanalysts dreams, sociologists and anthropologists societies and cultures, lawyers documents, critics poems, plays and pictures, priests and rabbis sacred texts.").

[14] *Id.* (manuscript at 83-84).

[15] *See id.* (manuscript at 82) (describing the difficulty of explaining one's own interpretation of a poem or a piece of music, yet observing that "the distinctive truth-seeking and argumentative phenomenology of interpretation survives").

may be constituted by the use of "concepts."[16] When he writes that moral reasoning is conceptual interpretation, he thus immediately refers to the "great variety of moral concepts" that we use in moral reasoning, such as "the concepts of reasonableness, for instance, honesty, trustworthiness, tactfulness, decency, responsibility, cruelty, shabbiness, insensitivity, deceit, and brutality, as well as the special political concepts of legitimacy, justice, liberty, equality, democracy and law."[17] Once again, shared concepts anchor interpretation in a public, social domain, but in one that tolerates significant disagreements. According to Dworkin, "[w]e develop our [different and distinctive] moral personalities through interpretations of what it is to be honest or reasonable or cruel, or what actions of government are legitimate, or when the rule of law has been violated."[18]

Third, interpretation is a creative, holistic activity, in which we seek a reflective equilibrium among all of our convictions and principles and our interpretations of practices, objects, and concepts.[19] What will count as the best interpretative understanding of anything in particular will depend on various candidate interpretations' coherence with our convictions about what is valuable in other domains or practices. As Dworkin puts it, "[t]here is no way that I can test the accuracy of my moral convictions except by deploying further moral convictions."[20]

Fourth, interpretation is pervasively anti-skeptical in its assumption that there is one right answer to questions of value or interpretive truth. Interpretation begins with practices, and the concepts that structure them, and with convictions that we already hold because there is, literally, nowhere else to start. There are no entities in the world, discoverable through the tools of science, that would either validate or falsify what we believe, and cannot help believing, about what is good and right and valuable. Nevertheless, truth is what we seek, with the concept of responsibility providing "a kind of moral epistemology."[21] Dworkin maintains that "[w]e achieve responsibility when and to the extent we order our convictions so that each supports the others in an overall network of value that we embrace authentically; we pursue responsibility so understood by interpreting abstract convictions so as to create an active integrity among them."[22]

---

[16] *See id.* (manuscript at 68) ("Conceptual analysis of an interpretive concept is therefore itself an exercise in moral theory. The concept of a moral principle or ideal is an interpretive concept.").

[17] *Id.* (manuscript at 101).

[18] *Id.*

[19] *See id.* (manuscript at 86, 99).

[20] *Id.* (manuscript at 65).

[21] *Id.* (manuscript at 66).

[22] *Id.*

II. GETTING DOWN TO CASES

In order to explore the meaning and implications of Dworkin's claim that moral and ethical reasoning are forms of conceptual interpretation, it will help to have some test cases in mind.

Case 1: You know that millions of Africans suffer grievously, and that many thousands are likely to die, from dislocation and malnutrition. You must decide whether to give money to charity, and, if so, in what amount and to which charity.[23]

Case 2: A boat capsizes offshore. At the front, two passengers cling to one life preserver. At the back, a single passenger clutches a second life preserver. Sharks encircle both the pair of passengers at the front and the single passenger at the back. You are on shore with a small boat. You have enough time to rescue either the two passengers at the front or the single passenger at the back, but not both. The single passenger is your wife. You must decide whom to rescue.[24]

Case 3: A friend whose book you have just read, and adjudged to be poorly reasoned and badly written, asks you if you think it is good. You must decide how to answer.[25]

III. WHAT IS CONCEPTUAL INTERPRETATION?

As I try to understand how reasoning about any of the cases that I laid out above should be conceived as "conceptual interpretation," three possible specifications of that claim stand out. The first is what I shall call a "weak, definitional" version. Its meaning and plausibility will come into view if we consider the thought process of an act utilitarian or of an adherent of a religious morality who cares only about God's will in reasoning about Case 1. Neither of these people is likely to understand herself as putting concepts or conceptual reasoning – about, to quote Dworkin again, such concepts as "decency, responsibility, cruelty, shabbiness, insensitivity, deceit, and brutality"[26] – at the forefront of her analysis. One will care only about what

---

[23] Dworkin discusses a similar case in *Justice for Hedgehogs. Id.* (manuscript at 179).

