JOURNAL *of* ETHICS
& SOCIAL PHILOSOPHY

VOLUME XVIII · NUMBER 3
October 2024

ARTICLES

DISCUSSION

JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY
http://www.jesp.org
ISSN 1559-3061

The *Journal of Ethics and Social Philosophy* (*JESP*) is a peer-reviewed online journal in moral, social, political, and legal philosophy. The journal is founded on the principle of publisher-funded open access. There are no publication fees for authors, and public access to articles is free of charge. Articles are typically published under the CREATIVE COMMONS ATTRIBUTION-NONCOMMERCIAL-NODERIVATIVES 4.0 license, though authors can request a different Creative Commons license if one is required for funding purposes. Funding for the journal has been made possible through the generous commitment of the Division of Arts and Humanities at New York University Abu Dhabi.

JESP aspires to be the leading venue for the best new work in the fields that it covers, and it is governed by a correspondingly high editorial standard. The journal welcomes submissions of articles in any of these and related fields of research. The journal is interested in work in the history of ethics that bears directly on topics of contemporary interest, but does not consider articles of purely historical interest. It is the view of the editors that the journal's high standard does not preclude publishing work that is critical in nature, provided that it is constructive, well argued, current, and of sufficiently general interest. *JESP* also endorses and abides by the Barcelona Principles for a Global Inclusive Philosophy, which seek to address the structural inequality between native and nonnative English speakers in academic philosophy.

JESP publishes articles, discussion notes, and occasional symposia. Articles normally do not exceed 12,000 words (including notes and references). *JESP* sometimes publishes longer articles, but submissions over 12,000 words are evaluated according to a proportionally higher standard. Discussion notes, which need not engage with work that was published in *JESP*, should not exceed 3,000 words (including notes and references). *JESP* does not publish book reviews.

Papers are published in PDF format at https://www.jesp.org. All published papers receive a permanent DOI and are archived both internally and externally.

Journal of Ethics and Social Philosophy
vol. 28, no. 3 · october 2024

https://doi.org/10.26556/jesp.v28i3.2838

HOW PRACTICES MAKE PRINCIPLES AND HOW PRINCIPLES MAKE RULES

Mitchell N. Berman

What gives law its content? If *q* is a legal norm, what makes that so? Many contemporary legal philosophers believe that answering this question is the discipline's most urgent task. Mark Greenberg, a leading antipositivist, maintains that dispute over "the determinants of the content of the law" makes out "a central—perhaps *the* central—debate in the philosophy of law."[1] Scott Shapiro, a leading positivist, agrees, emphasizing that we cannot resolve first-order legal questions unless we first "know which facts ultimately determine the content of all law."[2] The view is widespread.[3] This article offers a new general account of the determination of legal content. I call this theory "principled positivism."

The account is positivist because it maintains that legal norms are necessarily determined by the actions and mental states of persons (or by *facts about* such actions and mental states) and by moral notions only contingently, if at all. However, and in marked contrast to the reigning positivist theory that is associated with H. L. A. Hart, my account gives the weighted, contributory norms that the arch antipositivist Ronald Dworkin called "principles" a central role in the determination of legal "rules." In currently favored metaphysical terminology, legal practices fully ground legal principles, and legal principles partially ground legal rules.

This paper motivates, explicates, illustrates, and defends principled positivism. Section 1 sets the table. It briefly sketches a Hartian theory of legal content and then presents what I consider the two most formidable challenges to it, both pressed by Dworkin, positivism's fiercest critic.[4] The first challenge was raised in Dworkin's first attack against Hart's theory, "The Model of Rules I"

[1] Greenberg, "How Facts Make Law," 157 (reprinted and revised in Hershovitz, *Exploring Law's Empire*). As should be apparent, I derive my title from Greenberg's.

[2] Shapiro, *Legality*, 29 (emphases omitted).

[3] See, e.g., Plunkett and Shapiro, "Law, Morality, and Everything Else," 56; Stavropoulos, "The Debate that Never Was," 2090; Toh, "Jurisprudential Theories and First-Order Legal Judgments"; and Baude and Sachs, "Grounding Originalism," 1460.

[4] I clarify in what sense the theory I will be critiquing is "Hartian" in section 1.1 below.

("TMR I").[5] This objection, which I call the *challenge from principles*, maintains that Hartian positivism has difficulty accounting for the contributory, weighty, and conflicting norms that Dworkin called legal "principles." Exactly why, on Dworkin's analysis, Hart's account cannot accommodate principles is largely misunderstood. Drawing on a predecessor article, I explain that the crux of the challenge is not that Hart's account cannot deliver legal *principles* but that, insofar as it can, it cannot deliver legal *rules* due to the way that principles contribute to rules.[6]

Dworkin developed his second challenge in work that followed TMR I, most insistently when speaking as an American constitutional theorist. It maintains that because of pervasive disagreements among US justices and judges about matters of "constitutional interpretation," vastly fewer putative legal norms are "valid," or "exist," than sophisticated observers and participants believe on reflection there to be. I call this objection the *too-little-law challenge*. It is kin to a much better-known objection, the *challenge from theoretical disagreements*, that Hart's theory more easily rebuts.

Section 2 introduces an alternative to the Hartian theory of legal content designed to meet the challenges from principles and of too little law. The two key moves are: first, to allow for the determination of nonfundamental (i.e., derivative) legal norms by a means that does not require Hartian "validation"; and second, to allow for the determination, or "grounding," of fundamental legal norms in practices that fall short of judicial consensus. In presenting an account that has these twin virtues, this section explains (1) how "legally fundamental" weighted norms can be grounded directly in the messy, conflictual human practices that characterize modern, vast, and decentralized legal systems, (2) how such principles can interact or combine by nonlexical, aggregative means—that is, means not properly classified as "validation"—to determine the legal status of token acts and events, and (3) how the "decisive" and general legal norms customarily called "rules" fit into the picture.

Section 3 puts my account to work, showing how it meets Dworkin's challenges. It does so with the aid of two concrete disputes from American statutory and constitutional law. The first is the "snail darter case" that Dworkin discusses at length in *Law's Empire*.[7] The second is the constitutional right to recognition of same-sex marriage that was announced in *Obergefell v. Hodges*.[8]

[5] Dworkin, "The Model of Rules," reprinted and revised as "The Model of Rules I" in Dworkin, *Taking Rights Seriously*. Subsequent citations will be to the book.

[6] Berman, "Dworkin versus Hart Revisited."

[7] *Tennessee Valley Authority (TVA) v. Hill*, 437 U.S. 153 (1978).

[8] *Obergefell v. Hodges*, 576 U.S. 644 (2015).

* * * * *

This article aspires to contribute to general jurisprudence, not (directly) to American constitutional law or theory. But as section 3 makes clear, the disciplines are not crisply separable. That was one of Dworkin's core insights, memorably pronouncing jurisprudence "the general part of adjudication, silent prologue to any decision at law."[9] Insofar as the jurisprudential intervention this article undertakes is successful, implications for American legal interpretive theory are unavoidable. This one article—already near law-review length—does not draw forth and defend those implications. But readers whose interest in jurisprudence derives largely from its character as prologue will naturally wonder at what might follow. What follows is a positivist, pluralist, and dynamic theory of American constitutional law that I call "organic pluralism." Organic pluralism is a competitor to all forms of originalism. Principled positivism is its jurisprudential backbone.

1. HARTIAN POSITIVISM AND TWO DWORKINIAN CHALLENGES

This article could possibly start where section 2 does—with a presentation of the account I call principled positivism. But that account emerges within a tradition. And if it boasts any distinctive virtues, they can be grasped only with an understanding of the theoretical dialectic. This section supplies the necessary context.

Section 1.1 sketches the Hartian theory of legal content, emphasizing the ultimate rule of recognition's character as a social practice that grounds "fundamental" legal norms—the "ultimate criteria of validity"—and the role of those criteria in "validating" the legal norms that are "derivative." The remainder of the section identifies the most daunting obstacles that account faces. Section 1.2 introduces the most forceful challenge pressed by the early Dworkin: the "challenge from principles" lodged in TMR I. Common wisdom holds that Hartians have successfully rebutted that challenge.[10] I will argue that such optimism is based on a misunderstanding of Dworkin's argument, and that the challenge remains unrefuted.[11] Section 1.3 introduces the challenge from theoretical disagreements from *Law's Empire.* Here I argue, against Shapiro, that the challenge is easily met. Section 1.4 turns to a rarely discussed cousin to the challenge from

[9] Dworkin, *Law's Empire*, 90.

[10] See, e.g., Shapiro, "The Hart–Dworkin Debate," 35.

[11] This is the main work of Berman, "Dworkin versus Hart Revisited," a prequel to the current article. Sections 1.1 and 1.2 here summarize arguments developed at greater length there.

theoretical disagreements, what I call the *too-little-law challenge*. I argue that it is the later Dworkin's most formidable objection. This section's takeaway is that if positivists are to offer a complete theory of legal content, they must still engage with and defeat the challenge from principles and the challenge of too little law.[12]

1.1. From Socio-Normative Positivism to Hartian Legal Positivism

Before we get to legal norms, let us discuss social norms. At the time of writing, norms in most Western cultures direct that one should greet a new acquaintance by shaking hands, while norms in many Asian cultures direct that one should bow. Students at many schools observe a norm not to volunteer to answer instructors' questions. Let us say that the "content" of a norm is what the norm directs or provides. A norm's content is thus analogous to a word's meaning; it is what differentiates one norm (word) from another. Common theoretical wisdom about social norms includes three elements:

1. *Minimal realism* (the "metaphysically unambitious" thesis that "there really are ways that things might be" with respect to social norms and their contents, "and that our thoughts and sentences do sometimes correctly represent that reality");
2. *Thin normativity* (the view that these norms exhibit or exert a type or grade of normativity different in character or stringency from moral norms as conceived by traditional or "robust" moral realists and are not "truly" or "unconditionally" binding); and
3. *Positivism* (the idea that these norms are what they are and have the contents they do in virtue of certain behaviors and mental states (or by *facts about* those behaviors and mental states) undertaken by some members of the social groups to which the norms apply).[13]

Putting these elements together: (1) social norms in Mali really do direct that prepubescent girls should be subjected to genital mutilation; (3) this norm exists in virtue of certain behaviors and attitudes prevalent in Malian society; and (2) that Malian norms direct that parents should subject their daughters to genital mutilation does not entail that they *really* (robustly, unconditionally) should do so.

[12] This takeaway is important for any contemporary scholar interested in explaining legal content. It need not amount to a criticism of Hart, though, for providing an account of legal content was not his primary goal, if one at all.

[13] For minimal realism, see Van Roojen, *Metaethics*, 9–14. Thin normativity is the type of normativity that attaches to rules of etiquette and rules of a club, as famously explored in Foot, "Morality as a System of Hypothetical Imperatives." For elaboration, see, e.g., Berman, "Of Law and Other Artificial Normative Systems," 143–44; Finlay, "Defining Normativity"; and Wodak, "What Does 'Legal Obligation' Mean?"

There are different ways to make sense of the (minimal) reality of social norms and therefore of the mode by which social facts or practices determine norms' contents. But philosophers are increasingly treating norms as elements of social ontology to be explained metaphysically. And those who do are increasingly drawn to the language of "grounding," where grounding is a relationship of metaphysical determination by which more fundamental facts or entities explain, noncausally, less fundamental ones.[14] For example, physical, neurochemical states of the brain ground mental phenomena such as beliefs, intentions, and pain; microphysical properties such as molecular structure ground macrophysical properties such as hardness and conductivity. I will adopt this vocabulary for explaining norms, both social and legal, without further defense. That is, I will gloss the third element in the standard view of social norms—positivism—by saying that social norms are "grounded in" social practices.

Figure 1 depicts the determination of social norms by "social practices," by which I mean to embrace a potentially broad range of behaviors and accompanying mental states, such as believing and stating that the standard a norm captures is normative, using it to guide and justify one's own conduct, criticizing oneself and others for deviance, and so on. Practices are "social" when engaged in by (significant portions of) some identifiable subset of society; they need not be found through all of society. I designate the grounding relationship simply "G1," leaving its details entirely open.[15]

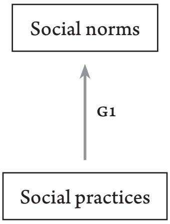


FIGURE 1 Social Norms Model

[14] I aim to remain as noncommittal as possible on controversial issues in metaphysical grounding. That said, I will generally take the grounding relata to be entities such as speech acts, practices, and artificial norms—not *facts about* speech acts, practices, or artificial norms. Compare, e.g., Rosen, "Metaphysical Dependence" (facts) with Schaffer, "On What Grounds What" (not facts). But I am not doctrinaire about this. When it facilitates exposition, I will sometimes speak about the grounding facts. I trust that nothing of substance in my argument depends on adopting one or another position on this intramural controversy.

[15] See, e.g., Brennan et al., *Explaining Norms*: "Norms ... are clusters of normative attitudes plus knowledge of those attitudes" (35). See also Bicchieri, *The Grammar of Society*: "Norms are supported by and in some sense consist of a cluster of self-fulfilling expectations" (ix).

For a legal positivist, complex institutionalized normative systems including law exhibit these same three properties. EU securities regulations, offside rules in soccer, Jewish dietary laws—they are all minimally realist, only thinly normative, and determined by (many would say "grounded in") social practices or facts.[16]

There is, however, one critical difference. All social norms are grounded *directly* in social facts: *q* is not a social norm of community *S* if not the object of some supportive practices.[17] Things are different in complex systems: at least some norms of such systems are *not* taken up by participants and might be entirely unknown to them. As American constitutional theorists William Baude and Stephen Sachs note, "we can be surprised by, mistaken about, or disobedient toward the law without it ceasing to *be* law."[18] So if legal norms are grounded in social facts, the mechanism by which facts determine law must be *indirect*, at least sometimes. The task for positivist theories of legal content, then, would be to explicate the indirect determination relationship that yields legal norms consistent with a scientific picture of the world.[19]

A natural thought is that if a positivist model of complex normative systems including law is to prove viable, it would likely involve two levels of determination, whereas the generic positivist model of social norms recognizes only one. On this positivist model of law, social practices ground fundamental legal norms, by G1 or a close analogue; and fundamental legal norms, together with whatever facts, practices, or phenomena the fundamental legal norms "point to" or otherwise make legally relevant, determine derivative legal norms, by a mechanism or relation D2 (figure 2). The fundamental legal norms that *are* directly grounded in social practices function as "normative intermediaries" in the determination of legal norms that are *not* directly grounded in such practices. For example,

[16] For the view that legal positivists should (and Hart did) accept minimal realism about legal norms, see Kramer, *H. L. A. Hart*, 30–31, 192–93. For the view that "positivism is best interpreted as a grounding thesis," see Chilovi and Pavlakos, "The Explanatory Demands of Grounding in Law," 900 (citing Tomasz Gizbert-Studnicki, David Plunkett, Gideon Rosen, and Nicos Stravropoulos as other proponents).

[17] As Cristina Bicchieri cautions, this does not mean that a social norm must be *heeded* to exist. Even if all members of a normative community *S* secretly flout *q*, *q* can still be a social norm of *S* so long as the members engage in such norm-supportive behaviors as urging others to comply with *q* or criticizing others (or themselves) for noncompliance. See Bicchieri, *The Grammar of Society*, 11.

[18] Baude and Sachs, "Grounding Originalism," 1473. See also Bix, "Global Error and Legal Truth"; and Sachs, "The 'Constitution in Exile' as a Problem for Legal Theory."

[19] See Plunkett and Shapiro, "Law, Morality, and Everything Else," arguing that jurisprudence is a branch of metanormative inquiry and that metanormative theory in general is concerned with explaining "how thought, talk, and reality that involve [normative notions] fit into reality" (49).

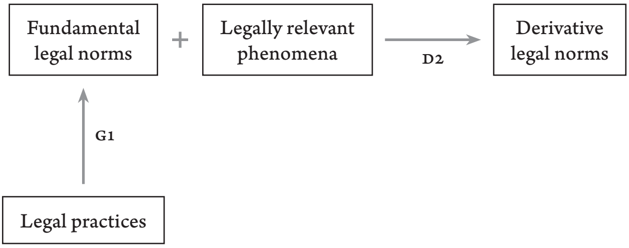


FIGURE 2 Generic Two-Level Legal Positivism

suppose that a fundamental legal norm, *F*, of legal system *S* provides that *r* is a legal rule of *S* if *r* corresponds to a specified type of communicative content of a specified type of text.[20] And suppose that *T* is a text of the specified type, and its relevant communicative content is *q*. Then the fact that a legal rule of *S* corresponds to *q* is jointly determined by *F* and the communicative content of *T*.[21]

The account that Hart presented in his masterwork, *The Concept of Law*, is easily understood as one way to put flesh on this skeletal legal positivist model. But I have learned that the closer "Hart" and "grounding" appear in the same sentence, the more important it becomes to emphasize just what I am and am not claiming.

Thus far, I have (a) said that a full-bore positivism about law should include a theory of legal content, (b) endorsed a metaphysical rendering of that project,

[20] Notice that *F* in this example functions more as a constitutive rule than as a regulative rule. See generally Searle, *Speech Acts*, 33–34. It serves to make something the case, not to require, direct, or prohibit. Persons who believe that every norm is an *ought* and thus that a notion or operator must purport to have action-guiding character to count as a "norm" (see, e.g., Himma, "Understanding the Relationship between the U.S. Constitution and the Conventional Rule of Recognition," 98) will resist my characterization of *F* as a legal norm. My linguistic intuitions about "norms" are more expansive and embrace elements or concepts within the normative domain or that bear specified relationships to norms that have a directive or deontic character. But this is a semantic dispute that need not detain us. If you would withhold the term "norm" from an abstract entity whose function is to metaphysically determine the content of action-guiding entities but not to guide action directly, you might call *F* and its kin "shnorms" or "auxiliaries to norms." My substantive points remain unaffected.

[21] Philosophers debate whether grounding is a single type of metaphysical determination, a group of related types, or just a comprehensive label for varied kinds of already recognized determination relationships. See generally Berker, "The Unity of Grounding." I am myself more persuaded that grounding is a genuine type of determination and one that obtains between practices and norms than I am that the determination of derivative legal norms by fundamental legal norms and the phenomena that they make relevant is also best conceived in terms of grounding. I signal the possibility of important differences in the two determination mechanisms by referring to the latter relationship as simply "determination"—denominated D2 rather than G2—and by representing D2 with a horizontal arrow rather than a vertical one, departing from the convention that represents grounding vertically.

and (c) embraced the grounding idiom for this metaphysical inquiry. Plainly, Hart did not speak in terms of "grounding"; it was not part of the then-prevailing philosophical lexicon. More importantly, vocabulary aside, it is contestable whether Hart's overall theory includes an account of legal content at all and, if so, whether it is one that can be translated into grounding terms. Perhaps he was offering only a theory of how preexisting extralegal norms get validated as legal.[22] Perhaps he was offering a theory of the validation of legal sources alone, not also of the norms that sources partially determine.[23] Perhaps he was offering a noncognitivist account of legal thought and talk and would reject minimal realism about legal content.[24]

But whatever Hart was up to, anyone who accepts minimal realism about legal content should see the need for a theory that explains how that content comes to be and that has the resources to adjudicate disputes about whether the content is *this* rather than *that*. Furthermore, for anyone who seeks such a theory and has positivist sensibilities, the search most naturally starts with Hart. And if we do look toward Hart with the aim to discern or develop a theory of legal content, a possible view emerges clearly enough. Roughly: it is the nature of a legal system that legal norms have the legal contents that they do in virtue of being validated by a set of (usually) sufficient conditions or "criteria" that are grounded in the ultimate "rule of recognition," a convergent practice among officials (chiefly judges) of identifying legal norms that the officials follow with the critical reflective attitude that Hart dubs the "internal point of view."[25] I will call this view the "Hartian theory of legal content" without worrying further about the extent to which Hart himself held it. I follow other scholars in speaking this way.[26] As the remainder of this section argues, Dworkin's criticisms of

[22] Cf. Gardner, "Legal Positivism": "Legal positivism is not a whole theory of law's nature, after all. It is a thesis about legal validity." (33).

[23] See, e.g., Waldron, "Who Needs Rules of Recognition?" 336.

[24] See Toh, "Hart's Expressivism and His Benthamite Project."

[25] See generally Hart, *The Concept of Law*, 100–17. Grant Lamond spins Hart's account in a metaphysical direction when maintaining that "the language of 'recognition' and 'identification' is not entirely apt: what the rule of recognition does is to *constitute* the rules as rules of the system, that is, it *makes* them rules of the system" ("The Rule of Recognition and the Foundations of a Legal System," 114). Yet the language of "recognition" and "identification" seems very apt insofar as what are being validated are preexisting norms external to this legal system. In such cases, the facts about legal practice that ground fundamental legal norms would not determine the norms' contents; those contents would be determined by whatever extralegal grounds ground the extralegal norms. (I take this thought from an anonymous referee.) But many norms in contemporary municipal legal systems are created by the legal system, not simply adopted from some other normative system. For them, "recognition" and "identification" do seem unfit, and "constitution," "determination," or "grounding" are better.

[26] See, e.g., Chilovi and Pavlakos, "Law-Determination as Grounding," who sketch "a ground-theoretic interpretation" of "Hartian positivism" according to which "rules of recognition

Hart are comfortably understood as targeting something very much like this account.