[24] *Id.* (manuscript at 180-82) (discussing a quandary wherein all three passengers are strangers, as well as a scenario in which the single passenger is the rescuer's wife).

[25] Dworkin suggests a similar case when he alludes to the plausibility of the claim that values conflict:

There is no reason to think, certainly not in advance of a great deal of reflection, that values are always nicely knit together in the mutually accommodating way that hedgehogs imagine. On the contrary, it seems more plausible that values conflict – as they certainly seem to do, for instance, when it would be an act of kindness to tell someone a lie or when the police can save some people from a terrible death only by torturing other people.

Id. (manuscript at 9).

[26] *Id.* (manuscript at 101).

would best promote overall happiness, while the other will think only about God's wishes.

Nevertheless, each, on Dworkin's account, will at least have interpreted the practice of morality. The utilitarian will have done so to determine that the point of morality is to promote utility, and the religious moralist to determine that the purpose of morality is to advance God's will. Dworkin suggests another way to uphold a weak, definitional version of the claim that moral and ethical reasoning are species of conceptual interpretation when he refers to "the concept of . . . what we ought to do."[27] In deciding what we ought to do, perhaps we all necessarily interpret "the concept of . . . what we ought to do." With a sufficiently capacious understanding of what a concept is (such that "what we ought to do" counts), and of what it means to engage in "conceptual interpretation" (such that determining when a concept applies suffices for the label to fit), the claim that all moral reasoning is conceptual reasoning seems tautologically true.

If this were all that Dworkin means when he says that moral reasoning is conceptual interpretation, however, his thesis would be relatively weak and uninteresting. It would tell us nothing about how a moral reasoner ought to proceed. Instead, his conclusion would be built into stipulated definitions of what concepts and conceptual reasoning are.

On a second possible interpretation, Dworkin's claim that moral reasoning is conceptual interpretation would simply be an iteration of his claim that interpretation is a holistic process in which we seek to bring all of our moral convictions into a mutually supportive harmony or reflective equilibrium.[28] To illustrate a "reflective equilibrium" interpretation of the claim that moral reasoning is conceptual interpretation, we can again imagine an act utilitarian and a religious moralist reasoning about what to do in Case 1. Each might reason unhesitatingly from her moral principles to a conclusion, but it is also possible that either might experience an initial dissonance between her principles and her tentative convictions about what would be best to do in the particular case. If so, a two-way, back-and-forth deliberation would ensue, in which either might adjust her specific conviction about what she ought to do in the case at hand, or alternatively, might revise her utilitarian or religious principles.

The problem with this account of Dworkin's claim that moral reasoning is conceptual interpretation is that it does not insist that good moral reasoning should be *conceptual* in any very interesting or distinctive way. It simply applies the Rawlsian idea of reflective equilibrium as the methodology of moral reasoning.[29]

---

[27] *Id.* (manuscript at 114).

[28] *See, e.g.*, *id.* (manuscript at 37-40, 66, 170).

[29] JOHN RAWLS, A THEORY OF JUSTICE 18 (rev. ed. 1999) ("It is an equilibrium because at last our principles and judgments coincide; and it is reflective since we know to what principles our judgments conform and the premises of their derivation.").