On this reading, the Hartian theory of legal content is a specification of generic two-level legal positivism in three respects. First, it replaces the vague generic reference to "legal practices" with Hart's signature theoretical innovations, the internal point of view and ultimate rule of recognition. Second, it conceptualizes the "fundamental legal norms" that are grounded in practice as "ultimate criteria of validity."[27] Third, and working hand in glove with the second, it posits that the determination mechanism is "validation."[28] (See figure 3.)

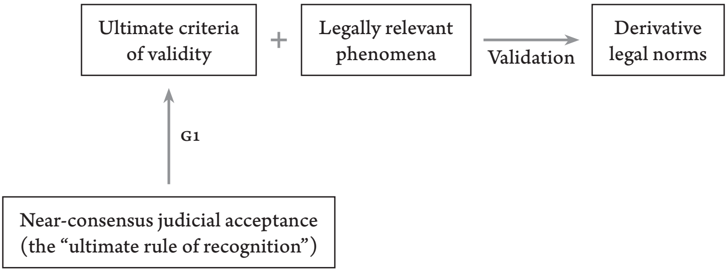

Ultimate criteria of validity
+
Legally relevant phenomena
Validation
Derivative legal norms
G1
Near-consensus judicial acceptance (the "ultimate rule of recognition")

FIGURE 3 Hartian Legal Postivism: First Pass

1.2. *A Problem for Validation: The Challenge from Principles*

Many legal theorists today accept the foregoing picture, at least in broad strokes. Ronald Dworkin did not. His target in the paper that would come to be known as the "The Model of Rules I" was legal positivism. His strategy was

play a double role" in that "they count as partial grounds of law" and "enable certain facts to be further grounds, and determine the way in which these facts contribute to legal content" (71–74). See also Greenberg, "What Makes a Method of Legal Interpretation Correct?": "Jurisprudential theories like those of Hart and Dworkin offer accounts of how the content of the law is determined at the fundamental level.... On Hart's theory, the content of the law is determined at the fundamental level by convergent practices of judges and other officials" (112–13).

[27] Scholars frequently use the term "rule of recognition" (often omitting the modifier "ultimate") to refer both to the social rule among judges of accepting criteria of legal validity and to the criteria themselves. Hart himself did not adhere to the distinction consistently. See, e.g., Hart, *Essays in Jurisprudence and Philosophy*, agreeing with Lon Fuller that the ultimate rule of recognition could be deemed "a political fact" but insisting that "[t]he propriety of this ... description [does] not exclude the classification of this phenomenon as an ultimate legal rule" (359). Still, I am persuaded that clarity is enhanced by keeping the notions separate, as I attempt to do here. (I am grateful to Brian Leiter for doing the persuading.)

[28] Chilovi and Pavlakos, "Law-Determination as Grounding" offers a similar analysis of Hart's account in terms of grounding. I explain the modest differences between our accounts in Berman, "Dworkin versus Hart Revisited," 560n41.

to demonstrate that positivism's most fully realized version—Hart's—could not make sense of legal principles as a logically distinct type of norm.

On this much, all agree—but on little else. It is not merely that commentators disagree about whether the *challenge from principles* (as I term it) succeeds. As is often the case when it comes to Dworkin exegesis, they do not all agree on exactly how the challenge even runs. I unpack Dworkin's argument at length elsewhere.[29] This section summarizes.

Standard understanding of Dworkin's argument starts with his proposed distinction between rules and principles. "Rules," Dworkin explains, "are applicable in an all-or-nothing fashion. If the facts a rule stipulates are given, then either the rule is valid, in which case the answer it supplies must be accepted, or it is not, in which case it contributes nothing to the decision."[30] Principles, in contrast, bear on a decision with variable "weight or importance" and are not decisive. Principles "incline a decision one way, though not conclusively, and they survive intact when they do not prevail."[31]

The problem for Hartian positivism, according to this challenge, is that it is a "model of rules" alone, not of principles as well. This is because Hart allows for legal norms to arise in only two ways: by being validated in accordance with the criteria of validity or by being the subject of convergent acceptance by officials, centrally judges. But, says Dworkin, principles cannot arise in either of these two ways. Principles cannot be determined by validation because they do not depend upon specifiable sufficient conditions; they cannot be validated by any "test that all (and only) the principles that do count as law meet."[32] Nor can they arise by acceptance because that would reduce the scope and significance of the rule of recognition; it "would very sharply reduce that area of the law over which [Hart's] master rule held any dominion."[33] Therefore, Hart's theory cannot accommodate legal principles.

As early critics of the essay showed, this argument is infirm in several respects.[34] While some flaws might be massaged away, many readers were

[29] Berman, "Dworkin versus Hart Revisited."

[30] Dworkin, *Taking Rights Seriously*, 24.

[31] Dworkin, *Taking Rights Seriously*, 35.

[32] Dworkin, *Taking Rights Seriously*, 40.

[33] Dworkin, *Taking Rights Seriously*, 43.

[34] For one thing, Dworkin offered two stabs at the distinction between rules and principles, not one. In addition to distinguishing rules and principles on the basis of their logical character, Dworkin also offered a substantive (or "normative") difference: principles concern "justice or fairness or some other dimension of morality" (Dworkin, *Law's Empire*, 22). However, the scholarly consensus is that "Dworkin's two accounts of principles do not mesh" (Lyons, "Principles, Positivism, and Legal Theory," 423) and that, if there is a distinction

wholly unpersuaded by what they took to be Dworkin's core thesis—namely, that legal principles cannot "come into being" either (directly) by being accepted or (indirectly) be being validated.[35] To the contrary, commentators thought it apparent that they can arise in *both* ways.

Take validation first.[36] Suppose the criteria of validity specified by the ultimate rule of recognition provide that [q is a legal norm if text T says q], and suppose further that what T says, among other things, is that "states should be paid special regard." It is not at all clear why that conjunction of facts would not validate some legal principle of federalism, the contours of which would be shaped in common-law fashion. Next take acceptance. Given that Hart allows that customary law can be law in virtue of being accepted, there is no obvious bar in Hart's theory to principles being accepted too.[37] Figure 4 represents the Hartian model as tweaked or clarified to respond to Dworkin's challenge: derivative legal principles can be validated by the ultimate criteria of validity; and just like those ultimate criteria, fundamental legal principles can also be directly grounded in the practices that Hart calls acceptance.

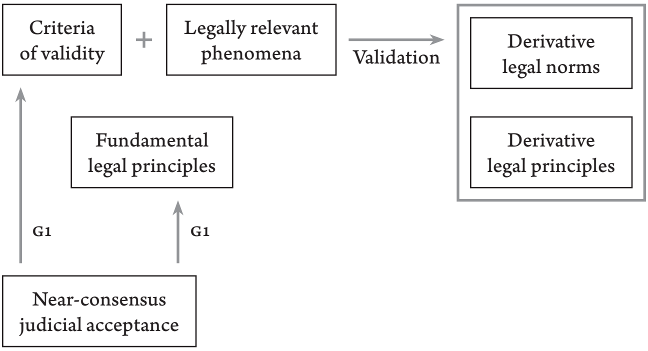

Criteria of validity
+
Legally relevant phenomena
Validation
Derivative legal norms
Derivative legal principles
Fundamental legal principles
G1
G1
Near-consensus judicial acceptance

FIGURE 4 Hartian Legal Positivism: Response to Dworkin

So far so bad for Dworkin's challenge, it seems. And yet even though Dworkin failed to fully corral his quarry, many theorists think that he was on the right track.[38] If so, the task is to make clearer what he was up to.

here, it resides in the vicinity of Dworkin's "logical" difference. For another, it appears probable that rules *can* conflict and have variable weight or importance (Soper, "Legal Theory and the Obligation of a Judge," 479–84; and Raz, "Legal Principles and the Limits of Law").

[35] Dworkin, *Taking Rights Seriously*, 20.

[36] See, e.g., Lyons, "Principles, Positivism, and Legal Theory," 425; Ten, "The Soundest Theory of Law," 524; and Hart, *The Concept of Law*, 261, 264–65.

[37] Raz, "Legal Principles and the Limits of Law," 853.

[38] See, e.g., Smith, "Dworkin's Theory of Law": "While many positivists thought that [Dworkin] over-stated or misunderstood the difference between rules and principles, most

Although Dworkin highlights his claim that Hartian positivism cannot explain the *existence* of legal principles, the true force of his challenge, I have argued elsewhere, is that it cannot explain their *function* or *operation*. As figure 4 indicates, the Hartian account, as modified to meet the challenge from principles, represents rules and principles (both fundamental and derivative) as coexisting in parallel, more or less. In the words of the inclusive positivist David Lyons, "principles supplement rules."[39] But principles have a function, which is to contribute to rules, not (merely) to supplement them; their role is to help constitute or metaphysically determine the rules that are not themselves grounded in official acceptance. And they do so, Dworkin charges, in a manner that the rule of recognition cannot accommodate: "rules … owe their force at least in part to the authority of principles … and so not entirely to the master rule of recognition."[40] This, finally, is the central thrust of Dworkin's challenge. "What really kills the model of rules in Dworkin's theory," Timothy Endicott rightly observes, "is not the proposition that there are some legal standards ['principles'] not identifiable by reference to a rule of recognition, but the proposition that all legal standards [including 'rules'] depend on standards that are not identifiable by reference to a rule of recognition."[41]

Unfortunately, Dworkin does not spell out precisely *why* determination of derivative rules by principles cannot be governed by the ultimate rule of recognition. One rare scholar who understood that Dworkin was targeting rules, not just principles, confessed to finding Dworkin's argument "puzzling."[42] Here I will try to make the logic and force of the challenge plainer. I will first lay it out succinctly and then say a little in defense of each of the argument's premises.

accepted that there *is* a difference between these two types of norm" (268). See also Alexander and Kress, "Against Legal Principles," observing that the Dworkinian distinction between rules and principles reflects "an entire jurisprudential tradition, a tradition that has shaped not only academic thought on these matters but also how lawyers and judges think and operate" (745). See also Ávila, *Theory of Legal Principles*.

[39] Lyons, "Principles, Positivism, and Legal Theory," 421.

[40] Dworkin, *Taking Rights Seriously*, 43. This way of putting things assumes that principles form part of a theory of legal content and not only of a theory of adjudication. See below text accompanying note 83. Dworkin spoke in both registers while being notoriously cavalier about the difference. See also Dworkin, *Taking Rights Seriously*: "The rules governing adverse possession may even now be said to reflect the principle [that nobody may profit from his own wrong] … because these rules have a different shape than they would have had if the principle had not been given any weight in the decision at all" (77). And also: "Unless at least some principles are acknowledged to be binding upon judges, requiring them as a set to reach particular decisions, then no rules, or very few rules, can be said to be binding on them either" (37).

[41] Endicott, "Are There Any Rules?" 203–4 (emphasis omitted).

[42] Bayles, *Hart's Legal Philosophy*, 167.

We will see that Dworkin's surprising contention that the rule of recognition cannot make sense of legal *rules* all depends on a crucial but entirely implicit distinction between two kinds of determination relationship, two general ways that determinants map onto resultants, or that grounded facts are grounded in grounding facts.

P1. There are two kinds of determination relationship: "lexical" and "nonlexical."

P2. If the Hartian account of legal content is true, then ordinary (derivative) legal rules are ultimately validated by (criteria grounded in) the ultimate rule of recognition.

P3. Validation is a lexical mode of determination.

P4. Principles contribute to the determination of rules nonlexically.

C. Therefore, the Hartian account of legal content is not true.

P1 is perhaps the most important of Dworkin's premises but also by far the least well developed. Fortunately, the core idea is highly intuitive: some determination relationships centrally involve such notions and operations as "if . . . then," necessity, and sufficiency, while others revolve around different notions, prominently including "greater than/less than," contribution, and thresholds. This is a familiar if undertheorized distinction from outside jurisprudence. Start with the treatment of moral principles in moral philosophy. As Jonathan Dancy observes, "there seem to be two ways of . . . getting a determinate answer to the question of what to do" when the principles that contribute to a decision conflict. One way "is to rank our principles lexically"; the other is "to think of principles as having some sort of weight" and adding them up.[43] "These two ways are different."[44] Or turn to legal practice, where lawyers recognize a distinction between "rules" and multifactor "balancing tests," the former dictating results by strict entailment and the former involving factors that combine or aggregate to dictate the legally proper result in a manner that eschews sufficient conditions and resists specification. Lastly, consider the difference between two accounts of conceptual "structure": the "classical" account that views concepts as definable by a set of necessary and sufficient conditions and the "cluster" account pursuant to which multiple criteria "count towards" or "bear upon" a concept's proper application in a given case, without any of the criteria being necessary or sufficient.[45] All these familiar dyads point to the same central division in the theory of determination. In the absence of a well

[43] Dancy, *Ethics without Principles*, 25.

[44] Dancy, *Ethics without Principles*, 25.

[45] See Margolis and Laurence, "Concepts."

settled nomenclature, but following Dancy, the labels "lexical" and "nonlexical" seem as good as any other.

After P1, the remaining premises are easy. P2 simply restates the Hartian claim that legal rules that are not accepted can exist only in virtue of being validated by the system's criteria of validity.[46] P3 reflects common scholarly characterization of Hartian validation as a process or function by which resultants are determined by satisfaction of a set of necessary and sufficient conditions.[47] P4 captures the point of insisting on principles' weightedness. As Stephen Perry encapsulates Dworkin's analysis, "the bindingness of a legal rule is nothing more than the collective normative force of the principles."[48] So even if principles could be grounded in judicial practice (as Dworkin denies), those principles combine to constitute rules, and their cumulative impact cannot be specified by a finite or tractable set of criteria.

Errol Lord and Barry Maguire, two philosophers of normativity who do not work in jurisprudence, argue that any normative theory must recognize "two central cross-cutting distinctions": the distinction between "strict" and "nonstrict" notions, and a second between "weighted" and "nonweighted" notions. Typically, nonstrict notions are weighted, and weighted notions help explain the strict.[49] For Lord and Maguire, reasons are the "paradigmatic" weighted and nonstrict normative notion—indeed, the only such notion they identify.[50] For a legal philosopher, however, Dworkin's principles are just as paradigmatic. They are weighted, nonstrict notions whose function is to contribute to a strict or decisive normative status, whereas rules are strict or decisive notions by nature whose function is to deliver decisive verdicts all by themselves (even if the decisive verdicts they purport to deliver are countermanded by others).

The surprising upshot of the challenge from principles, in short, is not that Hart's account cannot accommodate legal principles; it is that, thanks to the

[46] Hart, *The Concept of Law*, 110.

[47] See, e.g., Raz, "Legal Principles and the Limits of Law," 851; Himma, "Understanding the Relationship between the U.S. Constitution and the Conventional Rule of Recognition," 96; and Dworkin, *Taking Rights Seriously*, 62. This is not precisely right. Validation need not involve *necessary* conditions at all, and even supposedly sufficient conditions are not truly "sufficient" given Hart's embrace of defeasibility. See Berman, "Dworkin versus Hart Revisited," 560–62. But these quibbles aside, validation is a quintessentially lexical determination structure. As Hart explains, "To say that a given rule is valid is to recognize it as passing all the tests provided by the rule of recognition.... A statement that a particular rule is valid means that it satisfies all the criteria provided by the rule of recognition" (*The Concept of Law*, 103). See also Hart, *Essays in Jurisprudence and Philosophy*, 359.

[48] Perry, "Judicial Obligation," 225.

[49] Lord and Maguire, "An Opinionated Guide to the Weight of Reasons," 3–4.

[50] Lord and Maguire, "An Opinionated Guide to the Weight of Reasons," 3–4.

existence of fundamental legal principles and the nonlexical determination relationship that obtains between principles and rules, *the Hartian theory of legal content cannot explain legal rules*. The core of Dworkin's subtle argument in "The Model of Rules I" tasks positivists to explain how derivative legal rules can be partially determined by the workings of principles and not (only) by validation. The challenge from principles is, at heart, the *challenge of nonlexical determination*. It remains unrebutted.

1.3. A False Problem for Consensus: "Theoretical Disagreements"

Although positivists had not succeeded in blunting or even fully grasping his challenge from principles, by *Law's Empire*, Dworkin had fastened on a new leading argument against positivism, one that, like his first, does not depend upon the success of his own antipositivist account of law. The target of his earlier challenge, to repeat, was Hart's spin on the determination relationship that links fundamental and derivative legal norms—namely, that it involves *validation*, which is a lexical operation. Dworkin's new target was Hart's account of the practices—the ultimate rule of recognition—that ground the criteria of validity that function as fundamental legal norms. Hart makes clear that the rule of recognition depends upon a very substantial degree of judicial agreement on the criteria it picks out: "what is crucial is that there should be a unified or shared official acceptance."[51] Dworkin advanced two closely related arguments against this premise: the *challenge from theoretical disagreements* and the *challenge of too little law*. This section and the next tease these challenges apart and argue that the former, while well known and much engaged by scholars, scores no points against Hart, but the latter, though largely ignored, has far greater force.

According to the new challenge from theoretical disagreements, positivists are supposedly unable to make sense of disagreements among jurists about what the proximate grounds of derivative legal norms are, as distinguished from disagreements about whether those grounds obtain in a given case. They cannot make sense of such disagreements because, says Dworkin, positivism endorses "the 'plain fact' view of the grounds of law,"[52] pursuant to which, as Shapiro puts it, "the grounds of law in any community are fixed by consensus among legal officials."[53] Because "questions of law can always be answered by looking in the books where the records of institutional decisions are kept" and because legal actors must be taken to know this to be true, the existence of

[51] Hart, *The Concept of Law*, 115.

[52] Dworkin, *Law's Empire*, 7.

[53] Shapiro, "The Hart–Dworkin Debate," 37.

genuine theoretical legal disagreements is unintelligible on positivist premises.[54] Put in the Hartian vocabulary, Hart's account, argues Dworkin, cannot make sense of disagreements about what the criteria of validity are, as opposed to disagreements (what Dworkin terms "empirical" rather than "theoretical") about whether some criterion is satisfied.

Dworkin introduces the "snail darter case," TVA *v. Hill*, to illustrate. I will examine this case in greater depth later (in section 3.1), but the basics are enough for now. The case concerns interpretation of the federal Endangered Species Act (ESA), in particular whether the ESA required that construction of a nearly completed dam, for which millions of public dollars had already been expended, be terminated. The majority, in an opinion by Chief Justice Warren Burger, held that it did. Justice Lewis Powell, for himself and Justice Harry Blackmun, held that it did not.

As Dworkin reads the opinions, the disagreement between Burger and Powell flows from the "very different" theories "of legislation" that they adopt:

> Burger said that the acontextual meaning of the text should be enforced, no matter how odd or absurd the consequences, unless the court discovered strong evidence that Congress actually intended the opposite. Powell said that the courts should accept an absurd result only if they find compelling evidence that *it* was intended.[55]

This disagreement, Dworkin emphasizes, is entirely "about the question of law; they disagreed about how judges should decide what law is made by a particular text enacted by Congress when the congressmen had the kinds of beliefs and intentions both justices agreed they had in this instance."[56] His conclusion: this type of disagreement is unintelligible if Hart's theory is correct. A model that grounds law in official consensus is incompatible with the existence of genuine and sincere disagreements about legal fundamentals. In short, positivism maintains that "genuine argument about law must be empirical rather than theoretical."[57]

Notice that this argument relies upon two distinct premises: (1) that *q* is the law only if validated by criteria supported by official consensus; and (2) that the officials whose consensus grounds legal content know 1. Premise 2 is essential to Dworkin's argument because there is no difficulty explaining judges' sincere disagreements about what the legal fundamentals are if they do not fully

[54] Dworkin, *Law's Empire*, 7.

[55] Dworkin, *Law's Empire*, 23.

[56] Dworkin, *Law's Empire*, 23.

[57] Dworkin, *Law's Empire*, 37.

appreciate that what they are depends constitutively on judicial agreement. Yet Hart does not stipulate that those who are disagreeing know (or believe) that *q* is a legal norm if and only if the fundamental legal notions are the subject of judicial consensus. Whether judicial near-consensus grounds legal rules and whether participants know this to be true are separate questions. Hart's theory explicitly asserts the former but not the latter.

So Dworkin needs an argument to establish that participants to putative theoretical disagreements must know that the plain-fact view is true, hence cannot be genuinely uncertain about what our legally fundamental norms are. Dworkin supports this premise by attributing to his opponents the claim that "the very meaning of the word 'law' makes law depend on certain specific criteria, and that any lawyer who rejected or challenged those criteria would be speaking self-contradictory nonsense."[58] In *Hill*, "past legal institutions had not expressly decided the issue either way, so lawyers using the word 'law' properly according to positivism would have agreed there was no law to discover."[59]

But this attribution is baseless. Hart flatly insisted that there was "no trace" in his work of the idea that his rule of recognition and associated criteria of validity were baked into the word "law."[60] And most commentators have thought it plain that positivism is not in the business of defining words.[61] So the semantic sting cannot furnish what the challenge from theoretical disagreements needs. And the challenge fares no better if we replace Dworkin's semantic claim with a conceptual one. It is no part of Hart's theory that it is part of our concept LAW, if not our word "law," that legal norms are grounded in judicial consensus.[62]

[58] Dworkin, *Law's Empire*, 31.

[59] Dworkin, *Law's Empire*, 37.

[60] See Hart, *The Concept of Law*, 247.

[61] See, e.g., Leiter, "Beyond the Hart–Dworkin Debate": "if any argument is no longer worth discussing, it is this one" (31n49). See also Kramer, *H. L. A. Hart*, 207n2.

[62] What content Hart ascribed to our shared concept of law is surprisingly unclear given his monograph's title. My own view is that insofar as we share a concept of law, its core is that law concerns the set of norms delivered and sustained by legal systems, which are artificial normative systems established and maintained by political communities and designed to serve a potentially limitless range of functions, characteristically including resolving disputes among community members and preserving public order. I think this was close to Hart's own view at times and that he never meant to reduce the concept of law to the union of primary and secondary rules. See Hart, *The Concept of Law*, explaining that he has sought "to give an explanatory and clarifying account of law as a complex social and political institution with a rule-governed (and in that sense 'normative') aspect" (239). But I cannot pursue these claims further here.

1.4. A Genuine Problem for Consensus: "Too Little Law"

If, contra Dworkin, the existence of "theoretical disagreements" causes little trouble for Hart's view that the practices that ground fundamental legal norms must involve official consensus, a nearby argument that has attracted considerably less attention does. I call this Dworkin's challenge of too little law. The problem it poses for Hart is not that his account cannot explain genuine and sincere disagreements about the fundamental legal norms. (That is the subject of the challenge from theoretical disagreements.) It is that when judges do disagree on the fundamentals, neither side can be correct about what the law is. Even if Burger and Powell could have held their conflicting views sincerely, neither could have been right.

According to the orthodox reading of Hart, whenever the relevant officials (paradigmatically judges) fail to converge on some putative "criterion of validity" or whenever they agree that some criterion "counts" but fail to converge on how it fits within the rule of recognition's overall logic, to that extent, the criteria grounded in the rule are unable to perform their validating function. "Where there is no consensus, there is no law."[63] Unfortunately, in the mature legal systems we are most familiar with, these failures of convergence are likely to be common. The worry looms especially large in theoretical debates over American constitutional law. Many constitutional scholars believe that such failures and gaps thoroughly characterize American constitutional practice, that very few constitutional disputes that reach the US Supreme Court (and even the federal appellate courts) are determinately resolved by criteria that enjoy near-consensus judicial recognition.[64] In consequence, Hart's account seems to entail that there is much less (constitutional) law than appears correct to many sophisticated observers, even on reflection. This is the *too-little-law objection*: if Hart's account of law were correct, "it would follow that there is actually almost no law in the United States."[65] This was not a throwaway line: Dworkin pressed it for forty years.[66]

To this critique, the usual responses are available: "Not so!" and "So what?" Let us take them one by one.

The "Not so!" response is very tempting because, frequent repetition notwithstanding, Dworkin's charge of "almost no law" is obviously exaggerated.

[63] Barzun, "The Positive U-Turn," 1355.

[64] The most thorough study to reach that conclusion is Greenawalt, "The Rule of Recognition and the Constitution." See also Greenberg, "What Makes a Method of Legal Interpretation Correct?" 124; and Leiter, "Explaining Theoretical Disagreement," 1224.

[65] Dworkin, "Hart's Posthumous Reply," 2116.

[66] See Dworkin, *Law's Empire,* 10, and *Taking Rights Seriously,* 350.

But the question is not whether Dworkin's rhetoric matches the reality. It is whether the grain of truth behind the hyperbole is large enough to warrant being taken seriously. The answer to that question strikes me as plainly yes. Some American judges, including some on the Supreme Court, recognize originalist or textualist criteria of validity that render invalid major pieces of federal legislation and vast swaths of federal administrative regulations. If the Hartian account of legal content that I have sketched is the best positivist account of legal content available, then very many questions of federal statutory and regulative law are underdetermined, not just little pockets here and there.

That leaves the second response: "So what?" Unlike the first retort, this one acknowledges that American law is much less determined, much gappier, than American lawyers and legal scholars, let alone laypeople, routinely suppose. The "So what?" response simply denies that that fact undermines the Hartian account. Brian Leiter is perhaps the most notable champion of this rejoinder.[67] In his estimation, few if any controverted questions of American constitutional law *do* have legally correct answers, making what Dworkin thought a bug of Hart's theory a feature.

Leiter could be right, of course. But how bitter is the bullet to be bitten depends on how many considered casuistic judgments the diner would have to abandon. Ordinary thought and talk about law, including about American constitutional law, is cognitivist on its face. And very many speakers, including supposed sophisticates, routinely attribute determinate constitutional properties ("unconstitutional" being the most common) to acts even when the correctness of the attribution depends upon legal premises that we know to be controversial. Furthermore, my own considered judgments that thus-and-such is constitutionally prohibited or constitutionally permitted often survive despite my knowledge that my judgment rests on controverted premises. Simply put, it frequently feels to me, when "playing judge," that there are legally right answers to a good number of controversial cases. Furthermore, many colleagues report the same. Even if my judgment that there is law even where there is disagreement could be wrong, it is obviously not idiosyncratic to me and

[67] See Leiter, "Explaining Theoretical Disagreement." Bill Watson is with Leiter, advocating a Hartian account of the validation of legal sources married to the "standard picture" of law in which legal content is determined by the pragmatically enriched communicative contents of those sources. Watson, "In Defense of the Standard Picture." He has argued that that package explains the wide expanse of legal *agreement* better than other theories, including (in personal communication) principled positivism. See Watson, "Explaining Legal Agreement." My impression is that Watson overstates the degree of support for the standard picture in US legal practices and underestimates the ability of principled positivism, at least, to explain agreement. But his careful arguments warrant closer attention than can be afforded here.

Dworkin. On coherentist reasoning, these judgments, along with attributes of ordinary legal thought and talk, are enough to justify our treating the too-little-law objection as a genuine challenge for positivists, at least provisionally.[68] If positivists cannot amend Hart's account to make plausible that some legal propositions are true despite the lack of near consensus on their truthmakers, that some legal rules exist in the absence of uniform support for the principles that are their determinants, then Leiter's response remains available.

2. FROM HARTIAN POSITIVISM TO PRINCIPLED POSITIVISM

I have argued that Dworkin marshals two troubling objections to a Hart-inspired positivist account of legal content: that it cannot satisfactorily explain the existence *and operation* of legal principles (i.e., that they play a role in making legal rules what they are but do so in a fashion that does not involve "validation"); and that it does not allow for as much law as legal sophisticates believe there to be, even on reflection. If so, what follows? Dworkin's own conclusion is that we should abandon positivism.

This article pursues an alternative possibility. It is to revise or supplement Hart's account in a way that enables positivism (1) to accommodate genuine legal principles that participate in the nonlexical determination of derivative legal rules and (2) to allow for fundamental legal norms to emerge from legal practices that fall significantly short of consensus. Many leading positivists have long believed that Hart's account is overly regimented or incomplete and that

[68] An anonymous non-American reviewer worries that even if my judgment that law survives official dissensus is held by other American constitutional scholars, that view is too parochial to warrant being taken seriously by others. Rather, in their estimation, the fact of significant judicial dissensus on fundamentals "in the USA merely serves as evidence that that country has a defective/malfunctioning legal system," and my effort to articulate a positivist theory that would vindicate my and others' judgment that law survives dissensus only bolsters already well-warranted suspicions that the literature on American constitutional theory "is, basically, a systematically disingenuous discourse."

I find those judgments too harsh but not baseless. See Berman, "Our Principled Constitution," 1334. This cannot be the place to defend American constitutional theory writ large. I acknowledge that this article will hold greater interest for readers who antecedently believe that there is sometimes law even when judges disagree about legal fundamentals, and that that set of persons will possibly include American constitutional theorists disproportionately. But even scholars (of any nationality) who do not actively believe that claim should be more open to it than the reviewer's comments suggest. If you start off disposed toward a positivist account of legal content but open to the too-little-law challenge, then you have all the reason you need to give a non-Hartian account of legal content an honest hearing. If you then find my alternative account unpersuasive on other grounds, the exercise will still have returned value if it increases your confidence in what I am calling the Hartian theory of content.

some loosening, reworking, or supplementing would be required to render positivism a fully adequate theory of law.[69] This is my attempt to contribute to that effort by bringing a less tightly structured vision of legal content into crisper resolution.[70] Success in this endeavor would not disprove antipositivism but would make positivism vastly more eligible.[71]

[69] See, e.g., Soper, "Legal Theory and the Obligation of a Judge": "It may be that we have moved some distance from the view that a 'master test,' capable of actually identifying with some precision all standards relevant to legal decision, forms the core of a positivist's theory" (514). See also Schauer, "Amending the Presuppositions of a Constitution": "In referring to the ultimate rule of recognition as a *rule*, Hart has probably misled us.… The ultimate source of law … is better described as the practice by which it is determined that some things are to count as law and some things are not" (150–51). See also Kramer, *H. L. A. Hart*: "A satisfactory theory of law has to include a much better account of legal reasoning and interpretation than the account offered by Hart" (205). See also Bayles, *Hart's Legal Philosophy*, 170. John Gardner disagrees with Kramer when insisting that legal positivism is only "a thesis about legal validity." See note 23 above. I am with Kramer in believing that Gardner's characterization of legal positivism is stipulative and unduly narrow. A comprehensive or complete positivist theory of law would include a theory of legal content, whether or not that was of interest to Hart.

[70] Dworkin anticipated and dismissed a view that some might think resembles the one I am presenting. After arguing that principles cannot arise by validation or by acceptance, he offered this final possibility: "If no rule of recognition can provide a test for identifying principles, why not say that principles are ultimate, and form the rule of recognition of our law?" (*Taking Rights Seriously*, 43). The law of a jurisdiction would, on this view, be "all the principles … in force in that jurisdiction at the time, together with appropriate assignments of weight. A positivist might then regard the complete set of these standards as the rule of recognition of the jurisdiction" (43). "This solution," says Dworkin, "is an unconditional surrender. If we simply designate our rule of recognition by the phrase 'the complete set of principles in force,' we achieve only the tautology that law is law" (43–44).

My version of positivism, like that of Dworkin's imagination, holds that the complete set of principles, with their relative respective weights, constitutes the fundamental legal norms of a community. But that is where the commonality ends. Principled positivism does not treat the existence of such fundamental principles as a brute inexplicable fact but as metaphysically determined by the practices by which participants in a legal system take them up in legal decision-making. Furthermore, rather than relying upon a "rule of recognition" and the validation with which it is associated, principled positivism maintains that fundamental weighted principles determine derivative norms nonlexically. The view could be wrong and still wants for detail, but it does not approach a tautology.

[71] This is an important point about the dialectic. I started (in section 1.1) by assuming some claims about the nature or essence of law, including that legal norms are only thinly normative. I am trying to provide a better account than Hart's of the socio-factual grounding of legal norms so conceived. This way of proceeding cannot establish that my starting assumptions are correct, which is close to the nub of the disagreement between positivists and antipositivists. See Tripkovic and Patterson, "The Promise and Limits of Grounding in Law," 222–26. Nonetheless, my effort, if successful, does improve positivism's prospects in its battle with antipositivism because a choice between them depends on comparative

Here is the preview. Fundamental legal principles are grounded in practices more or less as ordinary social norms are: by dint of legal actors taking them up in legal decision-making. Their scopes and relative weights are grounded dynamically in argumentative legal practices. Individual principles bear constitutively on the legal status of a token act or event—that the act or event is legally permissible or impermissible, legally valid or invalid, etc.—by exerting force toward one status or the other. The force any one principle exerts is a function of two variables: the principle's own relative weight or importance within the legal system; and the extent to which the principle is "activated" by the presence of legal practices or other phenomena that the principle "turns upon" or makes legally relevant. The all-things-considered legal status of a token act or event is determined by the aggregate force of the activated principles (think vector addition) or by more complicated functions that, like the principles themselves, are also grounded in legal practices. Rules are reflections of the legal status of properly described act or event types; they describe the curvature of legal-normative space that is effected by the aggregative force of the principles.

That is a highly condensed summary. The key differences between this model and the Hartian model are two. They concern, first, how the fundamental legal norms—principles—bear on nonfundamental legal notions (in a nonlexical, aggregative manner) and, second, how those fundamental legal norms are themselves grounded in practices (by being taken up by legal actors and thereby embedded in the legal materials rather than by convergent agreement or acceptance). These two differences are what enable the full account to meet the two challenges that hamstrung Hart's theory. (See figure 5.)

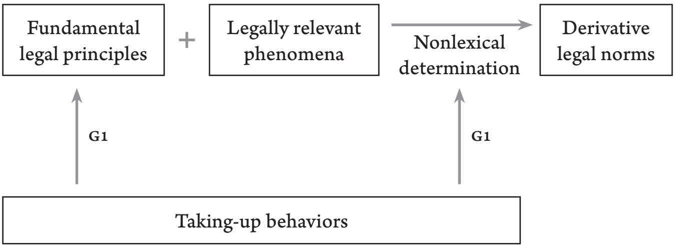


FIGURE 5 Principled Positivism

This section develops the picture in four steps. Section 2.1 explains how fundamental contributory norms—legal principles—are grounded in practice. Section 2.2 explains how these fundamental principles, along with all the facts, practices, or phenomena that they reference or make legally relevant, combine

tallies of overall plausibility points, as David Enoch argues with respect to competing metaethical theories. See Enoch, *Taking Morality Seriously*, 14–15.

by nonlexical aggregation to determine the legal properties (such as being legally permitted or prohibited, or legally valid or invalid) that attach to token acts and events and, in so doing, to determine derivative and "summary" legal rules. Section 2.3 explains why the determination function between fundamental principles and summary rules is what it is, or in virtue of what it has the particular form or content that it does. Section 2.4 adds a further clarification about legal rules, contrasting the summary conception introduced in section 2.2 with a second conception of "promulgated" rules. It explains how promulgated rules contribute to summary rules by operation of the fundamental legal principles.

2.1. How Legal Practices Ground Legal Principles

A legal principle exists in legal system *S* in virtue of being "taken up" by a legal agent or institution in a legally significant speech act (such as deciding judicial cases, enacting, signing, or vetoing legislation) that purports to invoke and rely upon such principle.[72] That's the basic idea, though of course it puts matters too simply. Let me elaborate.

What determines whose behaviors count and to what relative degree is not a brute fact constant across all legal systems but is itself a product of the recognitional attitudes and behaviors of members of the legal-normative community. Those persons who play privileged roles in the determination of the fundamental legal norms are those whom other participants in the practice recognize as having privileged law-determination roles. So whose speech acts matter and how much they matter are largely products of who members of the community take to matter. Think fashion. Whose fashion decisions matter is determined by those persons whom others in the fashion community (or proto fashion community) take to have capacity to set the fashion norms.

That said, legal actors disagree about our principles, both synchronically and diachronically. It is implausible that the single invocation of a putative legal principle by a single actor in the face of opposition is sufficient to render the putative principle a principle of the system or sufficient to endow the principle with the same importance as possessed by a principle that enjoys broad, longstanding, and durable support. So we ultimately need some handle on how patterns of acceptance and rejection, skepticism and enthusiastic embrace, all bear on the contents and relative importance of the resulting principle.

[72] Cf. Postema, "Classical Common Law Jurisprudence (Part 1)," arguing that, for "common lawyers . . . , the law in its fundament was understood to be not so much 'made' or 'posited'—something 'laid down' by will or nature—but rather, something 'taken up,' that is, used by judges and others in subsequent practical deliberation" (166).

While the answer is surely complex and likely possesses elements of a sorites problem, I do not think there is any deep mystery about how fundamental norms can be grounded in social practice, even as particulars elude us. As Rolf Sartorius suggested decades ago, fundamental norms arise within an institutionalized normative system when they have the type of "institutional support" to which Dworkin drew our attention: they are "embedded in or exemplified by numerous authoritative legal enactments: constitutional provisions, statutes, and particular judicial decisions."[73] The more a principle is taken up by the relevant actors and the more that subsequent legal decisions rely upon and reinforce the principles or the decisions they are understood to underwrite, the more secure is the principle's status as a legal norm of the system.

Undoubtedly, this basic picture calls for detail and refinement. Here, however, I want only to highlight two points. First, this is a positivist account because embeddedness is an explanatory, not justificatory, notion. It concerns, in some fashion, what judges (and others) do accept or how they do reason, not what they should accept or how they should reason.[74] Second, for a standard to be embedded in the legal materials does not require that it enjoy anything approaching the near-consensus support that Hart required and that some theorists hostile to the possibility of distinctly legal principles have thought essential to positivism.[75] As C. L. Ten emphasizes, an intelligible version of positivism may tolerate "considerable disagreement among judges about what rules and principles are embedded in the legal sources." But it is nonetheless "dependent on social practice—the practice of recognizing constitutional provisions, legislative enactments and judicial decisions, as well as what is embedded in them, as legal standards."[76] Indeed, "there is no important difference" between how Dworkin would assess fit "and the view of the legal positivist who extracts legal principles from legal sources in the manner [just] suggested.... Both appeal from the settled and explicit rules to what is embedded in them."[77]

[73] Sartorius, "Social Policy and Judicial Legislation," 154–55. See also Sartorius, *Individual Conduct and Social Norms*: "A principle is relevant if and only if, *and to the degree to which*, it enjoys what Dworkin aptly calls 'institutional support'" (193).

[74] Dworkin fails to appreciate this possibility in his response to Sartorius in "The Model of Rules II" (reprinted in Dworkin, *Taking Rights Seriously*, 66–68).

[75] See Alexander and Kress, "Against Legal Principles," 767–68.

[76] Ten, "The Soundest Theory of Law," 530.

[77] Ten, "The Soundest Theory of Law," 532. When further explicated, the notion of embeddedness will rely on some elements of coherence and support some versions of coherence theories of law. See Sartorius, *Individual Conduct and Social Norms*, 196–99. But I tread cautiously here, for existing coherence-based theories of law reflect at turns both unclarity and disagreement regarding the particular relata that must be brought into coherence. See generally Kress, "Coherence"; and Rodriguez-Blanco, "A Revision of the Constitutive and

The difference between a model in which the social-factual grounds involve the taking-up and embedding of principles (mine) and one that requires judicial near-consensus (Hart's) is illustrated by the familiar (putative) principles of American equal protection law customarily termed "colorblindness" and "antisubordination." They are frequently arrayed against each other in concrete legal disputes, especially concerning state-mandated preferences for racial minorities, making it possible that neither has ever attracted support from or been accepted by a super majority of judges or other legal elites. If legal principles depend for their existence on something approaching full agreement among members of one or another class of legal actors, then neither colorblindness nor antisubordination (however the latter may be glossed) would qualify as a principle of American law.

But many constitutional lawyers would resist that conclusion. Consistent with the alternative Sartorius-Ten account, many American constitutionalists would say that *both* are principles of our law. Each is a principle in virtue of having been invoked, relied upon, or used as legal justification for judicial rulings. And each has become further embedded in our law to the extent that the decisions that have taken it up serve as support for additional judicial decisions or are approved and championed by other legal (and popular) elites. Broadly, then, *q* may be grounded not only in acceptance or invocation of *q* itself but also in acceptance, as legally correct, of decisions or rulings that *q* is understood to explain. In such fashion does a principle become embedded in the law, regardless of whether a head count would establish that nearly all judges accept it.

The most common worry about this part of the picture is not that positivist legal norms cannot be embedded in this (admittedly gestural) manner but that such norms cannot have the dimension of weight. This is the chief objection to positivist legal principles that Larry Alexander and Ken Kress advance in their aptly titled article "Against Legal Principles."[78] As they summarize: "We cannot establish principles by agreement because we cannot establish their weights by agreement."[79]

There are two responses. The first is technical. As we will see in section 2.2, my account, unlike Dworkin's, does not require that the principles have varied weights. It could be that all fundamental principles have equal weight. All that is required is that their manner of determination (D2) is aggregative or, in any event, nonlexical.

Epistemic Coherence Theories in Law." See also Hurley, "Coherence, Hypothetical Cases, and Precedent."

[78] Alexander and Kress, "Against Legal Principles," 761–64.

[79] Alexander and Kress, "Replies to Our Critics," 925.

In fact, though, I believe that fundamental principles often do vary in importance or weight. Thus the second response. Alexander and Kress explicitly assume a form of positivism in which fundamental legal norms can arise only by agreement or consensus about that fundamental norm.[80] Once we soften this supposed requirement, as the Sartorius-Ten picture proposes, then it is no longer difficult to envision rough weights emerging from judicial practice. As I have elsewhere argued:

> The weights of principles, like their contents or contours, are brought about by members of the legal community taking them up and deploying them in legal reasoning and decision-making. Weights are relative to one another, and are given by what members of the legal community say about them and how they use them. They are also conferred, as it were, by battle—by the rules that are adjudged victorious, and thus made so, when principles press in opposing directions.[81]

Weights conferred in this manner will be rough at best (think: slight, moderate, weighty, very weighty, or nearly conclusive; *not,* e.g., 12 or .68) and change in organic fashion that is usually gradual. A principle's relative weight ebbs and flows, much as its contours constrict and expand. Compare the principles that partially constitute a person's psychological or deliberative profile. Each of us acts upon a different bundle of ethical and practical principles—principles that favor keeping promises, trying new experiences, planning for the future, promoting justice, respecting one's elders, and so forth. The principles that make out an individual's psychological profile are not arrayed in a tightly structured hierarchy, let alone once and for all. But they must exhibit a nontrivial degree of stability and consistency to underwrite personal integrity—in the sense of coherence, not moral worth. The same is true of legal systems, which is one kernel of truth underpinning Dworkin's theory of law as integrity.