It thus seems more plausible to adopt an interpretation of Dworkin's claim that moral reasoning is conceptual reasoning that is both more specific and, I believe, farther reaching than the reflective equilibrium interpretation. This third version, which I shall call a "*conceptual* equilibrium" interpretation, maintains that moral reasoning should afford a central role to concepts such as, to quote him again, "reasonableness, for instance, honesty, trustworthiness, tactfulness, decency, responsibility, cruelty, shabbiness, insensitivity, deceit, and brutality."[30]   This interpretation further maintains that good moral reasoning should aspire to achieve harmonious conceptual understandings such that there is no conflict among concepts marking genuine values and, what is more, that the judgments of value embedded in otherwise different concepts stand in a relation of mutual support with one another.

To see the difference between the bare claim that moral reasoning aims to achieve reflective equilibrium among general moral principles and specific moral convictions and the further claim that it aspires to achieve non-conflicting, mutually supportive interpretations of morally laden *concepts*, consider my Case 2, in which you need to determine whom to save after a boat has capsized. If you begin with a strong, pre-reflective conviction that you ought to save your wife, you might, on the reflective equilibrium model, attempt to develop an account of the principles bearing on duties of rescue. You might then work back and forth between plausible statements of those principles and a judgment about what you ought to do until your principles and your judgment about the particular case were intellectually acceptable and harmonious. But moral concepts (such as, once more, honesty, decency, trustworthiness, and so forth) would not necessarily play a central role in your quest for intellectual equilibrium.

Under a conceptual equilibrium model, by contrast, moral concepts should loom large in your analysis. In thinking about what is right or wrong, or what you are obliged or forbidden to do, you might, for example, consider the bearing and application of such concepts as love, devotion, trust, betrayal, and responsibility, all of which might push you toward the conclusion that you should save your wife and that it would be right to do so. Then you might take into account the concept of equality of moral worth, which might initially seem to tug the other way: how, consistently with respect for the equality of every person's moral worth, could you justifiably choose to save your wife in preference to two others whose lives are, objectively speaking, equally valuable? With this question having been raised, you, like Dworkin, might conclude upon further reflection that respect for every person's equal moral worth does not, when that ideal is properly understood, require you to prefer saving the two to saving your wife after all.[31]   If so, you would not, in Dworkin's view, necessarily have engaged in any discreditable compromises in order to reach your conclusion. To the contrary, it is by viewing our set of

---

[30] DWORKIN, *supra* note 1 (manuscript at 101).

[31] *See id.* (manuscript at 181).

moral concepts and convictions as a whole, and so interpreting them as to bring them into a harmonious alignment with one another, that we most responsibly seek to determine what we owe both to others and to ourselves.[32]

In my view, the best interpretation of Dworkin's claim that moral reasoning is conceptual reasoning equates it with the strong claim that moral reasoning involves the search for a broad, harmonious alignment of interpretations of concepts such that no conflict exists among them and that the judgments of value that they reflect are mutually supportive. In the remainder of this Essay, I shall therefore assume this interpretation to be correct. I am not wholly confident in doing so, however, and therefore want to leave my question about the precise meaning of Dworkin's claim in the status of a question.

There is an additional, even more important reason for my insistence that what Dworkin means when he says that moral reasoning is conceptual interpretation should be viewed as an open question. On what I shall call an "exclusive interpretation," the claim that moral reasoning is conceptual interpretation would hold that moral reasoning is pervasively conceptual and interpretive and includes no non-conceptual, non-interpretive elements. On a "non-exclusive interpretation," Dworkin would maintain only that all moral reasoning is conceptual and interpretive at least in part, but would not necessarily deny that it can sometimes include empirical and predictive aspects.[33] I shall take up the question whether Dworkin's claim that moral reasoning is conceptual interpretation should be regarded as exclusive or non-exclusive in the next Part of this Essay.