Return to our equality principles of colorblindness and antisubordination. If the disputes in which the two pull in different directions are reliably resolved in favor of colorblindness (assuming that other relevant principles are in rough

[80] Alexander and Kress, "Against Legal Principles," 767 and n106.

[81] Berman, "For Legal Principles," 254. The gist of my argument there is that Alexander and Kress marshal forceful objections to Dworkin's picture of legal principles as suboptimal moral principles that morally justify legal rules and outcomes but score no damage against a positivist picture in which legal principles, grounded in social facts, participate in the metaphysical determination of legal rules. Broadly similar verdicts are reached by Leiter, who argues that "Against Legal Principles" "is actually devoid of any arguments against the *existence* of legal principles" ("Explanation and Legal Theory," 906). See also Lawson, "A Farewell to Principles."

equipoise), that very pattern of decisions would make it the case that it is (for the time being) the weightier principle.

2.2. How Legal Principles Make Legal Rules

We now reach a further objection to a positivist picture that accommodates, let alone foregrounds, nonlexical determination—not that legal practices cannot deliver variably weighted principles but that any principles practices deliver cannot combine to determine anything resembling rules. They can of course be used by judges when deciding what to do or what rules to create. But they cannot combine to determine legal content that judges are able to discover or ascertain rather than make. The concern is just another instantiation of the demand that has been made of normative pluralists of all stripes, from W. D. Ross to Isaiah Berlin to Philip Bobbitt: to explain how the all-in derives from the contributory.[82] In the case of principled positivism, the challenge is to explain *how* legal "principles" (legal norms with possibly variable weights, grounded directly in practices of legal participants) combine to constitute or determine legal "rules" (determinate legal norms not directly grounded in taking-up practices) if *not* by collectively constituting a set of (usually) sufficient conditions. Baude and Sachs vividly formulate this challenge to a preliminary sketch of my account, wondering how a large number of variegated norms with diverse weights can determine or constitute more determinate legal norms (rules) "rather than merely make soup."[83]

The obvious answer, which I've been previewing for many pages, is "by aggregation." Rules and principles are types of norms; norms are kinds of forces or, at a minimum, can be fruitfully analogized to forces (they push or press or weigh or favor); and forces can combine by force addition.[84] This is Stephen Perry's approach. As Perry explains, "the principles that are relevant to a particular situation are assumed to be commensurable and capable of being

[82] Think of "the priority problem" that Rawls worries bedevils all forms of "intuitionism" (Rawls, *A Theory of Justice*, chs. 7 and 8). The same concern underwrites doubts that nonclassical accounts of concept structure are intelligible. See Davies, "The Cluster Theory of Art"; and Margolis and Laurence, "Concepts."

[83] Baude and Sachs, "Grounding Originalism," 1489 (criticizing Berman, "Our Principled Constitution"). See also Alexander, "The Banality of Legal Reasoning": "No one—not even lawyers—can meaningfully 'combine' fact and value, or facts of different types, except lexically. . . . Any non-lexical 'combining' of text and intentions, text and justice, and so forth is just incoherent, like combining *pi*, green, and the Civil War. There is no process of reasoning that can derive meaning from such combinations" (521).

[84] See Ross, *The Right and the Good*, 28–29.

aggregated, along their dimension of weight, so as to produce an overall balance of principles."[85]

Imagine a legal-normative field defined by the poles "is legally prohibited" and "is not legally prohibited." Then consider any token act or event, *x*, that is a proper subject of the predicates that define the field. Any given legal principle, P_n, will have no bearing on the status of *x*, or will bear constitutively for one of the polar properties or its opposite. The token *x* thus acquires the legal property or status that corresponds to the greater net force of the principles.

Figure 6 illustrates this dynamic, where the height of a vector arrow represents the principle's relative weight or importance, its direction represents whether it militates for or against the legal permissibility of the conduct at issue under the circumstances, and its length represents the extent to which the principle bears toward one normative pole or the other given the relevant facts.

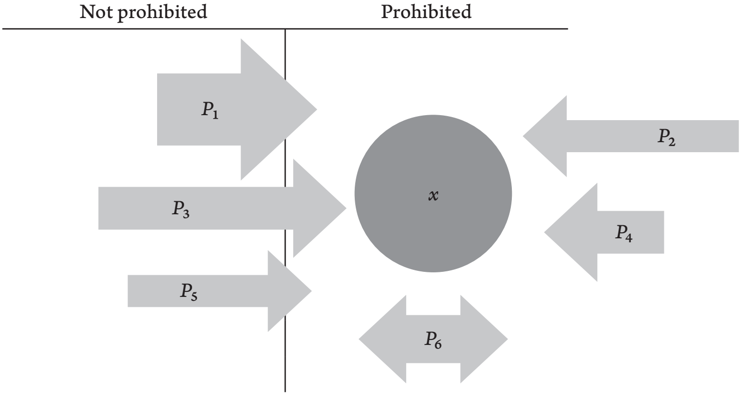


FIGURE 6 Nonlexical Determination of Rules by Principles (Intuitive Model)

Here are several things one can read off the graphic: P_1, P_3, and P_5 have the same "valence" with regard to *x*: they all bear toward its being prohibited. P_1 is a weightier principle (it possesses more potential force) than P_3 or P_5, but P_3 is more fully activated against the permissibility of *x* than P_1 (it exerts more of its potential). A two-headed arrow, representing principle P_6, has no net impact on the legal permissibility of *x*, either because it exerts itself equally in both directions at once or because it doesn't bear at all.

All the same information can be represented by a more orthodox representation of vector addition.[86] In this model, a principle's relative weight (a context-invariant property) is represented by its length, and the degree of its

[85] Perry, "Two Models of Legal Principles," 788. See also Perry, "Second-Order Reasons."

[86] I am grateful to my student Brandon Walker for urging me to deploy this standard model for representing vector addition.

activation (a context-variant property) is represented by the angle it describes relative to neutrality (here represented by the *y*-axis). The force that the principles exert collectively is determined by linking the arrows head to tail. If the chain of vector arrows starts at neutral, then the act or event *x* has the legal property or status that corresponds to the area of the plane where the chain ends. Figure 7 below captures the same information conveyed in figure 6 above.

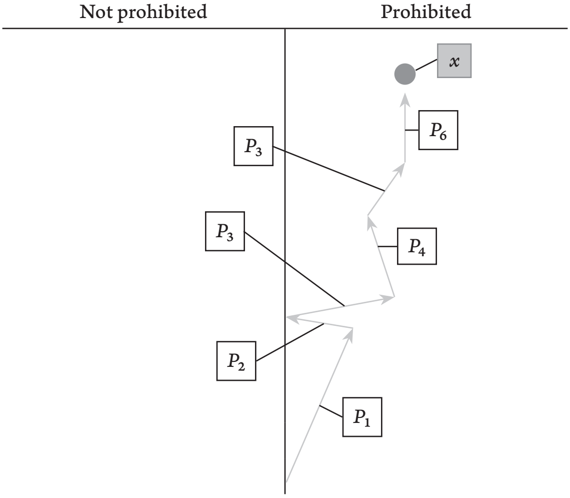


FIGURE 7 Nonlexical Determination of Rules by Principles (Orthodox Model)

A legal rule is a description of the legal status of a contiguous stretch of tokens that share the same legal status.[87] It reflects the normative status of an act *type*, where that status is derivative of the like statuses of all the tokens of that type. If [x_1 is prohibited] and [x_2 is prohibited] and [x_n is prohibited], there will be some description of the act type *X* for which it is true that [*X* is prohibited]. The rule [*X* is prohibited] is the summary of a range of instances of [x_n is prohibited] where each token prohibition obtains in virtue of the net bearing of the fundamental principles on x_n. On this view, says Perry, a rule "is regarded as nothing more than a device of convenience, a kind of *aide-mémoire* for recording the perceived aggregate consequences of the various principles that bear on the resolution of a specific kind of dispute."[88]

[87] Here and throughout, I have said that principles operate upon tokens, not types. I believe that this is a more promising way to explain how legal properties can be assigned to token acts or events themselves, as we should ultimately wish, and not only to descriptions of them. But many words could be expended on this question, and I do not believe that the substance of the argument changes if you think principles operate upon finely defined act or event types.

[88] Perry, "Judicial Obligation, Precedent and the Common Law," 225.

Perry is an antipositivist. But nothing about the summary picture of rules just sketched is obviously uncongenial to positivism. The supposed trouble for positivism arises when we return to the problem of weights. The objection now becomes not that principles cannot accrue weight or importance in the way described in section 2.1 but that, as that discussion emphasized, such weights can only be rough, and we need more determinacy if principles can jointly determine rules as the summary conception envisions. Perry encourages this line of argument, noting that "it is difficult to see how custom could be sufficiently nuanced as to be able to assign determinate weights to individual principles."[89]

Whether his doubts are well founded depends on how determinate principles' respective weights must be, and the answer to that question is supplied by functional considerations: the weight of principles must be as determinate as need be for principles to do their job tolerably well. So the objection to a positivist picture of the determination of rules by the aggregation or accrual of weighted principles reduces to the claim that, on any reasonably contestable legal question, some principles will press one way, some will press the other, and their net impact, and thus the legal upshot, will too frequently be underdetermined, metaphysically and epistemically.[90] Thus would principles require more finely specified weights than practice can be expected to deliver.

I do not find this objection persuasive. For one thing, we should not assume that a roughly equal number of principles will routinely bear for and against competing candidate legal rules. In many cases, the sheer number of principles pointing one way will dwarf the number pointing against.[91] As significantly, the total force that a principle exerts on a given legal question is not determined exclusively by its weight. I have already noted that the force a principle exerts in a given context toward a determinate legal status (e.g., valid, prohibited, permitted) is a function of two variables, not one: the weight of the principle, and

[89] Perry, "Two Models of Legal Principles," 794. As a second reason to doubt a positivist account predicated on the accrual of principles, Perry also agrees with Dworkin "that legal principles are in any event not treated by common law judges as rooted purely in custom." Perry, "Two Models of Legal Principles," 794, citing Dworkin, *Taking Rights Seriously*, 43–44 and 64–65. But the fact that judges invoke moral arguments when trying to establish that a putative principle is a legal principle of the jurisdiction, or has this or that weight, does not prove that those arguments are good ones, that they do go toward establishing what they purport to establish. As I argue in Berman, "Dworkin versus Hart Revisited," judicial practices ground principles, while the fact that judges believe these principles are morally good causally explains the judicial practices that are the grounds (574–76).

[90] See, e.g., Sartorius, *Individual Conduct and Social Norms*, 193–94.

[91] Cf. Fallon, "A Constructivist Coherence Theory of Constitutional Interpretation," arguing that the recognized "modalities" of American constitutional argument usually align, or can be viewed as aligning, even in hard cases.

the extent to which the principle is (as I call it) "activated."[92] Take a possible legal principle that provides that *historical practice matters*. The total force this principle exerts in favor of the putative legal fact [*x* is legally permitted] will depend on how long and widespread the practice of *x*-ing has been. A principle that gives effect to some communicative content of a text activates more fully the clearer that content is. Weight may be constant across contexts—though *not* over time—while *activation* is context sensitive.[93] Given the role played by context-variant activation, the net force of principles may well yield rules determinately even when particular principles' relative context-invariant weights are highly uncertain—which is not to deny that some underdeterminacy, possibly substantial, will remain.[94]

The difference between what I am calling "weight" and "activation," though widely overlooked, is of great importance. Alexander and Kress, the arch-critics of legal principles, assert that, "because principles' weights vary in different concrete contexts, a complete account of principles requires differing weights for every conceivable context."[95] That is mistaken. What is required is that the *force* that a principle exerts can vary across contexts, not that its *weight* does. An analogy: the mass of a body and thus the gravitational force it has the capacity to exert is not contextually variant, though the gravitational force that it does exert on an object in a given context also depends on its distance to that object, which is context-variant. This is a pregnant comparison, for artificial normative systems can be conceptualized in terms of normative fields, analogous to gravitational fields. Normative fields are created and sustained by a convergent practice among participants or "subscribers" in more or less the way described by Hart's rule of recognition. Principles are constituted by the taking-up behav-

[92] Cf. Alexy, "Formal Principles," defining the "concrete weight" that a principle exerts in context as a function of, *inter alia*, the principle's "abstract weight" and the "intensity of interference" with the principle under the circumstances.

[93] The temporal inconstancy of principles follows from the facts that they and their weights are grounded in human behaviors and that human behaviors are inescapably dynamic.

[94] To be clear, I am addressing the worry that the balance of principles will be underdeterminate in a great many cases—many more than would be consistent with widespread judgments among sophisticates regarding the actual extent of legal underdeterminacy. I am not responding to Dworkinian anxiety that there will be *some* underdeterminacy and therefore that the picture I present leaves some room for judicial discretion. I share the common judgment that a positivist "can reject the model of rules yet accept the doctrine of judicial discretion" (Lyons, "Principles, Positivism, and Legal Theory," 422). Just as significantly, the thought that discretion begins where already determined law ends is untrue to the relevant phenomenology. When struggling toward the law in difficult cases, judges do not experience a clean divide between (1) trying to ascertain existing law and (2) creating new legal norms. See Sartorius, "Social Policy and Judicial Legislation," 156–60.

[95] Alexander and Kress, "Replies to Our Critics," 924–25.

iors of the system's subscribers (or of some subset). Principles operate within the normative field much as masses do within a gravitational field. Rules are articulable descriptions of stretches of the curvature of the normative field that the principles effect.[96]

One final analogy, this time from the study of Multiple-Criteria Decision Analysis (MCDA) and Multi-criteria Analysis (MCA) in such fields as decision theory, management science, and fuzzy logic. As the names suggest, MCDA and MCA concern how decision makers should reach overall assessments about the relative value ranking of options that implicate a multiplicity of criteria, factors, or attributes.[97] Although not yet well known in law and legal theory, the field is many decades in development, and its tools and methods are routinely deployed across industry, finance, science, and governance, on questions ranging from how to build an investment portfolio to where to locate an airport to which students to admit to a graduate program.[98] The simplest and most widely used of all MCDA and MCA models is simple additive weighting (SAW) and its variants.[99] Wrinkles aside, a decision maker employing SAW "directly assigns weights of relative importance to each attribute" and then obtains a total score "for each alternative by multiplying the importance weight assigned for each attribute by the scaled value given to the alternative on that attribute, and summing the products of all attributes."[100] The simple model I adapted from Perry as an example of how principles can aggregate to determine summary rules is little more than the conversion of a powerful, widely used decision-making protocol into a model of the metaphysics of artificial normative systems.

2.3. *On the Determination of the Determination Function*

The argument to this point explains how variably weighted norms grounded in legal practice, by being taken up and further embedded, could aggregate to determine decisive summary norms, and not *only* to be used by judges to

[96] I doubt that this model of determination is properly classified as aggregation, which helps explain why I locate the critical distinction among modes of determination (section 1.2) at a higher level of generality—between lexical and nonlexical rather than between validation and aggregation.

[97] A useful introduction and overview is Goodman and Wright, *Decision Analysis for Management Judgment*.

[98] See Lindell, *Multi-criteria Analysis in Legal Reasoning*, who notes that "while the volume of literature in its own field of knowledge is extensive, there is very little written in legal literature about MCA and fuzzy logic" (8–9) and speculates that the literature's relative formal and scientific language has impeded its reception by lawyers and legal scholars.

[99] See, e.g., Abdullah and Adawiyah, "Simple Additive Weighting Methods of Multi-criteria Decision Making and Applications."

[100] Lindell, *Multi-criteria Analysis in Legal Reasoning*, 48.

make law when existing legal content is underdetermined (or is believed to be underdetermined). But even if determination of this sort is possible, is it actual? What would make it the case that principles do aggregate in this fashion, either generally or in a given legal system? After all, an aggregative system could take many forms. It could incorporate thresholds or eschew them. It could involve more complicated operators, such as the multipliers, enablers, and defeaters familiar from current theories of practical reasoning.[101] It could be only partially aggregative, including lexical features too. What makes it the case that a given legal system *S* maps principles to all-in legal facts—and thus to summary rules—*this* possible way rather than *that* possible way? If it is true that *R* is a rule of *S* if the aggregate force of principles favoring *R* exceeds the aggregate force of principles favoring ¬*R*, in virtue of what would this be so? What determines the determination function between fundamental norms and derivative ones?

The answer, I think, has two components. The first traces once again to insights supplied by an antipositivist—this time Mark Greenberg. Greenberg persuasively argues that it is part of the nature of law and legal systems that the determination relationship between practices (or practice facts, in the terminology that Greenberg prefers) and legal norms (or facts) must satisfy what he calls "the rational-relation doctrine," which provides that "the content of the law is in principle accessible to a rational creature who is aware of the relevant law practices."[102] Macrophysical properties such as hardness and brittleness are determined by microphysical facts involving the arrangement of a substance's molecules. That determination relationship can be brute: it can be a fact about the universe that this or that arrangement of molecules grounds this or that macrophysical property even if it were opaque to us why *this* arrangement determines *that* property. Law, Greenberg argues, is different. "That the law practices support *these* legal propositions over all others is always a matter of *reasons*—where reasons are considerations in principle intelligible to rational creatures."[103]

Greenberg emphasizes that the rational-relation doctrine does not itself resolve the debate between positivism and antipositivism: "it is an open

[101] See generally Lord and Maguire, eds., *Weighing Reasons*; and Dancy, *Ethics without Principles*, ch. 3. The example best known to legal scholars is Raz's "exclusionary reasons" (*Practical Reason and Norms*, 35–48).

[102] See Greenberg, "How Facts Make Law," 237.

[103] Greenberg, "How Facts Make Law," 237. As he further explains, "lawyers believe that when they get [the law] right, the reasons they discover are not merely reasons for believing that the content of the law is a particular way, but the reasons that *make* the content of the law what it is. . . . Lawyers take for granted that the epistemology of law tracks its metaphysics. And the epistemology of law is plainly reason-based" (239).

question whether there are non-normative, non-evaluative facts that could constitute reasons for legal facts—and indeed whether there are value facts that could do so."[104] I agree. But he is driven to antipositivism because, he believes, "it turns out that value facts are needed to make *intelligible* that law practices support certain legal propositions over others."[105] That I deny. I see no reason to anticipate that determination of legal facts by aggregation of principles grounded in practice leaves an intelligibility deficit.[106] Rather, the rational-relation doctrine itself—understood as an aspect of law's nature—strongly favors some mappings over others. The more complex a mapping, the greater it threatens the ability of participants in legal practice to reason from the contributory to the all-in. Because no mechanism or mapping is more intuitive or intelligible than simple aggregation, we might expect it to be the default mode in a complex, comprehensive, and decentralized legal system. It is no surprise that simple additive weighting is widely heralded as the most user-friendly and "robust" of MCA models.[107]

Second and notwithstanding, to describe simple aggregation as the likely default in a mature, complex, and decentralized legal system is not to deny that such a system could incorporate other mappings. I suspect that they can and do. What determines the particulars of a mapping is the same broad type of practice facts that ground the principles themselves. That is, the taking-up behaviors of participants ground not only the fundamental principles of a legal system but also the "meta-principles" that bear on their interaction. Or, to shift terminology, helping to establish the particular mapping of principles to rules that obtains in a given legal system is one possible function of what Andrei Marmor calls "deep conventions."[108] For example, if a "meta-principle" or "deep convention" were to arise in *S* to the effect that there is a uniquely right legal answer to (almost) all legal questions, that would have a bearing on how principles in *S* accrue: it would exert pressure toward mappings that facilitate more determinate rules and against mappings that would yield greater indeterminacy. This is why figure 5 depicts practices as playing a role in the determination of not only fundamental legal principles but also the determination function that maps such principles to derivative legal rules.

[104] Greenberg, "How Facts Make Law," 233.

[105] Greenberg, "How Facts Make Law," 240.

[106] Here I am in broad agreement with Chilovi and Pavlakos, "The Explanatory Demands of Grounding in Law." I interpret Greenberg as arguing for explanation in their "weak sense," and I share their judgment that positivism can supply it.

[107] See, e.g., Lindell, *Multi-criteria Analysis in Legal Reasoning*, 47.

[108] See generally Marmor, *Social Conventions*, ch. 3.

These practices, moreover, are responsive to ordinary human needs and interests. As a thought experiment, suppose that legal system *S* begins life with only a single determinant at the fundamental legal level—that is, a single determinant that is directly grounded in practices: [for all *p*, *p* is a rule of *S* if the constitutional text says *p* (or if *p* is entailed by what the text says)].[109] It is exceedingly unlikely that a mature or complex legal system will recognize only a single legal factor. This is because some legal rules that arise by application of a single factor will prove unacceptable to most judges (or they will be unacceptable to many citizens, and judges change their practices in response to social unrest or dissatisfaction when it exceeds a certain level). Suppose, for example, that what the text says yields legal rules such as [states are permitted to racially segregate the public schools], [states are permitted to establish official churches], or [the federal government lacks power to regulate sources of air pollution]. Discomfort with such outcomes can be sufficiently broad and intense to cause judges to recognize and accept additional factors. The system will evolve from recognizing a single factor to recognizing a plurality of factors, such as, for purposes of illustration: [what the text originally meant], [what the text means to an ordinary contemporary reader], [what the authors of the text intended to do or accomplish], [what our stable practices have been], [what the courts have held], [what justice requires], etc.