IV. WHY "RECONSTRUCT" MORAL REASONING AS CONCEPTUAL REASONING?

My second set of questions about Dworkin's thesis that moral reasoning is conceptual reasoning begins with a query concerning why he might idealize a picture of moral reasoning as the search for a broad conceptual equilibrium. Consider again the case in which you are standing on the shore and must decide whether to save your wife or two other people. In what sense would it bring you closer to a moral ideal if we imagine that, before saving your wife, you sought an equilibrium in your understandings of all of the concepts that might potentially be pertinent to a moral evaluation of your action – including,

---

[32] Dworkin also pursues a conceptual equilibrium approach to moral reasoning in arguing that there is no conflict among liberty, democracy, and equality when these concepts are examined jointly and cast in the best light. *Id.* (manuscript at 220). When we consider liberty, democracy, and equality against the even deeper background values or principles that these concepts serve, we should, he argues, reject any interpretation that would put democracy at odds with liberty. *Id.* Similarly, we should reject any conception of liberty that would result in the conclusion that well-functioning democracies regularly deprive their citizens of liberty. *See id.*

[33] *Id.* (manuscript at 112) ("We need some way of describing truth that is generous enough to include both interpretive and empirical domains . . . .").

for example, love, prudence (perhaps you would need to risk your own life), gratitude, reciprocity, betrayal (which you want to avoid), equality of moral worth, callousness, and indifference?[34]

As one imagines the cascade of evaluative concepts that you might potentially take into account, and with respect to which you might hope to achieve a conceptual equilibrium, it is tempting to say, in a variation of Bernard Williams's celebrated quip, that anyone who thought this way before setting out to save his wife would have not one, but many, "thought[s] too many."[35] Any actual person who tried not to engage in morally fraught action without first conducting a conceptual inquiry with the aim of achieving a broad conceptual equilibrium would be an absurd caricature of over-intellectualized self-consciousness. The thought thus arises that perhaps we do you no favor if, in imagining how you might respond to the case of your drowning wife, we "reconstruct"[36] your thinking as involving a search for conceptual equilibrium.[37]

We should not jump too quickly to that conclusion, however. The danger on the other side, which *Justice for Hedgehogs* emphasizes, is the risk of inviting a well-founded charge of moral irresponsibility.[38] There are some things you could not rightly do to save your wife – for example, wresting the strangers' life preserver away from them to be able to give it to her. This being so, the thought that a particular course of action to save your wife is not morally forbidden is not easily dismissed as "one thought too many." And if this thought belongs in your calculus, others may need to intrude too. More to the point, if the most reliable and responsible form of moral reasoning requires searching for conceptual equilibrium, then we may indeed flatter you if we imagine you engaging in such a search. This, quite apparently, is Dworkin's view. Convictions about the nature and importance of moral responsibility also appear to undergird his assertion that society generally follows a "division of labor" with respect to moral deliberation, and that the exemplars of good

---

[34] Dworkin makes clear that there is a considerable gap between his account of moral reasoning and most people's actual thought processes with respect to moral issues. *See id.* (manuscript at 201). He says, for example, that he does not "intend an absurdity: that when a friend asks you for financial help and you are reluctant you ponder the underlying point of friendship to decide whether you must." *Id.* But he seems to suggest that your response to that friend will, and should, be informed by an interpretive understanding the concept of friendship. *Id.*

[35] BERNARD WILLIAMS, MORAL LUCK 18 (1981).

[36] *See* DWORKIN, *supra* note 1 (manuscript at 107).

[37] Dworkin is quite clear that he intends his theory to apply as much to everyday, practical moral reasoning as to philosophers' systematic reflections. *See id.* (manuscript at 117) ("Since moral concepts are interpretive, both quotidian moral reasoning and high moral philosophy are interpretive exercises.").

[38] *See id.* (manuscript at 10) ("[W]hether or not our convictions are true, we are intellectually responsible in holding those convictions and morally responsibility [sic] in acting on them.").

moral reasoning are moral philosophers – those who, in cool moments, most thoughtfully and imaginatively interpret the practices of morality and ethics and carry out the most intensely rigorous searches for conceptual equilibrium.[39]

Dworkin's admiring claims about moral philosophers and the division of labor inspire more questions, however. In pondering the suggestion that moral philosophers are paragons of moral reasoning because moral reasoning is a species of the conceptual reasoning at which philosophers are adept, we might be moved to reconsider the question of what it means to say that moral reasoning is conceptual interpretation. In doing so, I shall begin by provisionally accepting Dworkin's view that good moral reasoning includes – or at least relies on – a well-thought-out alignment of pertinent moral concepts. But might it also include more?

If moral reasoning consists *exclusively* of conceptual interpretation, then moral philosophers who are skilled at conceptual analysis are almost certain to be better moral reasoners than the rest of us, and we will put ordinary moral thinking in the best light if we imagine that it follows, however crudely, the track that a philosopher would pursue in writing a journal article. Matters may look slightly different, however, if we consider that even if moral reasoning were inherently and necessarily both conceptual and interpretive in part, it might not necessarily be conceptual and interpretive in its entirety. Much practical moral reasoning requires a kind of instinctive, frequently unconscious sizing up of the facts of a situation, including others' psychological states and their likely perception of particular acts or gestures as kind, supportive, funny, creative, insightful, and so forth. In my experience, there are many non-philosophers whose lives exhibit shining moral excellence, not because they are profound conceptual reasoners, but because of their capacities to make insightful and sometimes imaginative psychological and empirical judgments about the effects of possible actions and about ways to transform moments or lives for the better. They perceive possibilities, and means for realizing those possibilities, that others – sometimes including first-rate philosophers – do not.

Given the obvious dependence of moral reasoning on pertinent psychological facts and accurate empirical predictions, I incline toward a non-exclusive interpretation of Dworkin's claim that moral reasoning is conceptual interpretation. But I lack confidence in this judgment, in part because acknowledgment that moral reasoning includes diverse elements, not all within the distinctive expertise of philosophers, might work subtly to undermine the claim that good moral reasoning necessarily includes distinctly conceptual interpretive elements (other than in the sense that I dismissed above as weak and uninteresting). The potential for subversion arises if I am correct in my intuition – which I shall not attempt to defend here – that the kind of empirical and psychological sizing up of situations routinely exhibited by non-philosophers who lead lives of moral excellence does not map well onto a

---

[39] *See id.* (manuscript at 71).

claim that the other relevant dimension of moral reasoning inherently involves a quest for a relatively broad conceptual equilibrium. To be more precise, it would be my intuition that a non-philosopher who quickly sizes up a situation and then acts bravely or kindly would typically be better described as having perceived what she could do effectively to rescue someone from a bad predicament or to alleviate someone's distress, without too much risk or cost to herself, and as having determined that she had reason to do so, than as having made a judgment related to the best interpretation of a range of potentially pertinent concepts. If ordinary moral assessments of what one ought to do in light of the facts are more naturally conducted in the currency of *reasons* than of concepts, then why insist that all moral reasoning is necessarily about the meanings of concepts?

Although I may appear to have ranged far, I have now returned to the question with which I began this Part: why does Dworkin idealize a picture of moral reasoning as the search for a broad conceptual equilibrium, apparently in all cases? There are probably some cases in which moral reasoning is exclusively conceptual. A ready example comes from political philosophers, including Dworkin, who want to determine whether liberty can ever conflict with equality or democracy.[40] But I am far from certain that we flatter ordinary non-philosophers engaged in practical moral reasoning, much less assimilate their thinking to an ideal that illuminates what is often excellent in non-philosophers' moral thought, if we imagine them as being always and necessarily engaged in attempts at conceptual analysis.