If this is right, the next question concerns what will be the character or mode of the function that maps the plurality of factors to decisive legal norms in a system that has, in virtue of the speech acts of the relevant legal actors, established a plurality of fundamental legal determinants. The standard view among legal positivists, following Hart (or their reading of Hart), is that the plurality of grounds are necessarily arrayed into a lexical ordering, which can be represented as a complex if-then statement.[110] I draw attention to the alternative possibility that the factors are weighted and determine derivative legal norms by aggregate force, akin to the way that simple additive weighting is understood to underwrite or recommend a decision. No doubt the mix that emerges in any legal system is contingent on a great many variables—size and heterogeneity of the population, responsiveness of the legal system to the pop-

[109] Cf. Hart, *The Concept of Law*, 100–1.

[110] Some orthodox positivists might object that this reading of Hart is a misreading and that his notion of "validation" does not presuppose what I have called lexical determination. I address this objection elsewhere, noting that many theorists are skeptical that nonlexical determination is workable and that if Hart means to embrace it, neither he nor his followers address those concerns. See Berman, "Dworkin versus Hart Revisited," 576–77. In any event, as noted earlier (notes 22 and 23 and accompanying text), I am more interested here in the state of jurisprudential thinking than in Hart exegesis.

ulace, age of the system, scope of the system's regulatory reach, amenability of the central legal instruments to prompt purposive change, and so forth. You can speculate as well as I about what practices are likely to emerge under what conditions.

But one advantage of the nonlexical model warrants emphasis: it demands less coordination among the participants whose behaviors ground the determination. Lexical determination requires that any condition sufficient to confer legal status must enjoy clear majority endorsement or acceptance, else two contradictory rules could both be valid law. Were acceptance by a (substantial) minority of judges sufficient to ground the rule that p is the law if C_1, and acceptance by a different (substantial) minority sufficient to ground the rule that q is the law if C_2, then p and q would both be the law if C_1 and C_2 jointly obtain, even if p are q are mutually incompatible. That would be untenable. Nonlexical determination by weighted principles can deliver law when practices are less uniform. If a minority of judges take up and thus ground principle P_1, and a different minority of judges take up and thus ground a conflicting or inconsistent principle P_2, the consequence is only that they might cancel each other out in a given case, each rendering the other constitutively inert. The conflicting principles would *not* thereby determine conflicting normative verdicts, as would be true of lexical determination.[111] This is important because it shows that it's no happy accident that principled positivism can address *both* Dworkinian challenges to Hart's version. While opening positivism to nonlexical determination directly addresses Dworkin's challenge from principles, that adjustment at the same time permits a relaxation of the demand that the fundamental legal materials enjoy supermajority official support, which is a precondition to meeting the challenge of too little law.

At this point, it seems to me we have all the rudiments of a positivist account of legal content adequate to meet Dworkin's two challenges. Fundamental norms are grounded in speech acts of legal actors. These norms gain rough variable weights in essentially the same way that they gain their contents. Weighted norms can determine the legal status of tokens by simple weighted aggregation or by more complicated interactions, as the nature of legal systems and the meta-principles or deep conventions of the system collectively determine. Rules reflect or capture a describable set of tokens that share legal status. Is this a complete account? No. Does detail remain to be filled in? Sure. But that is true of every extant theory of legal content.[112] The present task is

[111] I discuss conflicts between principles at greater length in Berman, "Religious Liberty and the Constitution," 889–94.

[112] Greenberg acknowledges that his own affirmative antipositivist constitutive theory ("the moral impact theory") depends upon a not yet developed account of "the legally proper

not to try to prove out principled positivism in a single article but to make it a plausible and promising candidate, worthy of attention by jurisprudents and other metanormative philosophers.

Scholars attuned to this account will find plenty of judicial support for it. Elsewhere, I show that principled positivism makes sense of many and significant constitutional decisions by the US Supreme Court, favored by liberals and conservatives alike.[113] But the account is not particular to the US legal system. A revealing recent example from Britain is the unanimous opinion of the UK Supreme Court holding that Prime Minister Boris Johnson's advice to the Queen to prorogue Parliament was legally invalid, rendering the purported prorogation a nullity.[114] That conclusion rested on two planks. First, "the United Kingdom ... possesses a Constitution, established over the course of our history by common law, statutes, conventions and practice," and that Constitution "includes numerous principles of law, which are enforceable by the courts in the same way as other legal principles."[115] Second, "the boundaries of a prerogative power relating to the operation of Parliament are likely to be illuminated, *and indeed determined*, by the fundamental principles of our constitutional law."[116] The view, in short, is that the fundamental legal principles are embedded in legal practice, and they combine or interact to determine legal rules. The Court could then ascertain what the rule governing prorogation is once it identified what the UK's fundamental constitutional principles are. To be sure, the Court's analysis was controversial.[117] But the surface conformity

way" that legal institutions act to change "the moral profile." See Greenberg, "The Moral Impact Theory of Law," 1323.

[113] Berman, "Our Principled Constitution"; Berman and Peters, "Kennedy's Legacy"; and Berman, "Religious Liberty and the Constitution."

[114] *R (Miller) v. the Prime Minister*, [2019] UKSC 41.

[115] *R (Miller) v. the Prime Minister*, par. 39.

[116] *R (Miller) v. the Prime Minister*, par. 38 (emphasis added).

[117] See, on the one hand, e.g., Craig, "The Supreme Court, Prorogation and Constitutional Principle" (finding the decision "correct and compelling"); Twomey, "Article 9 of the Bill of Rights 1688 and Its Application to Prorogation" (averring "the Court has taken an approach consistent with its previous jurisprudence ... and has not altered its course for political or any other reasons"); Young, "Deftly Guarding the Constitution" (describing the decision as "a carefully reasoned judgment, respectful of the constitutional and institutional limits of the judiciary, which protects the foundations of our constitution including representative democracy") (internal quotation marks omitted); Konstadinides, O'Meara, and Sallustio, "The UK Supreme Court's Judgment in *Miller/Cherry*" (approving of the decision as "grounded in classic constitutional and legal principles"); Caird, "The Politics of Constitutional Interpretation in the UK" (dismissing criticisms that the ruling was "improper" and noting "that all exercises of constitutional interpretation, when undertaken by a constitutional actor, are political"); Grogan, "The Rule of Law, Not the Rule of

of that analysis with the central elements of principled positivism can lend support to that theory of legal content even if the Court got this dispute wrong.[118]

2.4. Of Promulgated Rules and Summary Rules

The preceding analysis explains how principles aggregate to ground legal rules via their power to determine, nonlexically, the legal status of act and event tokens. You might worry that this gets things backwards, that the legal property or status that a token act or event possesses should be a function or *consequence* of the applicable legal rule, if there is one, not a determinant or *input* to the applicable legal rule. I address that concern here by distinguishing two kinds of rule: what I call "summary" (or "resultant") rules and "promulgated" (or "contributory") rules.

A summary rule reflects the actual normative state of affairs. The preceding subsections explain its emergence. A promulgated rule, in contrast, is an effort to change the normative state. To a first approximation, the promulgated rule is what is said or asserted in a statute. Resultant rules are summaries of the aggregate impact of principles, whereas promulgated rules are among—possibly chief among—the facts upon which principles operate.

Take a statute in legal system *S* that asserts that "*q* is prohibited." This assertion acquires normative force from underlying principles that are activated by or give effect to communicative contents of statutes. If the only fundamental legal principle in *S* provides that legal norms are all and only what authoritative

Politics" (deeming the decision "clearly follow[ing] from principle" and the judgment's criticisms "unfounded"); Sedley, "In Court" (celebrating the decision and claiming that the Court "has re-lit one of the lamps of the United Kingdom's constitution: that nobody, not even the Crown's ministers, is above the law"). See in contrast, e.g., Endicott, "Making Constitutional Principles into Law," 177–78 (arguing that the Supreme Court was wrong "to decide when Parliament must be in session" because "the fact that Parliament should meet as appropriate does not support the conclusion that the law requires it to meet as appropriate"); Finnis, "The Unconstitutionality of the Supreme Court's Prorogation Judgment" (describing the judgment as "undercut[ting] the genuine sovereignty of Parliament," "wholly unjustified by law," and "a historic mistake, not a victory for fundamental principle"); Fisher, "No Politics Please," 144–45 (claiming that the Supreme Court referenced "inadequate" justifications in *Miller II* to "procure [...] obliquely an effect which could be achieved *directly* only by open departure from prior authority"); and Tierney, "Turning Political Principles into Legal Rules" (ascribing a "political view" to the decision "that led to the identification first of a constitutional principle and then the creation of a legal rule that served to normativise this principle even to the point of constraining a prerogative of sovereignty").

[118] For an example from a civil law country, see the 2018 decision from France's Constitutional Council holding that the principle of *fraternité* barred prosecution under a statute making it a crime to help migrants entering the country illegally. *Conseil Constitutionnel*, decision no. 2018-717/718 QPC, July 6, 2018.

legal texts assert, then (conflicting assertions aside), it would be a derivative legal rule in *S* that *q* is prohibited. There would be no daylight between the promulgated rule and the summary rule, in which case our inclination to treat the promulgated rule as *the* rule (unmodified) would be vindicated.

In complex mature legal systems, however, fundamental norms are plural and (very likely) weighted. Almost certainly, fundamental principles will provide that communicative contents of statutory texts have great legal force. (The text will be among the "legally relevant phenomena" that, as figure 5 represents, combine with the principles to determine derivative legal facts.) Thus, and again, the status of tokens will be substantially shaped by the promulgated rules. But because other principles are in play, it might not be the case that every token's status is what the promulgated rule directs, in which case the summary rule will depart, if only a little, from the promulgated one. This is why summary (resultant) rules closely track but are not identical to promulgated (contributory) ones.

3. PRINCIPLED POSITIVISM AT WORK

This section turns to concrete legal disputes. It aims to advance understanding of principled positivism by illustrating how it can explain legal content, even in disputed cases, and to better reveal some of the account's relative merits. Section 3.1 discusses the US Supreme Court's decision in *TVA v. Hill*, the "snail darter case" that we encountered in section 1.3, in connection with Dworkin's ill-fated challenge from theoretical disagreements. I will show that principled positivism makes the disagreements in that case perfectly intelligible. Section 3.2 turns to the Supreme Court's same-sex marriage decision, *Obergefell v. Hodges*, a textbook casualty of Dworkin's too-little-law challenge. Here I show that principled positivism can deliver law where Hartian positivism cannot.

3.1. Snail Darters Revisited: Explaining Theoretical Disagreements

The federal Endangered Species Act of 1973 (ESA) is one of the nation's signature environmental protection statutes. It directs the secretary of the interior to identify threatened species and their critical habitats and imposes extensive public and private obligations and prohibitions that such designations trigger. Section 7 provides that all federal departments and agencies shall "tak[e] such action necessary to insure that actions authorized, funded, or carried out by them do not jeopardize the continued existence of such endangered species and threatened species or result in the destruction" of such habitats.[119]

[119] 16 U.S.C. § 1536.

In 1967, The Tennessee Valley Authority (TVA), a federally owned corporation, had started constructing a dam on the Little Tennessee River to generate hydroelectric power and to promote regional economic development. Six years in, scientists discovered in the river a previously unknown species of perch, the snail darter. In 1975, two years after the act's enactment and eight years after construction of the Tellico Dam had commenced, the secretary of the interior listed the snail darter as endangered and the Little Tennessee as its critical habitat. The issue was thus posed: does the ESA require that construction on the dam cease when nearing completion, after public expenditures of nearly $80 million?

In *TVA v. Hill*, a divided Supreme Court held that it does. As discussed earlier (in section 1.3), that decision serves in *Law's Empire* as a central recurring example designed to cause trouble for positivism and to furnish support for Dworkin's own competing antipositivist theory, "law as integrity." The thrust is that the disagreement between Chief Justice Warren Burger's majority opinion and Justice Lewis Powell's principal dissent (joined by Justice Harry Blackmun) is inexplicable on positivist premises but makes perfect sense if viewed through Dworkin's competing theory of law.

I argued earlier that Hartian positivists can explain the disagreement. Because Hart's theory does not require that the participants whose behaviors constitute the rule of recognition understand its workings, both Burger and Powell could have been genuinely unaware that neither side's "theory of legislation" could be legally correct given its rejection by the other. But that does not mean that the challenge is entirely inert. Even if Hart's account does not require that judges understand how his system works and even though knowledge cannot be attributed to them on purely semantic bases, one might nonetheless think that if, as the Hartian theory maintains, derivative legal rules are validated by criteria grounded in judicial near-consensus, many sophisticated participants, including Supreme Court justices, would ferret that out. So theoretical disagreements of the sort that supposedly mark *Hill* are somewhat surprising and disconcerting, even if possible.

Principled positivism can explain these disagreements better than Hartian positivism can. To see how, we need a fuller understanding of the opinions than Dworkin's abbreviated summary conveys. Burger did not quite adopt what Dworkin called "the excessively weak version" of intentionalism in statutory interpretation, pursuant to which judges are obligated to follow clear "acontextual" statutory meaning unless "the legislature actually intended the opposite result."[120] And Powell did not quite reason that courts must avoid an absurd

[120] Dworkin, *Law's Empire*, 22.

result unless it is clear that the legislature intended it. Instead, both opinions recognize the same three principles as existing in our legal system and as at least potentially bearing on the legal status of the token act. These principles concern communicative contents of the statute, legal and application intentions of the enacting legislature, and the public good (as an ordinary person or legislature would view it).[121] Because principles lack canonical formulation, these, like all, can be rendered in diverse ways. But here's a first try: *what the statutory text means matters*; *legal intentions of the enacting legislature have force*; *absurd results should be avoided.* Perhaps the justices disagree about these principles' relative weights. More conspicuously and consequentially, however, they disagree about the extent to which each principle was activated.

Let us take the principles one at a time. The justices' disagreement over the meaning of section 7 is straightforward. As the majority saw things, "the explicit provisions of the Endangered Species Act require precisely [that dam construction cease]. One would be hard pressed to find a statutory provision whose terms were any plainer."[122] Powell thought otherwise. Agreeing with the majority that "the starting point in statutory construction" is the statutory text, he found the language "far from 'plain.'"[123] His thought (expressed somewhat obscurely) appears to be that section 7 would more clearly direct the result the majority ruled that it did if it explicitly enjoined federal agencies to take action "necessary to insure that actions authorized, funded, carried out, *or completed* by them do not jeopardize" endangered species or their habitats. But that is not what the section says. Therefore, it "can be viewed as a textbook example of fuzzy language, which can be read according to 'the eye of the beholder.'"[124]

Now turn to the justices' views about congressional intent. This is more subtle and requires unpacking. Recall that the ultimate issue in a litigated case is particular, not general; it concerns tokens, not types. In this case, the issue was whether the ESA required cessation of the Tellico Dam project. What content would congressional intent need to have to underwrite an affirmative answer? Consider three possibilities, in order of increasing generality. Congress might have intended that section 7 would apply (1) even to the Tellico Dam project, (2) even to projects that are close to completion at the time that the secretary of the interior lists a species as endangered or its habitat as critical, or (3) even when its application would incur great immediate or localized costs. All

[121] For introductions to differences among types of intention—semantic, communicative, legal, application—see Berman, "The Tragedy of Justice Scalia," 796–99.

[122] *TVA v. Hill*, 437 U.S. at 173.

[123] *TVA v. Hill*, 437 U.S. at 205 (Powell, J., dissenting).

[124] *TVA v. Hill*, 437 U.S. at 202 (Powell, J., dissenting).

members of the Court agreed that the Congress that enacted the ESA lacked any intention with content 1 or 2.[125] At the same time, the majority insisted, and the dissent did not deny, that the enacting Congress did have intention 3.[126] What divided the majority and dissent was whether intention 3 entailed or encompassed intention 1.

Burger thought that it did because intention 1 plainly falls within intention 2, and 2 does not differ in any material way from other subclasses of cases that fall under 3. Powell thought that the slide from 3 to 2 (and thereby to 1) is more fraught than the majority recognizes.[127] Nearly completed projects comprise a subclass of cases captured by 3, but one with distinctive features not shared by all subclasses of 3, namely that the costliness and thus *potential* absurdity of abandoning nearly completed projects is manifest. What should the government do in such cases? Spend additional funds to undo what it has already done? Leave a nearly completed but unusable dam standing, as a constant reminder to the community of the costs it has already sustained for promised benefits that will never materialize?[128] Because abandoning nearly completed projects might reasonably strike citizens and their representatives as more foolish or costly than not starting them, notwithstanding the economic logic that renders "sunk-cost" reasoning fallacious, congressional intent 3 does not entail congressional intent 2 and therefore does not entail congressional intent 1. It followed, according to Powell, that there was no actual congressional intention relevant to this dispute—no intention either that completion of the Tellico Dam project would be illegal or that it would not be.[129]

[125] See *TVA v. Hill*, 437 U.S. at 207–8 (Powell, J., dissenting).

[126] See, e.g., *TVA v. Hill*, 437 U.S.: "The dominant theme pervading all Congressional discussion of the proposed [Endangered Species Act of 1973] was the overriding need to devote whatever effort and resources were necessary to avoid further diminution of national and worldwide wildlife resources" (at 177, citation omitted).

[127] See *TVA v. Hill*, 437 U.S., criticizing the majority for "nowhere mak[ing] clear how the result it reaches can be 'abundantly' self-evident from the legislative history when the result was never discussed" (at 207, Powell, J., dissenting).

[128] See *TVA v. Hill*, 437 U.S.: "Few members of Congress will wish to defend an interpretation of the Act that requires the waste of at least $53 million … and denies the people of the Tennessee Valley area the benefits of the reservoir that Congress intended to confer. There will be little sentiment to leave this dam standing before an empty reservoir, serving no purpose other than a conversation piece for incredulous tourists" (at 210, Powell, J., dissenting).

[129] Powell actually sends conflicting signals on just this point. Much of his analysis aims to establish that Congress lacked an actual intention that the act would "apply to completed or substantially completed projects." *TVA v. Hill*, 437 U.S. at 196 (Powell, J., dissenting). But some language suggests the stronger conclusion that Congress possessed an actual intention that the Act *not* apply to such projects. See, e.g., *TVA v. Hill*, identifying "strong

So much for the opinions' disagreements regarding the first two principles or considerations: statutory plain meaning and the legislature's legal intention. What about the third, *avoid absurdity* (or *comport with common sense*)? Having concluded that the weightiest considerations do not clearly resolve this dispute—they do not activate as forcefully against the dam's completion as the majority believed—Powell embraced *avoid absurdity* enthusiastically. While acknowledging this principle's subordinacy to the first two, Powell nonetheless found it greatly activated.[130]

The majority is more circumspect, not surprisingly. Having determined that the most important principles pressed forcefully and in concert against permissibility, it did not need to examine the possible import of a palpably less weighty principle. Still, the majority opinion intimates that *avoid absurdity* would have some force in a dispute with respect to which meaning and intent were more equivocal.[131]

In sum, here is how the dispute looks through a principled positivist lens. Burger believed that the "meaning" of the statute and the enacting Congress's legal intent are both pellucid and that both direct that dam construction must cease. Whether or not this result would flout common sense, the *avoid absurdity* principle could not possibly overcome the combined force of the textualist and intentionalist principles. Powell believed that the statutory meaning was

corroborative evidence that the interpretation of § 7 as not applying to completed or substantially completed projects reflects the initial legislative intent" (at 210). I think that the former and weaker proposition better accords with Powell's opinion as a whole. Note, for example, his conclusion that "I had not thought it to be the province of this Court to force Congress into otherwise unnecessary action by interpreting a statute to produce a result no one intended" (at 210–11). Had he really endorsed the more aggressive position regarding congressional intent, this passage should have read "… to force Congress to produce a result contrary to what it intended."

[130] *TVA v. Hill*, 437 U.S., arguing that "where the statutory language and legislative history … need not be construed to reach [a result that disserves the public interest], I view it as the duty of this Court to adopt a permissible construction that accords with some modicum of common sense and the public weal" (at 196, Powell, J., dissenting).

[131] This too is modestly ambiguous. Burger's opinion can be read to suggest that *avoid absurdity* is a subordinate principle of our legal system that can have effect when the actual legal intention of the enacting legislature is uncertain. See *TVA v. Hill*, 437 U.S., observing that "Congress has spoken in the plainest of words, making it abundantly clear that the balance has been struck in favor of affording endangered species the highest of priorities," and asserting that judicial "appraisal of the wisdom or unwisdom of a particular course consciously selected by the Congress is to be put aside in the process of interpreting a statute" (at 194). Or it could be read to deny that it is a principle of our legal system at all: "in our constitutional system the commitment to the separation of powers is too fundamental for us to pre-empt congressional action by judicially decreeing what accords with 'common sense and the public weal'" (at 195).

much less clear than Burger did and that Congress did not actually intend the legal results that Burger claimed. At the same time, he thought, *avoid absurdity* pressed very strongly in the other direction. Because the principles that militated against the legal permissibility of completing the dam did so with much less aggregative force than the majority believed, the principle that militated forcefully in favor of the permissibility of project completion could carry the day. Figures 8, 9, and 10 represent these competing positions, cleaned up a bit.

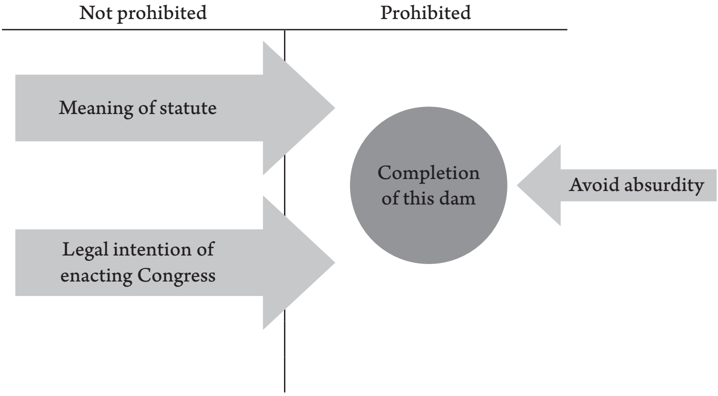


FIGURE 8 *TVA v. Hill*, per the Majority

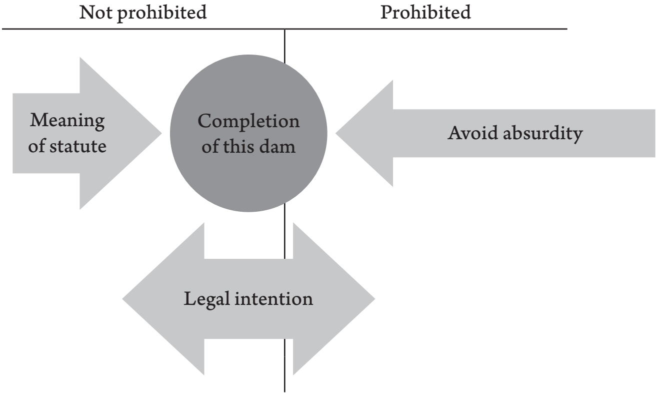
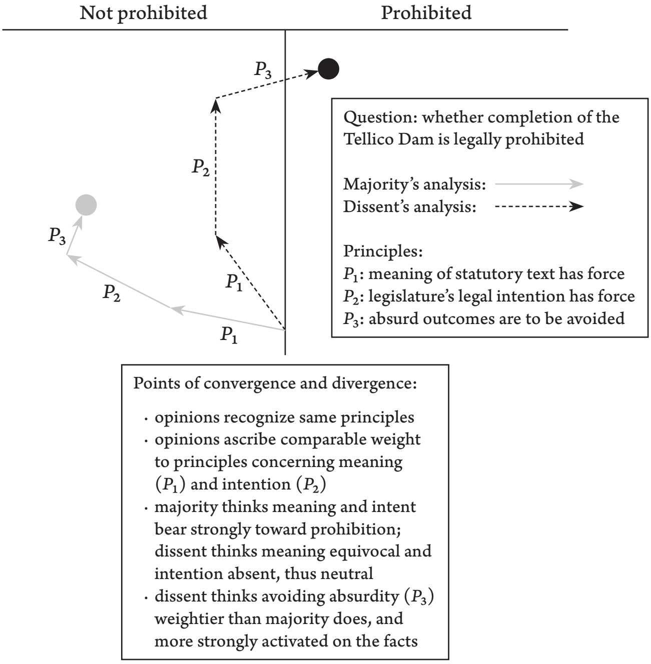


FIGURE 9 *TVA v. Hill*, per the Dissent

3.2. *Same-Sex Marriage before* Obergefell*: Delivering More Law*

Consider lastly whether states are constitutionally required to recognize same-sex marriages on the same terms as they recognize opposite-sex marriages. Call the affirmative proposition *same-sex marriage*. The Supreme Court took up the


Not prohibited
Prohibited
P_3
P_2
P_1
P_3
P_2
P_1
Question: whether completion of the Tellico Dam is legally prohibited
Majority's analysis:
Dissent's analysis:
Principles:
P_1: meaning of statutory text has force
P_2: legislature's legal intention has force
P_3: absurd outcomes are to be avoided

Points of convergence and divergence:

- opinions recognize same principles
- opinions ascribe comparable weight to principles concerning meaning (P_1) and intention (P_2)
- majority thinks meaning and intent bear strongly toward prohibition; dissent thinks meaning equivocal and intention absent, thus neutral
- dissent thinks avoiding absurdity (P_3) weightier than majority does, and more strongly activated on the facts

FIGURE 10 *TVA v. Hill*, Both Opinions

question in 2015 in *Obergefell v. Hodges*.[132] When it did, many people believed that the Court should rule for the plaintiffs on the (minimally realist) ground that *same-sex marriage* was already true (though not authoritatively *declared* to be true). Was it? Was this a compelling claim or even a plausible one?[133]

Recall my earlier contention in section 1.2 that Hartian validation depends upon satisfaction of any (complex) criterion that concordant acceptance picks out as sufficient. As it operates in Hart's account (and putting defeasibility aside), q is a norm of legal system S if C_1 or C_2 or C_3 or … C_n, where each condition C can itself be a complex combination of conjuncts and disjuncts and is grounded in the practices that make out the rule of recognition of S.[134]

[132] *Obergefell v. Hodges*, 576 U.S. 644 (2015).

[133] This section draws from Berman, "Our Principled Constitution," 1406–8; and Berman and Peters, "Kennedy's Legacy," 366–68. Readers of those earlier efforts will notice that the diagrams I use here to represent the bearing of principles on the legal status of act or event tokens differ from the ones used in those earlier articles. I previously explained that the two representations are interchangeable (Berman, "Our Principled Constitution," 1394n219) and have come now to believe that the diagrams in this paper are preferable on balance.

[134] For an argument that these criteria need not refer only to matters of "pedigree" rather than content, see Berman, "Dworkin versus Hart Revisited," 572–74.

An orthodox Hartian sympathetic to *same-sex marriage* even prior to its endorsement in *Obergefell* might reason along the following lines: *q* is a legal norm in the US if:

C_1: [the Supreme Court has held *q* in a nonoverruled decision]

or

C_2: [*q* is the plain original meaning of a provision of the constitutional text, and no decision of the Supreme Court (not itself overruled) holds or clearly says $\neg q$]

or

C_3: [the authors and ratifiers of the constitutional text intended to codify *q*, the nation has observed a consistent practice of respecting *q*, and both *q* and $\neg q$ are comparably compatible with the ordinary meaning of the constitutional text and with all (nonoverruled) Supreme Court holdings]

or

C_4: [*q* is required by a posture of equal respect for human dignity, and *q* is not clearly contradicted by any (nonoverruled) Supreme Court decision]

or

C_5: [*q* best promotes human flourishing and is not contradicted by the contemporary naive meaning of any provision of the constitutional text]

or

$\ldots C_n$

The problem for any Hartian who believes that the ruling in *Obergefell* was legally correct (and that a contrary ruling would have been legally incorrect) is that the sufficient conditions that plausibly are supported or recognized by a convergent consensus among judges—conditions such as C_1, C_2, and C_3—do not plausibly validate *same-sex marriage*, while conditions that do plausibly validate *same-sex marriage*—conditions such as C_4 and C_5—are pretty clearly not the object of a judicial consensus.[135] Of course, it could be that before *Obergefell*

[135] This exercise suggests why the Hartian rule of recognition is better understood as picking out sufficient conditions (subject to vagueness and defeasibility) rather than conditions that are both necessary and sufficient. (See note 47 above.) Even were it plausible that a

was decided, *same-sex marriage* was false as an account of existing law. On the orthodox Hartian account, however, *same-sex marriage* is not merely false but *obviously* false, a nonstarter. And many sophisticated observers will find that conclusion highly doubtful.[136] Principled positivism would earn a feather for its cap if it could make *same-sex marriage* plausible, even if not demonstrably correct.

The first step is to identify the fundamental legal principles that might bear on this legal issue. This is lawyers' work. But the very considerations that a Hartian American constitutional lawyer thinks figure somehow into internally complex validity criteria will often strike a principled positivist as independent fundamental legal principles. Such principles will give legal force to: original and current communicative contents of the ratified text; legal intentions of authors and ratifiers; judicial decisions; federalism; stable and accepted political practices; and moral principles concerning equality, liberty, respect for human dignity, and so forth. These principles obtain not because they are accepted by all or nearly all judges but because they have the type of "institutional support" to which Sartorius and Ten already drew our attention—they are "embedded in or exemplified by numerous authoritative legal enactments: constitutional provisions, statutes, and particular judicial decisions."[137]

To get a flavor for how principles embed in legal materials and practice, consider the legal principle *respect human dignity*. In his *Obergefell* dissent, Justice Thomas diagnosed "the flaw" in the majority's reasoning as being "of course, . . . that the Constitution contains no 'dignity' Clause."[138] True, it does not. But fundamental principles are extratextual, and the dignity principle that Justice Kennedy's majority opinion rested upon was well embedded in our constitutional law by the time *Obergefell* rolled around. Kennedy himself had relied heavily upon the principle in a handful of majority opinions that vindicated claimed constitutional rights of gay and lesbian people.[139] But as Leslie Meltzer Henry has shown, the principle (or, as she argues, a cluster of relatively distinct dignity-based principles that share a family resemblance) has been taken up in several hundreds of Supreme Court decisions over many decades and across

judicial consensus has picked out some criteria as sufficient, there is patently no consensus among American judges that those criteria are the *only* sufficient ones.

[136] Do not be misled by this one example: principled positivism and organic pluralism are not partisan. I have shown elsewhere that they support many conservative results. See Berman, "Our Principled Constitution," 1393–411, and "Religious Liberty and the Constitution."

[137] Sartorius, "Social Policy and Judicial Legislation," 154–55.

[138] *Obergefell v. Hodges*, 576 U.S. at 735 (Thomas, J., dissenting).

[139] See *United States v. Windsor*, 570 U.S. 744 (2013) at 770–75; *Lawrence v. Texas*, 539 U.S. 558 (2003) at 574–76; and *Romer v. Evans*, 517 U.S. 620 (1996).

the doctrinal waterfront.[140] It has undergirded successful claims to freedom of expression and personal liberty and to protection from excessive punishment, unreasonable searches, compelled self-incrimination, discrimination on the basis of race or sex, and more.[141] As Sartorius emphasized, "a fundamental test for law defined in terms of such notions as coherence and institutional support obviously goes well beyond reporting concordant judicial practice."[142]

In short, let us suppose the American legal system comprises many principles that bear on *same-sex marriage*, either for or against. If the principles came with finely individuated weights, it might be both true and reasonably discoverable that their net force weighed for (or against) *same-sex marriage*. But in our real world, the skeptic thinks, a model of rules constituted by the cumulative impact of many weighted principles delivers essentially the same underdeterminacy as does the established Hartian model in which rules are validated by a single master rule.

Yet this is precisely the skeptical conclusion that close attention to the distinct attributes of weight and activation (section 2.2) aims to dispel. In particular, constitutional principles concerning the pursuit of happiness and concerning the state's obligation to respect the inherent equal dignity of all persons within its jurisdiction (which principles include or lie adjacent to principles of anti-subordination) are activated very substantially in favor of *same-sex marriage*: the ability to enter into the legal institution of marriage with one's life partner is of tremendous instrumental value; and the exclusion of same-sex couples from this important and highly salient legal institution significantly demeans, degrades, and insults gay, lesbian, and bisexual people. At the same time, none of the principles that plausibly weigh against *same-sex marriage* activate very substantially. The constitutional text does not clearly state that states are free

[140] Henry, "The Jurisprudence of Dignity."

[141] *Cohen v. California*, 403 U.S. 15 (1971) at 24 (robust freedom of expression rooted in "the premise of individual dignity and choice upon which our political system rests"); *Planned Parenthood v. Casey*, 505 U.S. 833 (1992) at 851 ("choices central to personal dignity and autonomy are central to the liberty protected by the Fourteenth Amendment"); *Roper v. Simmons*, 543 U.S. 551 (2005) at 560 (the Eighth Amendment's ban on cruel and unusual punishment "reaffirms the duty of the government to respect the dignity of all persons"); *Rochin v. California*, 342 U.S. 165 (1952) at 174 (the Fourth Amendment proscribes unreasonable searches and seizures because they are "offensive to human dignity"); *Miranda v. Arizona*, 384 U.S. 436 (1966) at 460 (the Fifth Amendment's privilege against self-incrimination is founded on "the respect a government … must accord to the dignity and integrity of its citizens"); *Rice v. Cayetano*, 528 U.S. 495 (2000) at 517 ("race is treated as a forbidden classification [because] it demeans the dignity and worth of a person to be judged by ancestry"); *Roberts v. United States Jaycees*, 468 U.S. 609 (1984) at 625 (sex discrimination is forbidden because it "deprives persons of their individual dignity").

[142] Sartorius, *Individual Conduct and Social Norms*, 207.

to disregard same-sex unions; nobody who played an important role in drafting or ratifying portions of the constitutional text did so with an actual legal intention to authorize states to withhold recognition from same-sex unions; the most on-point judicial precedent was a one-sentence summary dismissal (entitled to little weight on standard case law principles); and so on.[143] If this is approximately correct, the net force of constitutional principles grounded in institutional practice metaphysically determined *same-sex marriage* even before *Obergefell* was decided. (See figure 11.)

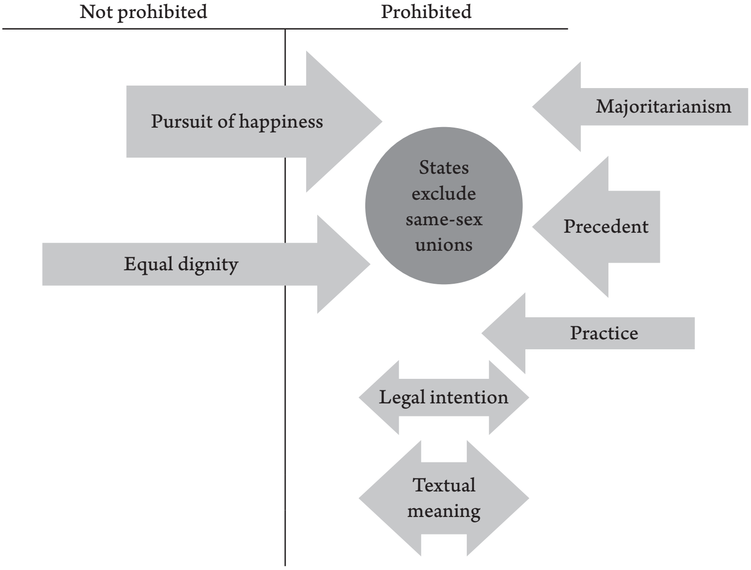


FIGURE 11 *Obergefell*, per Principled Positivism

I do not claim that this brief discussion and accompanying diagram are nearly sufficient to establish fully that *same-sex marriage* was a derivative legal rule of American constitutional law even before *Obergefell* so held. That is a lengthy task—and one for first-order constitutional scholarship, not legal philosophy. Rather, by explaining how that plausibly could be, I demonstrate how principled positivism differs from and likely improves upon Hartian positivism with respect to the challenge of too little law.[144] The example can thus serve as

[143] *Baker v. Nelson*, 409 U.S. 810 (1972).

[144] Admittedly, even if one is persuaded that a model of determination by net vector force yields a legally determinate rule in *this* dispute, while the orthodox Hartian model does not, that still would not establish that it yields more determinacy all things considered; some disputes that appear determinate on the Hartian account might become underdeterminate through the principled positivist lens. This is not something we can net out *a priori*. Still, two points merit emphasis. First (see section 2.3 above), I do not rule out that

proof of concept even for those who disagree with the constitutional bottom line it endorses.

Thirty-five years ago, the American constitutional theorist Richard Fallon focused attention on what he dubbed the "commensurability problem": the fact that American constitutional practice recognizes a variety of kinds of argument—arguments based on meanings of the text, framers' intentions, historical practices, values, and so forth—but lacks an agreed upon means of reconciling them "in a single, coherent constitutional calculus."[145] His proposed solution to the problem had two parts. First, judges should "assess and reassess the arguments in the various categories in an effort to understand each of the relevant factors as prescribing the same result."[146] Second, if attempts to massage or strongarm the diverse constitutional arguments into "constructive coherence" fails, judges should rank the arguments hierarchically and reach the judgment that accords with "the highest ranked factor clearly requiring an outcome."[147] Before elaborating and defending his own solution, however, Fallon flagged what he thought a surprising gap in the literature: the absence of any "powerfully argued balancing theory" that would deliver unique results from discordant factors or principles without lexical ordering.[148] Without favoring such approaches, he nonetheless thought they clearly merited more attention than scholars had paid.[149]

Now, principled positivism is not *exactly* what Fallon was looking for. Fallon presented his commensurability problem as a problem in American constitutional law, not in general jurisprudence, and the theories he contemplated—the "constructivist coherence theory" that he advocated as well as the alternative "balancing theory" that he only imagined—are proposed solutions to that problem. Even more significantly, Fallon sought a "methodology" that judges could follow when engaged in constitutional interpretation, whereas principled positivism is a theory of legal content, not a theory about how anybody ought to

the system includes lexical arrangements as well. My account, albeit hardly simple, surely simplifies a yet more complex reality. Second, by far the best way to get a good grasp of the workings, virtues, vices, and plausibility of this competing account is to investigate a large variety of actual and hypothetical legal disputes with an insider's knowledge and perspective. I attempt some of that elsewhere (Berman, "Our Principled Constitution"; and Berman and Peters, "Kennedy's Legacy") but do not pretend that my efforts to date are conclusive. Thanks to Ruth Chang for pressing me on this point.

[145] Fallon, "A Constructivist Coherence Theory of Constitutional Interpretation," 1190.

[146] Fallon, "A Constructivist Coherence Theory of Constitutional Interpretation," 1193.

[147] Fallon, "A Constructivist Coherence Theory of Constitutional Interpretation," 1193–94.

[148] Fallon, "A Constructivist Coherence Theory of Constitutional Interpretation," 1228.

[149] Fallon, "A Constructivist Coherence Theory of Constitutional Interpretation," 1229–30.

do anything at all. Because these are theories about different things, principled positivism, as such, cannot quite fill Fallon's bill.[150] That acknowledged, one would expect there to be a road to travel from general jurisprudential theories of legal content to jurisdiction-specific theories of proper judicial reasoning, and the preceding discussion suggests that the road from principled positivism to a theory of how US judges should reason in constitutional cases will be reasonably direct. Principled positivism is thus a general theory of legal content that, if sound, supplies the jurisprudential substrate for the "balancing theory" of American constitutional law that we have sorely lacked.

4. CONCLUSION

What makes it the case that the law has the content that it does? Insofar as Hartian positivism addresses this question at all, it holds that norms are "validated" as legal by satisfying sufficient criteria that are picked out by, thus grounded in, a convergent practice among legal officials that Hart termed the "ultimate rule of recognition." Principled positivism maintains, in contrast, that decisive and derivative legal norms ("rules") are (also) determined by the accrual or aggregation of fundamental weighted norms (what Dworkin called "principles") that are grounded in their being "taken up" by legal practitioners in legal decision-making.

Nomenclature aside, the critical differences are two. First, principled positivism allows, as the Hartian theory of legal content denies, that the social-factual grounds of fundamental legal norms ("principles" in one case, "criteria of sufficiency" in the other) can be unspecifiable and characterized by nontrivial dissensus. Second, principled positivism provides that principles "bear on" derivative norms in a weighted and aggregative fashion that cannot be fully captured by the language and machinery of validation. These two differences might strike some persons as modest. They are not. As this article shows, their payoffs are great, for they combine to defang the two most forceful objections that Dworkin leveled against Hart's own account—that it cannot make sense of the existence and functions of legal principles and that it cannot determine nearly as much law as legal sophisticates believe there to be. If this alternative to the Hartian theory of legal content is closer to correct, it makes a profound

[150] See Berman, "Our Principled Constitution," 1328–32 (distinguishing "prescriptive" from "constitutive" theories of constitutional interpretation); Sachs, "Originalism" (distinguishing "decision procedures" from "standards"); and Berman, "Keeping Our Distinctions Straight" (comparing the two sets of distinctions).

difference—not only to legal philosophers but to all who would understand or ascertain our law.[151]

University of Pennsylvania
mitchberman@law.upenn.edu

REFERENCES

Abdullah, Lazim, and C. W. Rabiatul Adawiyah. "Simple Additive Weighting Methods of Multi-criteria Decision-Making and Applications: A Decade Review." *International Journal of Information Processing and Management* 5, no. 1 (2014): 39–49.

Adler, Matthew, and Kenneth Einar Himma, eds. *The Rule of Recognition and the U.S. Constitution*. Oxford: Oxford University Press, 2009.

Alexander, Larry. "The Banality of Legal Reasoning." *Notre Dame Law Review* 73, no. 3 (1998): 517–34.

Alexander, Larry, and Ken Kress. "Against Legal Principles." *Iowa Law Review* 82, no. 3 (1997): 739–68.

———. "Replies to Our Critics." *Iowa Law Review* 82, no. 3 (1997): 923–55.

Alexy, Robert. "Formal Principles: Some Replies to Critics." *International Journal of Constitutional Law* 12, no. 3 (July 2014): 511–24.

Ávila, Humberto. *Theory of Legal Principles*. Dordrecht: Springer, 2007.

Barzun, Charles L. "The Positive U-Turn." *Stanford Law Review* 69, no. 6 (2017): 1323–88.

Baude, William, and Stephen E. Sachs. "Grounding Originalism." *Northwestern University Law Review* 113 (2019): 1455–91.