V. WHAT IS THE RELATION OF CONCEPTUAL REASONING TO THE UNITY-OF-VALUE THESIS?

My third question about Dworkin's thesis that moral reasoning is conceptual interpretation involves its relationship to his additional thesis asserting what he calls "the unity of value."[41] Does treating moral reasoning as conceptual reasoning guarantee the conclusion that the truth about what is right and good and valuable is "one big thing,"[42] as he puts it? Or, alternatively, is the conclusion that value is unitary one that must be "earned" through a series of case-by-case conceptual inquiries in which all genuine values are actually tested for consistency and mutual supportiveness and through which the thesis of "the unity of value" emerges as a demonstrated conclusion?

Although these are substantive questions, they are obviously bound up with definitional ones. In an interesting and important passage in *Justice for Hedgehogs*, Dworkin equates the unity-of-value thesis with "value holism," which he defines as "the hedgehog's faith that all true values form an interlocking network, that each of our convictions about what is good or right or beautiful plays some role in supporting each of our other convictions in each

---

[40] *See generally, e.g.*, RONALD DWORKIN, SOVEREIGN VIRTUE (2000).

[41] *See* DWORKIN, *supra* note 1 (manuscript at 7).

[42] *Id.*

of those domains of value."[43]   If our values did not form an interlocking network, we could not practice conceptual interpretation in the sense of searching for conceptual equilibrium.   Nevertheless, the premise that our values are interlocking and that they play some role in supporting one another leaves open many further questions about what exactly the unity-of-value thesis is – and, in particular, about whether it denies the existence of one or another kind of possible conflicts among values.

To begin to explore some of the ways in which values might or might not (be thought to) conflict, it may be useful to consider my Case 3,[44] in which a friend presses me on the question whether I think his bad book is good. On the surface, the case appears to me to involve a conflict between the values of honesty and kindness: I am inclined to believe that honesty counsels one response, kindness another. Does the unity-of-value thesis categorically deny that this is, or could be, the case?

On the strongest possible interpretation, the unity-of-value thesis would deny that values can ever conflict in any way, even in application. On this interpretation, if honesty and kindness both are genuine values, then it is, or ought to be, possible for me to arrive at interpretations of both such that I need never choose which takes priority over the other under particular conditions. Perhaps, when the concept of kindness is correctly understood, it is really not unkind to tell my friend the truth. Or perhaps it is not really dishonest to please the friend when no harm will come of a flattering report and giving a blunter answer would be unkind. Either way, if the unity-of-value thesis denies the possibility of value conflict in application, then a harmonious conceptual alignment that precludes such conflicts should be available.

Prior to *Justice for Hedgehogs*, Dworkin had argued, as he continues to argue, that there is no conflict among the values of liberty, equality, and democracy – even in application – when these concepts are correctly understood.[45] Based on his prior analyses of these three concepts, I assumed, upon first encountering his assertion of "the unity of value" at the beginning of *Justice for Hedgehogs*, that he held the very strong view that there can be no conflict among genuine values even in application.[46]

After reading *Justice for Hedgehogs* in its entirety, however, I am not certain that Dworkin embraces the strong view that values can never conflict even in application. I am also extremely doubtful that this thesis could be sustained even on the assumption that moral reasoning is conceptual reasoning. With respect to Dworkin's view, there is a potentially telling passage in *Justice for Hedgehogs* in which he writes that it might be "the right thing" for

---

[43] *Id.* (manuscript at 76-77).

[44] *See supra* Part II.

[45] *See, e.g.*, DWORKIN, *supra* note 40, at 121, 182, 184-90.

[46] *See* DWORKIN, *supra* note 1 (manuscript at 130) (distinguishing between "total freedom," which is not valuable as an end in itself, and negative liberty, which, when rightly understood, is).

someone to "act cruelly on some occasion."[47] This certainly sounds like a case of value conflict, at least in application. Doing the right thing is an important value. Non-cruelty or the avoidance of cruelty seems to be a genuine value as well. Other examples of values conflicting in application, even if they do not conflict in principle, come from the realm of politics. Public safety is a value, but rights represent values too. Accordingly, we sometimes need to sacrifice some degree of public safety in order to protect people's rights.