Bayles, Michael D. *Hart's Legal Philosophy: An Examination*. Dordrecht: Kluwer

[151] For very helpful conversations and comments on predecessors to this paper, I am indebted to Tom Adams, Larry Alexander, Hrafn Asgeirsson, Emad Atiq, Brian Bix, Ruth Chang, Samuele Chilovi, Hanoch Dagan, John Deigh, Mike Dorf, Oran Doyle, Ben Eidelson, David Enoch, Dick Fallon, John Golden, Mark Greenberg, Alex Guerrero, Brian Hutler, Felipe Jiménez, Jeffrey Kaplan, Guha Krishnamurthi, Brian Leiter, Errol Lord, Andrei Marmor, Marcin Matczak, Stephen Perry, David Plunkett, Joseph Raz, Connie Rosati, Gideon Rosen, Steve Sachs, Larry Sager, Fred Schauer, Steve Schaus, Stefan Sciaraffa, Scott Shapiro, Rebecca Stone, Kevin Toh, Mark van Roojen, Daniel Wodak, two anonymous referees for this journal, and workshop audiences at the Oxford Seminars in Jurisprudence, Cornell Law, University College London, UCLA Law, Penn Philosophy, and the Surrey Centre for Law and Philosophy. I started work on this paper while serving as a Laurance S. Rockefeller Fellow at the Princeton University Center for Human Values and gratefully acknowledge the center's support.

Academic Publishers, 1992.

Berker, Selim. “The Unity of Grounding.” *Mind* 127, no. 507 (July 2018): 729–77.

Berman, Mitchell N. “Dworkin versus Hart Revisited: The Challenge of Non-lexical Determination.” *Oxford Journal of Legal Studies* 42, no. 2 (Summer 2022): 548–77.

———. “For Legal Principles.” In *Moral Puzzles and Legal Perplexities: Essays on the Influence of Larry Alexander*, edited by Heidi M. Hurd, 241–59. Cambridge: Cambridge University Press, 2019.

———. “Keeping Our Distinctions Straight: A Response to Originalism—Standard and Procedure.” *Harvard Law Review Forum* 135 (2022): 133–50.

———. “Of Law and Other Artificial Normative Systems.” In *Dimensions of Normativity*, edited by David Plunkett, Scott Shapiro, and Kevin Toh, 137–44. Oxford: Oxford University Press, 2019.

———. “Our Principled Constitution.” *University of Pennsylvania Law Review* 166 (2018): 1325–420.

———. “Religious Liberty and the Constitution: Of Rules and Principles, Fixity and Change.” *University of Pennsylvania Journal of Constitutional Law* 26, no. 4 (2024): 851–928.

———. “The Tragedy of Justice Scalia.” *Michigan Law Review* 115 (2017): 783–808.

Berman, Mitchell N., and David Peters. “Kennedy’s Legacy: A Principled Justice.” *Hastings Constitutional Law Quarterly* 46 (2019): 311–84.

Bicchieri, Cristina. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press, 2006.

Bix, Brian H. “Global Error and Legal Truth.” *Oxford Journal of Legal Studies* 29, no. 3 (Autumn 2009): 535–47.

Brennan, Geoffrey, Lina Eriksson, Robert E. Goodin, and Nicholas Southwood. *Explaining Norms*. Oxford: Oxford University Press, 2013.

Caird, Jack Simson. “The Politics of Constitutional Interpretation in the UK.” Policy Exchange: Judicial Power Project, October 1, 2019. https://judicialpowerproject.org.uk/jack-simson-caird-bingham-centre-for-the-rule-of-law-the-politics-of-constitutional-interpretation-in-the-uk/.

Chilovi, Samuele, and George Pavlakos. “The Explanatory Demands of Grounding in Law.” *Pacific Philosophical Quarterly* 103, no. 4 (December 2022): 900–33.

———. “Law-Determination as Grounding: A Common Grounding Framework for Jurisprudence.” *Legal Theory* 25, no. 1 (March 2019): 53–76.

Craig, Paul. “The Supreme Court, Prorogation and Constitutional Principle.” *Public Law* (2020): 248–54.

Dancy, Jonathan. *Ethics without Principles*. Oxford: Clarendon Press, 2004.

Davies, Stephen. "The Cluster Theory of Art." *British Journal of Aesthetics* 44, no. 3 (July 2004): 297–300.

Dworkin, Ronald. "Hart's Posthumous Reply." *Harvard Law Review* 130, no. 8 (June 2017): 2096–140.

———. *Law's Empire*. Cambridge, MA: Belknap Press, 1986.

———. "The Model of Rules." *University of Chicago Law Review* 35, no. 1 (1967): 14–46.

———. *Taking Rights Seriously*. Cambridge, MA: Harvard University Press, 1978.

Endicott, Timothy. "Are There Any Rules?" *Journal of Ethics* 5, no. 3 (September 2001): 199–219.

———. "Making Constitutional Principles into Law." *Law Quarterly Review* 136 (2020): 175–81.

Enoch, David. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press, 2011.

Fallon, Richard H., Jr. "A Constructivist Coherence Theory of Constitutional Interpretation." *Harvard Law Review* 100, no. 6 (1987): 1189–286.

Finlay, Stephen. "Defining Normativity." In *Dimensions of Normativity*, edited by David Plunkett, Scott Shapiro, and Kevin Toh, 187–219. Oxford: Oxford University Press, 2019.

Finnis, John. "The Unconstitutionality of the Supreme Court's Prorogation Judgment." London: Policy Exchange, 2019.

Fisher, James C. "'No Politics Please, We're British': *R (Miller) v the Prime Minister; Cherry and Others v. Advocate General for Scotland* [2019] UKSC 41." *Hibernian Law Journal* 19 (2020): 140–60.

Foot, Philippa. "Morality as a System of Hypothetical Imperatives." *Philosophical Review* 81, no. 3 (July 1972): 305–16.

Gardner, John. "Legal Positivism: 5½ Myths." In *Law as a Leap of Faith: Essays on Law in General*, 19–44. Oxford: Oxford University Press, 2012.

Goodwin, Paul, and George Wright. *Decision Analysis for Management Judgment*. 4th ed. Chichester: Wiley, 2009.

Greenawalt, Kent. "The Rule of Recognition and the Constitution." *Michigan Law Review* 85 (1987): 621–71.

———. "The Rule of Recognition and the Constitution." In Adler and Himma, *The Rule of Recognition and the U.S. Constitution*, 33–66.

Greenberg, Mark. "How Facts Make Law." *Legal Theory* 10, no. 3 (September 2004): 157–98.

———. "How Facts Make Law." In *Exploring Law's Empire*, edited by Scott Hershovitz, 225–64. Oxford: Oxford University Press, 2006.

———. "The Moral Impact Theory of Law." *Yale Law Journal* 123 (2014):

1288–323.

———. “Response: What Makes a Method of Legal Interpretation Correct? Legal Standards vs. Fundamental Determinants.” *Harvard Law Review Forum* 130 (2017): 105–24.

Grogan, Joelle. “The Rule of Law, Not the Rule of Politics: Commentary on the *Cherry/Miller No. 2* Judgment.” *Verfassungsblog* (blog), October 1, 2019. https://verfassungsblog.de/the-rule-of-law-not-the-rule-of-politics/.

Hart, H. L. A. *The Concept of Law*. 2nd ed. Oxford: Clarendon Press, 1994.

———. *Essays in Jurisprudence and Philosophy*. Oxford: Clarendon Press, 1983.

Henry, Leslie Meltzer. “The Jurisprudence of Dignity.” *University of Pennsylvania Law Review* 160 (2011): 169–233.

Himma, Kenneth Einar. “Understanding the Relationship between the US Constitution and the Conventional Rule of Recognition.” In Adler and Himma, *The Rule of Recognition and the U.S. Constitution*, 95–121.

Hurley, S. L. “Coherence, Hypothetical Cases, and Precedent.” *Oxford Journal of Legal Studies* 10, no. 2 (July 1990): 221–51.

Konstadinides, Theodore, Noreen O’Meara, and Riccardo Sallustio. “The UK Supreme Court’s Judgment in *Miller/Cherry*: Reflections on Its Context and Implications.” UK Constitutional Law Association blog, October 2, 2019. https://ukconstitutionallaw.org/2019/10/02/theodore-konstadinides-noreen-omeara-and-riccardo-sallustio-the-uk-supreme-courts-judgment-in-miller-cherry-reflections-on-its-context-and-implications/.

Kramer, Matthew H. *H. L. A. Hart: The Nature of Law*. Cambridge: Polity Press, 2018.

Kress, Ken. “Coherence.” In *A Companion to Philosophy of Law and Legal Theory*, 2nd ed., edited by Dennis Patterson, 521–38. Oxford: Wiley-Blackwell, 2010.

Lamond, Grant. “The Rule of Recognition and the Foundations of a Legal System.” In *Reading H. L. A. Hart’s “The Concept of Law,”* edited by Luís Duarte d’Almeida, James Edwards, and Andrea Dolcetti, 97–122. Oxford: Hart Publishing, 2013.

Lawson, Gray. “A Farewell to Principles.” *Iowa Law Review* 82 (1997): 893–903.

Leiter, Brian. “Beyond the Hart–Dworkin Debate: The Methodology Problem in Jurisprudence.” *American Journal of Jurisprudence* 48, no. 1 (2003): 17–51.

———. “Explaining Theoretical Disagreement.” *University of Chicago Law Review* 75, no. 3 (2009): 1215–45.

———. “Explanation and Legal Theory.” *Iowa Law Review* 82 (1997): 905–9.

Lindell, Bengt. *Multi-criteria Analysis in Legal Reasoning*. Cheltenham: Edward Elgar Publishing, 2017.

Lord, Errol, and Barry Maguire. “An Opinionated Guide to the Weight of Reasons.” In *Weighing Reasons*, edited by Errol Lord and Barry Maguire. Oxford:

Oxford University Press, 2016.

Lord, Errol, and Barry Maguire, eds. *Weighing Reasons*. Oxford: Oxford University Press, 2016.

Lyons, David. “Principles, Positivism, and Legal Theory.” *Yale Law Journal* 87, no. 3 (1977): 415–47.

Margolis, Eric, and Stephen Laurence. “Concepts.” *Stanford Encyclopedia of Philosophy* (Winter 2019). https://plato.stanford.edu/archives/sum2019/entries/concepts/.

Marmor, Andrei. *Social Conventions: From Language to Law*. Princeton: Princeton University Press, 2009.

Perry, Stephen R. “Judicial Obligation, Precedent and the Common Law.” *Oxford Journal of Legal Studies* 7, no. 2 (Summer 1987): 215–57.

———. “Second-Order Reasons, Uncertainty and Legal Theory.” *Southern California Law Review* 62, no. 3–4 (1989): 913–94.

———. “Two Models of Legal Principles.” *Iowa Law Review* 82, no. 3 (1997): 787–819.

Plunkett, David, and Scott Shapiro. “Law, Morality, and Everything Else: General Jurisprudence as a Branch of Metanormative Inquiry.” *Ethics* 128, no. 1 (October 2017): 37–68.

Postema, Gerald J. “Classical Common Law Jurisprudence (Part I).” *Oxford University Commonwealth Law Journal* 2 (2002): 155–66.

Rawls, John. *A Theory of Justice*. Cambridge, MA: Belknap Press of Harvard University Press, 1971.

Raz, Joseph. “Legal Principles and the Limits of Law.” *Yale Law Journal* 81, no. 5 (1972): 823–54.

———. *Practical Reason and Norms*. 2nd ed. Princeton: Princeton University Press, 1990.

Rodriguez-Blanco, Veronica. “A Revision of the Constitutive and Epistemic Coherence Theories in Law.” *Ratio Juris* 14, no. 2 (June 2001): 212–32.

Rosen, Gideon. “Metaphysical Dependence: Grounding and Reduction.” In *Modality: Metaphysics, Logic, and Epistemology*, edited by Bob Hale and Aviv Hoffmann, 109–35. Oxford: Oxford University Press, 2010.

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

Sachs, Stephen E. “The ‘Constitution in Exile’ as a Problem for Legal Theory.” *Notre Dame Law Review* 89, no. 5 (2014): 2253–98.

———. “Originalism: Standard and Procedure.” *Harvard Law Review* 135, no. 1 (2022): 777–830.

Sartorius, Rolf. *Individual Conduct and Social Norms: A Utilitarian Account of Social Union and the Rule of Law*. Encino, CA: Dickenson Publishing Company, 1975.

———. “Social Policy and Judicial Legislation.” *American Philosophical Quarterly* 8, no. 2 (April 1971): 151–60.

Schaffer, Jonathan. “On What Grounds What.” In *Metametaphysics: New Essays on the Foundations of Ontology*, edited by David Chalmers, David Manley, and Ryan Wasserman, 347–83. Oxford: Oxford University Press, 2009.

Schauer, Frederick. “Amending the Presuppositions of a Constitution.” In *Responding to Imperfection: The Theory and Practice of Constitutional Amendment*, edited by Sanford Levinson, 145–61. Princeton: Princeton University Press, 1995.

Searle, John R. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press, 1969.

Sedley, Stephen. “In Court.” *London Review of Books*, October 10, 2019.

Shapiro, Scott J. “The Hart–Dworkin Debate: A Short Guide for the Perplexed.” In *Ronald Dworkin*, edited by Arthur Ripstein, 22–55. Cambridge: Cambridge University Press, 2007.

———. *Legality*. Cambridge, MA: Harvard University Press, 2011.

Smith, Dale. “Dworkin’s Theory of Law.” *Philosophy Compass* 2, no. 2 (March 2007): 267–75.

Soper, E. Philip. “Legal Theory and the Obligation of a Judge: The Hart/Dworkin Debate.” *Michigan Law Review* 75, no. 3 (1977): 473–519.

Stavropoulos, Nicos. “The Debate that Never Was.” *Harvard Law Review* 130, no. 8 (2017): 2082–95.

Ten, C. L. “The Soundest Theory of Law.” *Mind* 88, no. 352 (1979): 522–37.

Tierney, Stephen. “Turning Political Principles into Legal Rules: The Unconvincing Alchemy of the *Miller/Cherry* Decision.” Policy Exchange: Judicial Power Project, September 30, 2019. https://policyexchange.org.uk/blogs/stephen-tierney-turning-political-principles-into-legal-rules-the-unconvincing-alchemy-of-the-miller-cherry-decision/.

Toh, Kevin. “Hart’s Expressivism and His Benthamite Project.” *Legal Theory* 11, no. 2 (June 2005): 75–123.

———. “Jurisprudential Theories and First-Order Legal Judgments.” *Philosophy Compass* 8, no. 5 (May 2013): 457–71.

Tripkovic, Bosko, and Dennis Patterson. “The Promise and Limits of Grounding in Law.” *Legal Theory* 29, no. 3 (September 2023): 202–28.

Twomey, Anne. “Article 9 of the Bill of Rights 1688 and Its Application to Prorogation.” *UK Constitutional Law Association*, October 4, 2019. https://ukconstitutionallaw.org/2019/10/04/anne-twomey-article-9-of-the-bill-of-rights-1688-and-its-application-to-prorogation/.

Van Roojen, Mark. *Metaethics: A Contemporary Introduction*. New York: Routledge, 2015.

Waldron, Jeremy. “Who Needs Rules of Recognition?” In Adler and Himma, *The Rule of Recognition and the U.S. Constitution*, 327–49.

Watson, Bill. “Explaining Legal Agreement.” *Jurisprudence* 14, no. 2 (2023): 221–53.

———. “In Defense of the Standard Picture: What the Standard Picture Explains that the Moral Impact Theory Cannot.” *Legal Theory* 28, no. 1 (March 2022): 59–88.

Wodak, Daniel. “What Does ‘Legal Obligation’ Mean?” *Pacific Philosophical Quarterly* 99, no. 4 (December 2018): 790–816.

Young, Alison. “Deftly Guarding the Constitution.” Policy Exchange: Judicial Power Project, September 29, 2019. https://judicialpowerproject.org.uk/alison-young-deftly-guarding-the-constitution/.

Journal of Ethics and Social Philosophy
vol. 28, no. 3 · october 2024

https://doi.org/10.26556/jesp.v28i3.2927

SULKING INTO SEX

BLAME, COERCION, AND CONSENT

Sumeet Patwardhan

Sometimes, people sulk when their partners refuse sex. For instance, they might angrily pout, initiate the silent treatment, or manifest some other form of conspicuous, blame-laden withdrawal. To avoid this sulking, those on the receiving end sometimes submit to sex that they do not want.[1] Consider, for example, the following narrative from an online forum:

> *Cuddle*: I wanted a cuddle, and I told [my husband] that's all I wanted. He got frisky and started pushing it. This is not uncommon, since becoming parents I've often just let him go ahead because he sulks and I'm too tired, it's easier just to let him have his three minutes, and then I get some peace and he's happy.... This weekend I might have even felt like it if only he'd started with a bit of nice chatting and cuddling. But he went straight for the finishing line, as per usual. Then he got in a sulky, "victim" mood, rolled over, and refused to cuddle. And accused me of seeing somebody else!... Underneath it I think he just feels like I ought to do it whether I want to or not. And that is making me angry. Very angry. And very, very tired.[2]

The sulking that this woman faces is clearly morally problematic. So is the sex induced by that sulking. Indeed, both the sulking and the resulting sex seem to wrong her; she is owed rectification—perhaps just an apology, perhaps more. But what explains this intuition? It cannot be that all sexual pressures are wrongful, let alone wrongful in the exact same way. After all, sexual pressures are diverse: violence, disappointed sighs, the prospect of divorce, seductive flirting, peer group opinions, family expectations for children, economic

[1] For data on the prevalence of nonphysical sexual pressures, see Smith et al., "The National Intimate Partner and Sexual Violence Survey," 2–3, 15–16. I have not found data specific to sulking, but the number of online stories of sulking into sex indicates that it is common.

[2] Anonymous, "Being Made to Feel Bad." Netmums Forum, March 20, 2017, https://www.netmums.com/coffeehouse/drop-clinic-984/domestic-abuse-41/1636124-being-made-feel-bad-about-no-sex.html.

incentives, and the list goes on. Accordingly, let me rephrase the question: How can we explain the intuition about cases like Cuddle without overgeneralizing?

The first aim of this article is to answer this question. I start by arguing that even attempting to blamingly sulk someone into sex—blamingly sulking *for* sex—often imposes *wrongful blame*. Next, I argue that succeeding at blamingly sulking someone *into* sex often undermines their consent via coercion. This imposes the further wrong of *nonconsensual sex*.[3]

Both arguments cut against the current literature. Sarah Conly and Alan Wertheimer claim that pressures like sulking do not wrong a victim.[4] Conly, Wertheimer, and Kimberly Kessler Ferzan all claim that pressures like sulking are not consent undermining.[5] And Robin Morgan seems to claim that *all* sexual pressures undermine consent—a position avoided by my arguments.[6] Still, my arguments usefully extend to other sexual pressures that involve blame, demonstrating the continuity between subtle pressures like sulking and more overt pressures like threats of violence. They even extend to sulking within nonsexual interactions. I thereby offer a novel, striking explanation of the wrongfulness of blamingly sulking for and into sex—an explanation that generalizes without overgeneralizing.

The second aim of this article is to bring out three broader lessons for the literature on consent and coercion. To start, if we disregard the nuances of different sexual pressures—especially subtle pressures like sulking—we risk overlooking key moral features of those pressures. We run the same risk, moreover, if we ignore how such pressures unfold within the unique dynamic of close relationships. This risk increases if we consider only hypothetical, "cleaned-up" cases rather than first-person testimonies. In sum, the relative abstraction of contemporary discussions of consent and coercion has led scholars to neglect the wrongfulness of subtle sexual pressures. For this reason, I focus on blame-laden sulking within close relationships, leaving robust discussion of other sexual pressures for other papers within my broader research program. For the same reason, I draw heavily from real stories.

The article proceeds as follows. In section 1, I characterize sulking for and into sex. Typical cases such as Cuddle are prolonged, pervasive, habitual,

[3] Some philosophers, like David Archard in "The Wrong of Rape," define "rape" as "nonconsensual sex." Others, like Ann J. Cahill in *Rethinking Rape*, do not. I need not take a position here, so I avoid the term.

[4] Conly, "Seduction, Rape, and Coercion," 114–15; and Wertheimer, *Consent to Sexual Relations*, 183.

[5] Conly, "Seduction, Rape, and Coercion," 114–15 and 119; Wertheimer, *Consent to Sexual Relations*, 183; and Ferzan, "Consent and Coercion," 954–56, 971–80, 994–95, 1002–7.

[6] Morgan, "Theory and Practice," 165.

situated within a close relationship, and laden with blame. In section 2, I argue that blamingly sulking at someone for sex often wrongs them. It imposes numerous harms to pressure them to respond to the blame, even though they have done nothing morally wrong in the first place. In section 3, I first articulate some sufficient conditions for consent-undermining coercion. Next, I show that they are often satisfied in cases of blame-laden sulking into sex. In section 4, I examine some implications. I discuss the nature and gravity of nonconsensual sulking into sex, explain why it should not always be criminalized, and describe why the framework of consent is useful. I then extend my argument to myriad sexual and nonsexual pressures.