As I suggested above in connection with Case 3, an adherent of the strongest version of the unity-of-value thesis might push back in all of these cases. Perhaps rather than thinking that it is "right" to be "cruel," we should revise our understanding of cruelty so that an act that was ultimately right is not and could not really be cruel after all. Perhaps public safety is a genuine value only insofar as it can be achieved consistently with respect for rights. And, as I noted above, perhaps we should consider whether there is any real unkindness in responding to a friend's query by telling him that his book is bad – or no real dishonesty in flattering him.

As I try to work through a series of specific examples, however, I become convinced that the thesis that genuine values can never conflict in application is untenable. I cannot help believing both that if I tell my friend the truth about his book, I will be unkind, and that if I do what kindness requires, I will speak dishonestly.[48] I thus believe that Dworkin ought not to embrace the claim that genuine values can never conflict in application; certainly none of his arguments persuades me of it. Moreover, given Dworkin's observation that it might be right to be cruel sometimes, I am, as I said, not at all confident that even he wishes to deny the possibility that values might sometimes conflict in application when he asserts the unity-of-value thesis in *Justice for Hedgehogs*.[49]

---

[47] *Id.* (manuscript at 116).

[48] Although I am unsure whether Dworkin would agree or disagree with this particular judgment, he affirms very clearly that if I cannot help believing something "even after the most sustained reflection," then I must affirm my "undefeated conviction" and find an account of why I am justified in believing as I do:

[I]n the end raw, undefeated conviction must play a decisive role in any honest search for an equilibrium epistemology; there may be propositions that we find we cannot but believe even after the most sustained reflection. Then we must not pretend not to believe them but must instead struggle to find an account of why we are justified, in spite of the difficulties, in believing what we do.

Id. (manuscript at 40).

[49] Some of Dworkin's other remarks reinforce my uncertainty. In one passage he writes:

We must find convictions that we believe and that do fit. This is an interpretive process, as I said, because it seeks to understand each part and strand of value in the light of other strands and parts. Almost no one can manage this fully, and there is no guarantee that we will all be able, even together, to manage it even very well.

Id. (manuscript at 77); *see also id.* (manuscript at 240) ("A good government is democratic, just, and efficient, but these are not the same qualities and it is sometimes important to ask,

If an interpretation of the unity-of-value thesis that denies the possibility of any value conflicts whatsoever is too strong to be maintained, other interpretations need to be considered. On a relatively weak "no-conflict-in-principle" interpretation, the unity-of-value thesis might insist only that all genuine moral values exist in a relationship of sufficient harmony so that there is no necessary conflict among the explanations of why each of them, separately, is indeed valuable. If this version of the thesis holds, then it must be possible to give an interpretive account of what honesty is, and why it is valuable, that does not conflict with an interpretive account of kindness and why kindness is valuable. Nevertheless, the fact that honesty and kindness do not conflict in principle – in the sense that there may be no necessary incompatibility in their respective definitions or in the accounts of why each is valuable – does not guarantee that there will be no conflicts in application. On this weaker interpretation, it would be consistent with the unity-of-value thesis to acknowledge, in Case 3, that it may be impossible for me to answer my friend in a way that is both honest and kind.

Although this version of the unity-of-value thesis might be appealing to some, it seems too weak for Dworkin, for it leaves open the possibility that when values conflict in application, there might be no right answer to the question what people ought to do or believe. Dworkin leaves no doubt in *Justice for Hedgehogs* about his continuing embrace of the "one right answer" thesis.

A third possible interpretation of the unity-of-value thesis would conjoin the claim that there are no conflicts in principle among genuine moral values with the one-right-answer thesis. So interpreted, the unity-of-value thesis links to the thesis that moral reasoning is conceptual interpretation in the following way: when we search for conceptual equilibrium within our web of beliefs, there will always be one right answer to the ultimate question of what we ought to do or believe (even if, in some cases, the right answer might be that two courses of action are equally good or equally bad).[50] Thus, for example, if it is right to be cruel, then the right answer to what one ought to do is that one ought to be cruel. Similarly, even if one cannot simultaneously be both honest and kind in Case 3, there may be a right answer to the question whether one ought to be honest or instead ought to be kind on a particular occasion. To express the point in terms that I employed earlier, we might think that we should always be able to achieve a reflective equilibrium among our moral convictions about what is ultimately right or wrong for us to do, even if we

---

for instance, whether some constitutional arrangement that is likely to make a community more just or efficient must nevertheless be resisted because it is undemocratic.").

[50] *See id.* (manuscript at 25) ("Perhaps it is neither true nor false that abortion is wicked or that the American Constitution condemns all racial preference or that Beethoven was a greater creative artist than Picasso. But if so, this is not because there can be no right answer to such questions for reasons prior or external to value, but because that *is* the right answer . . . as a matter of sound moral or legal or aesthetic judgment.").

cannot always achieve a *conceptual* equilibrium in which genuine values never conflict in application. In other words, the "one big thing" the hedgehog knows is that our values are sufficiently commensurable and sufficiently integrated into a mutually reinforcing web of belief for all moral questions to have one right answer.

What we might now call the one-right-answer interpretation of the unity-of-value thesis is strong enough to be controversial. Furthermore, it may well be entailed or presupposed by Dworkin's thesis that moral reasoning is conceptual interpretation, at least when that thesis is fleshed out by his general theory of interpretation.

There is a further reason to expect that this interpretation of the unity-of-value thesis – which insists that our web of beliefs is sufficiently united to yield one right answer to moral questions, but which leaves open the question whether values within the web might otherwise conflict in application – would be congenial to Dworkin. Consistent with his large, recurring claim in *Justice for Hedgehogs* that there cannot be any interestingly Archimedean, outside-the-fray perspective on ultimate moral questions,[51] this interpretation locates debates about whether genuine values ever conflict in application squarely within the domain of moral argument.

There are also reasons, however, to believe that Dworkin might resist an interpretation of his unity-of-value thesis that reduces it to the conjunction of the claim that genuine values do not conflict in principle and the one-right-answer thesis. To begin with, at least some of the "foxes" that Dworkin takes as his intellectual adversaries – those who insist that truth within or about the domain of value is better characterized as many little things than as one big thing – might be prepared to accept this version of the "unity-of-value" thesis. To echo a formulation that I used earlier, a fox could think that although values sometimes conflict in application, so that there can be no choice among them without occasion for moral regret, there may nevertheless be one right answer to the question which value should prevail in particular contexts.[52]

An interpretation of the unity-of-value thesis as accepting the possibility that values might conflict in application, but as insisting that there is always one right answer to the question of how conflicts should be resolved, would also have what, for Dworkin, would be a very odd consequence. The hedgehog claims to know one big thing: that our values are sufficiently linked, mutually supporting, and commensurable so that there is always one right answer to moral questions. But if the unity-of-value thesis acknowledges that values can conflict in application, then it, by itself, tells us nothing about when conflict can or will occur or about how conflicts should be resolved. On this interpretation, moral wisdom would thus require conjoining the insight of the hedgehog with the insights of the fox. Even if the hedgehog is right that all genuine values cohere in principle and that there is always one right answer to

---

[51] *See id.* (manuscript at 37).

[52] *See supra* note 5 and accompanying text.

the question of what we ought to do, we would need the fox to tell us when values do and do not conflict in application and what we should do when conflict occurs. This conclusion strikes me as entirely plausible, but I am far from confident that Dworkin – who sets out to argue the hedgehog's case against the fox's – would welcome it.

In any event, my question, which began as one about the relationship of the claim that moral reasoning is conceptual interpretation to the unity-of-value thesis, has evolved into a puzzlement about what exactly the unity-of-value thesis is and how strong Dworkin intends it to be.