1. Sulking for and into Sex

> *Cycle*: [My boyfriend] came upstairs with me and started undressing me, but I let him know I was tired. He got pouty and pouty [sic] and left. He gets pouty and sulks any time I say no.... The last time we had sex was 7 days ago. It's not like months are passing.... It makes me feel even less excited about having sex, because I'm nervous about whether I'll WANT to have sex. So it's a vicious cycle. I feel nervous, like I have to want sex.... [I] feel like shit for not wanting to have sex.[7]

This narrative from Reddit illustrates three key features of sulking.[8] First, sulking is a triadic relation between a sulker, a sulkee, and a frustrated goal of the sulker—here, the girlfriend's having sex. By sulking, the sulker communicates to the sulkee that they want them to provide one or more of the following forms of support: to resolve the sulker's frustrated goal; to distract them from it; or to comfort them about it. In Cycle, the boyfriend wants the girlfriend to resolve his frustrated goal by having sex with him. To communicate this desire, he sulks.

This leads us to the second key feature of sulking: the peculiar way it communicates a desire for support. Unlike ways of seeking support that orient towards the supportive person—like crying on their shoulder—sulking involves withdrawal. But because sulkers seek support, they must remain within the scope of the sulkee's attention. To achieve this peculiar "proximate withdrawal," sulkers employ conspicuously limited verbalization, offering curt responses or pointed

[7] u/fakepalindrome_ (username), "Boyfriend Gets Pouty if I Don't Want to Have Sex." Reddit, September 8, 2014, https://www.reddit.com/r/relationships/comments/2fuhgm/boyfriend_gets_pouty_if_i_dont_want_to_have_sex/.

[8] This section draws from a similar account of sulking offered by psychologists Anita Barbee and Michael Cunningham. See Barbee and Cunningham, "An Experimental Approach to Social Support Communications," 393–95, 407.

silence. As in Cycle, sulkers often employ nonverbal forms of withdrawal as well: angry sighing; defiant body language; pouting or frowning; flat affect; manifest focus away from the sulkee; physical movement away from the sulkee; or reluctance to socialize. A sulker's use of withdrawal, I suspect, is one source of resistance to viewing sulking as coercive. In our popular imagination, coercion involves "approach" behaviors; this article resists that picture.

The affective core of sulking—its third key feature—is anger, rather than anxiety or sadness. Because of this, sulking frequently involves (un)conscious blame. It is no coincidence that the girlfriend in Cycle feels "like she has to want sex," feels "like shit for not wanting to." She feels guilty for saying no due to her boyfriend's sulky blame. Such blame is often difficult to challenge. Since sulking involves limited verbalization, sulkees frequently lack an explicit rebuke to challenge.[9] Even when the sulker does issue a rebuke, they will often prevent challenges, e.g., through silence. By preventing challenges, the sulker can avoid admitting their distress and thereby save face. Indeed, entertaining challenges would draw the sulker into precisely the engagement they seek to avoid: a conversation.

As a final observation, note that even though sulkers are often self-aware, they can certainly sulk unknowingly.

I can now formulate an account of sulking for and into sex. Someone (*A*) sulks someone else (*B*) *into* sex just in case:

1. *A*, knowingly or not, sulks at *B for* sex. That is:
   a. *A* proximately withdraws from *B* verbally and perhaps also emotionally, mentally, physically, and/or socially;
   b. primarily because *A* is angry about a frustrated goal;
   c. at least in part to communicate to *B* that *A* wants support for that frustrated goal;
   d. where the support *A* wants includes sex with *B*.[10]
2. *B* agrees to sex with *A*, at least in part because of 1, and *A* and *B* have sex.

While the above conditions are *necessary* features of sulking into sex, cases of sulking into sex also have four *characteristic* features, which I will discuss below. I will focus mostly on cases that have these features, in order to attend to sulking in its most typical form.

[9] Miceli, "How to Make Someone Feel Guilty," 96.

[10] Sometimes, the sulker does not want the sulkee to agree to sex in the moment; they want the sulkee to agree to their next sexual advance. For simplicity's sake, I do not focus on such cases, but my arguments easily extend to them.

To begin, a sulker is typically in a close relationship with the sulkee. People tend not to want support from strangers or acquaintances. Even when they do want it, they often do not pursue it because strangers will likely refuse or fail to support them. Even when they do pursue it, they tend to be more verbal, to avoid being misinterpreted. Hence, sulking in general, including sulking for and into sex, is far rarer between strangers or acquaintances.

Second, sulking for and into sex is typically blame laden: the sulker blamingly sulks at the sulkee for not having sex. It is certainly possible for someone to sulk for sex without blaming the sulkee. But because sulking is almost always embedded within close relationships and because feelings of sexual entitlement can easily arise within close sexual relationships, sulking for and into sex tends to involve blame for sexual refusal.[11]

Third, a sulkee usually recognizes when a sulker is blaming them for something—even if they do not always recognize what for. In some cases, this is because a sulker clearly communicates the blame or the views motivating it. As one sulkee has described, "[My husband] thinks its [sic] his right to have sex at least once a day but would like it twice a day."[12] In other cases, sulkees may recognize the blame on their own: "I knew he was mad. . . . In his mind he's the victim and always has been. . . . I'm the bad guy."[13] Such recognition is not surprising. As Victoria McGeer observes, we social creatures are disposed to pick up on others' attitudes towards us.[14] Hence, recognition of blame is common across blame-laden forms of sulking, including but not limited to blame-laden sulking for and into sex.

Fourth, sulking into sex is typically prolonged, pervasive, and habitual. Proximate withdrawal aims to make interpersonal engagement with the sulker contingent on the sulkee's support. If this withdrawal were brief, the sulkee would not be incentivized to submit to sex. Accordingly, when a sulkee does submit, the sulking tends to be prolonged.[15] The difficulty of challenging sulky blame, as observed earlier, is another reason that blame-laden sulking into sex

[11] For an argument that connects blame even more closely to withdrawal behaviors, see Bennett, "The Varieties of Retributive Experience," 149–52.

[12] Jo B(1113) (username), "Different Sex Drives May Lead to Separation." Netmums Forum, March 16, 2020, https://www.netmums.com/coffeehouse/life-504/family-other-relationships-50/1893925-different-sex-drives-may-lead-separation.html.

[13] Anonymous poster, "Boyfriend is Playing the Victim." Reddit, June 12, 2021, https://www.reddit.com/r/vaginismus/comments/nyg2qy/boyfriend_is_playing_the_victim/ (post since deleted).

[14] McGeer, "Civilizing Blame," 181–82.

[15] For some stories of particularly lengthy sulking, see "Emotional Abuse in Sulking Silence when Sexual Demands Go Begging"; and Sugar and Mitchell, "Sulking for Sex."

tends to be prolonged. Sulking into sex also tends not to stay compartmentalized. Instead, it pervades different parts of life. This is because "getting in a mood"—cooking in a sulk, going on a walk in a sulk, etc.—can give a sulkee powerful incentive to submit. Finally, like other strategies of seeking support, sulking is often habitual. As one sulkee recounts, "Things will be okay for a while, but then he reverts to the same behaviour."[16] Sulking into sex therefore frequently involves a kind of prolonged, pervasive, and habitual detachment that is anathema to us social creatures.

In sum, I will focus mostly on typical cases of sulking for and into sex—cases in which a sulker blamingly sulks at someone close to them for not having sex; the sulkee knows that the sulker is blaming them; and the sulking is pervasive, prolonged, and habitual.

Sulking into sex does manifest another typical feature worth mentioning. As you might notice, in almost all the real stories I discuss, a man sulks at a woman. This is no accident. Gender and patriarchy influence the prevalence of heterosexual relationships; the frequency at which different people feel entitled to (sulk for) sex; the costs that different people incur upon refusing sex; and more. My arguments, however, will not concentrate exclusively on cases of men sulking women into sex. This is because I aim to offer a more general account of the wrongs of sulking for and into sex—an account that can illuminate how cases of sulking that do not involve a man sulking at a woman can still be wrongful.[17]

Having elucidated some necessary and typical features of sulking for and into sex, let me emphasize: this elucidation is valuable independent of my later arguments. This is because it provides a foundation for further examining the nature and ethics of sulking. Indeed, this is one major reason that I focus specifically on sulking. To my knowledge, philosophers have said little to nothing about this peculiar behavior involving saying little to nothing—and yet there is so much to say.

[16] McDermott, "My Partner Wants Sex Every Night and Sulks if I Don't Agree."

[17] For examples of such cases, see u/Sam_Fort (username), "BF Sulks if I Don't Give Him Sex Every Night." Reddit, July 28, 2021, https://www.reddit.com/r/relationships/comments/otarsg/bf_sulks_if_i_dont_give_him_sex_every_night/; Price, "A Few Words about Sexual Coercion in the Wake of the Aziz Ansari Accusations"; and McDermott, "My Girlfriend Sulks if We Don't Have Sex and It's Bringing Back Painful Memories." For data on sexual victimization perpetrated by women, see Stemple, Flores, and Meyer, "Sexual Victimization," 303.

2. BLAMINGLY SULKING FOR SEX

> *Guilty*: I have been married for 12 years. . . . We met when I was 19 and carefree. We had sex multiple times a day. Since then life got crazy, and my sex drive went down. At a minimum we make love once a week. Our max at the moment is probably 4 times. I literally reject him 10 times a day/night. Not because I'm nasty but because I'm bloody tired! I work full time in child protection for DHS. It's a stressful role, plus 3 kids, a house etc. He will sulk and complain for hours after I say no. I'm just so over it. I'm ready to walk away because I'm sick of the guilt![18]

Guilty, from an anonymous user of an online forum, is a typical case of sulking for sex. But is it a typical case of sulking *into* sex? Does the husband's sulking ever get his wife to submit? I do not know, but to deem that he has wronged her, we do not need to know—or so I will argue in this section. That is, I will argue that even attempting to blamingly sulk someone into sex—blamingly sulking *for* sex—often wrongs them. That argument is built on three premises, as follows:

P1. In many cases of blamingly sulking for sex:
  a. the sulker and sulkee are in a close relationship;
  b. the sulker blamingly sulks at the sulkee for not having sex with them;
  c. the sulkee recognizes that they are being blamed; and
  d. the sulkee's not having sex with the sulker is not wrong.

P2. To "misdirectedly blame" someone is to blame them for something that is not wrong.

P3. Misdirectedly blaming someone who is close to the blamer and who recognizes that they are being blamed often wrongs them.

C1. In many cases of blamingly sulking for sex, the sulker wrongs the sulkee.

I have already defended the first three parts of P1 and will defend the fourth shortly. Afterwards, I will defend P3 at more length. Unlike these substantive

[18] Anonymous poster, "Seriously Considering Leaving My Husband Due to the Sulking Because of What He Believes to Be 'Not Enough Sex,'" Stay at Home Mum, 2016, https://www.stayathomemum.com.au/ask-sahm/question/734fee81-6367-48ea-9097-a4e21c59ab24/.

premises, P2 simply defines "misdirected blame." "Well-directed blame," in contrast, is blame directed towards something that is wrong.[19]

Some philosophers have discussed views that seem to conflict with P1d. For example, Scott Anderson argues that people can create sexual obligations by *promising*, say, to have sex after the kids are asleep.[20] Richard Hull, moreover, explores whether sex that minimally harms one person but greatly benefits another is required by *beneficence*—though he does not take a stand.[21] Finally, Alan Soble suggests that people can have *distributive* obligations, say, to reciprocate sexual pleasure.[22]

In many sexual interactions, however, these views fail to apply. Take Guilty. The wife has not promised to have sex at her husband's desired frequency. Nor does her refusal fall afoul of beneficence or distributive justice, given that she is exhausted and stressed. Hence, even if the views above are all true, we can still affirm P1d. In many cases of blamingly sulking for sex, refusing sex is not wrong.

P3 states that blame that is both misdirected and recognized—though it need not be recognized *as* misdirected—often wrongs a blamee close to the blamer. Other forms of blame, such as misdirectedly blaming a stranger, might also be wrongful. But for reasons discussed in section 1, such cases are not my focus. Additionally, P3 is neutral about what blame involves: a judgment, emotion, desire, intention, functional role, etc.[23] Having clarified P3, I can now defend it.

2.1. Wrongful Misdirected Blame

Targets of misdirected blame face seven characteristic, interconnected harms. Blame—whether misdirected or not—usually involves the blamer *negatively morally assessing* and *directing negative emotions towards* the blamee. Relatedly, blame that is recognized often causes the blamee to *morally criticize themselves* and *feel negative emotions* like guilt.[24] We care about how close relations view and feel about us and about how we view and feel about ourselves, so recognized

[19] If blame for suberogatory acts should also count as well-directed, my arguments can be extended to show that sexual refusal is rarely if ever suberogatory.

[20] Anderson, "On Sexual Obligation and Sexual Autonomy," 123–32. Contrast Liberto, "The Problem with Sexual Promises," 394–403.

[21] Hull, "Have We a Duty to Give Sexual Pleasure to Others?" 10–11.

[22] Soble, *Sexual Investigations*, 53–58. See also Wertheimer, *Consent to Sexual Relations*, 258–76. Contrast Srinivasan, "Does Anyone Have the Right to Sex?"

[23] Tognazzini and Coates, "Blame."

[24] Carlsson, "Blameworthiness as Deserved Guilt," 91; and Fricker, "What's the Point of Blame?" 173.

blame often harms us.[25] Moreover, blame, like wrongdoing, regularly imposes *relational harms*. The blamer and blamee cease to be in a relationship in which they both have and recognize that they have good will for each other.[26] Such a relationship gives them faith that each other will follow shared norms. Damaging this relationship, then, hinders goods of reliable norm compliance like safe vulnerability and mutual respect.[27] To stop these harms, the blamee must usually deny the act, excuse it, justify it, or atone for it. This *reparative labor* often takes time, energy, and social sensitivity.[28] Finally, if the misdirected blame persuades the blamee, they *gain a false moral belief* that their action is wrong, which can restrain them from living as they desire.[29] Misdirected, recognized blame in a relationship does not always cause all seven of these harms. But almost always, it causes at least some of them.

When blame is well directed, it can still cause some of these harms, like negative moral assessments. But it does so in the presence of justifying moral considerations: the moral improvement of the blamee; the reparation of past harms and damaged relationships; the prevention of future harm; etc. When blame is misdirected, however, it frequently lacks justifying moral considerations. There might be exceptions. For instance, on complex moral issues that require taking a stand, it might be worth it to risk levying misdirected blame. But oftentimes, there are not moral considerations that justify levying misdirected, recognized blame on a close partner.

Harming someone in the absence of justifying moral considerations wrongs them. This claim leaves open which moral considerations are enough to justify a given harm and whether harmless wronging is possible. Accordingly, I take this weak claim to be widely shared; I will not robustly defend it.

From this claim, we get to P3: misdirected, recognized blame directed at a close partner often wrongs them. When the husband in Guilty blamingly sulks at his wife "for hours," when his blame makes her feel "sick of the guilt," when she is "ready to walk away" from their twelve-year marriage—and all she has done is to reject some sexual interactions—she is not the victim of some cosmic tragedy.[30] She is the victim of a wrong.

Sarah Conly objects to P3. She claims that threats of "emotional pain," such as sulking or misdirected blame, impose "pressure of a sort an honorable

[25] McGeer, "Civilizing Blame," 166–67 and 181–82.

[26] Hieronymi, "The Force and Fairness of Blame," 144n30 and 145n34.

[27] McGeer, "Civilizing Blame," 163, 174.

[28] Hieronymi, "The Force and Fairness of Blame," 124–25.

[29] Fricker, "What's the Point of Blame?" 181.

[30] See the anonymous "Stay at Home Mum" post cited above note 18.

person wouldn't," but they do not "[go] beyond [one's] rights."[31] She explains, "It is the nature of family relations that you may use your relationship to (try to) impose on other family members, at least up to a point. . . . We are vulnerable to our families, but that vulnerability is the price you pay for having an emotional relationship."[32]

At what point is it wrongful to leverage relational ties to impose on one's partner? Conly does not give a comprehensive answer, instead discussing various examples. For instance, she thinks that it is permissible to threaten to break up with a partner unless they change, as long as the change bears on the relationship's health and is not itself immoral.[33] In contrast, threatening violence to induce change is clearly impermissible.[34] Threats of emotional pain, Conly suggests, are akin to permissible threats of a break-up.

Conly's reasoning neglects that emotional pains are heterogeneous. It can certainly be okay to impose some emotional pains (for example, a painful but important expression of disappointment). But I have just argued that some other emotional pains—specifically, instances of misdirected blame—often wrong the victim. The fact that some forms of emotional pain are the "price you pay" for a relationship does not entail that every form of emotional pain is similarly permissible. Accordingly, we should reject Conly's objection.

Alan Wertheimer advances a different objection. He says, "People are sometimes justified in being angry with others ... [but] even when expressions of anger are not justified, it does not follow that one's behavior is rights-violating [or obligation-violating]. Some boorish behavior is part of the rough and tumble of life."[35] One interpretation of this objection is as follows. First, angry

[31] Conly, "Seduction, Rape, and Coercion," 114–15. Conly's arguments concern threats of "emotional pain" writ large, which she also describes as ways of "using the strength of family ties to [one's] own ends." Additionally, she mentions a laundry list of pressure tactics that fit under this category: guilt tripping, sneering, contemptuously castigating, coaxing, cajoling, wheedling, importuning, haranguing, berating, and browbeating. For this reason, I take her comments to apply to sulking and misdirected blame, even though she does not explicitly mention them.

[32] Conly, "Seduction, Rape, and Coercion," 115.

[33] Conly, "Seduction, Rape, and Coercion," 110. For discussion of the complex ethics of break-up threats, see Liberto, "Threats, Warnings, and Relationship Ultimatums," as well as Ferzan, "Consent and Coercion," 977–78.

[34] Conly, "Seduction, Rape, and Coercion," 118.

[35] Wertheimer, *Consent to Sexual Relations*, 183. Note that Wertheimer's comments are explicitly about "unjustified anger" writ large but are nonetheless relevant. After all, recall that sulking's affective core is anger. And when sulking involves *misdirected* blame, the anger involved in such sulking is thereby unjustified. Furthermore, note that I add "obligation-violating" to the quote because Wertheimer switches between "rights talk" and "obligation talk" throughout his piece.

misdirected blame imposes minor harms. Second, minor harms are merely "boorish"; they are not wrongs. If they were, everyone would walk on eggshells to avoid them. Moreover, we would not easily let go of these harms. Instead, wrongdoers would make costly amends; third parties would expend effort to support the victim. These actions would likely be more significant than the minor harm suffered! In other words, we have an interest in avoiding excessive duties of diligence, rectification, and victim support. This interest stops minor harms from being wrongs.

I do not find this objection convincing. Even when breaching a promise imposes minor harms, it can still wrong the promisee. This is because keeping a promise does not always require excessive diligence, and breaching a promise need not lead to excessive rectification and victim support. Similar reasoning applies to misdirected blame. Oftentimes, avoiding misdirected blame requires only thinking before you blame, not walking on eggshells. Similarly, rectification can involve a brief apology; support can involve a brief reassurance that the victim is not to blame. Hence, our interest in avoiding excessive duties of diligence, rectification, and victim support should not stop minor harms from being wrongs.

In any case, misdirected blame often imposes major harms, at least when it is prolonged, pervasive, habitual, and directed at a close partner. My argument concerns exactly such cases. Accordingly, even if minor harms are not wrongs, one may adopt a duly restricted version of P3 without undermining my conclusion. Wertheimer's objection thereby fails to refute my argument.

2.2. *Wrongful Sulking*

P3 leads to my conclusion. Contra Conly and Wertheimer, a sulker for sex does not just evince bad character or impose nonwrongful harm. They often wrong the sulkee via misdirected blame. Hence, the husband in Guilty does not just have a nondirected duty to become more virtuous. He also has a directed duty of atonement to his wife, like a duty to apologize.[36]

Some sexual pressures do not involve misdirectedly blaming a close partner. Thus, my argument does not imply that all sexual pressures are wrongful, let alone wrongful in the exact same way. But some sexual pressures can involve misdirected blame, for instance, aggressive shouting, and verbal jabbing. My arguments usefully extend to such pressures.

Importantly, sulkers who levy blame might commit additional wrongs. To take one example, their behavior might transform sexual "invitations" into

[36] Radzik, *Making Amends*.

"demands."[37] To take another, if they prevent a sulkee from challenging their blame, their blame might be inappropriately peremptory.[38] Putting these points aside, this section suffices to establish that it is wrongful to blamingly sulk at someone *for* sex—independent of whether one sulks them *into* sex.

3. Blamingly Sulking into Sex

> *Tried*: When I tell [my husband] no, he fucking pouts about it. His mood is off for hours or even the rest of the day. I've tried explaining to him why I'm not interested, and I've told him how his sulking is annoying and makes me feel bad. I wonder how he would feel if he knew how many times I've consented to sex just because I don't want to have to deal with his pouting. . . . I've tried explaining to him how I feel touched out. He doesn't get it. I've tried explaining to him that when I have a million things to do sex is the last thing on my mind. He just doesn't get it. It makes me so angry. I feel like I have to choose between his grumpy mood or having sex even when I don't want to.[39]

In Tried, from another Reddit thread, the husband does not just blamingly sulk at his wife *for* sex; he does not just wrong her via *misdirected blame*. He sulks her *into* sex, and so he wrongs her further. He *coercively undermines her consent*. To make this argument—and to generalize beyond this case—I will start by identifying sufficient conditions for consent-undermining coercion. In the rest of the section, I will show that blame-laden sulking into sex often meets these conditions.

3.1. Consent-Undermining Coercion

Consent to an activity is the normative power to release another person from a duty not to infringe on the relevant domain of your authority.[40] Importantly, I might agree to something without my agreement counting as morally transformative consent. For example, if I am coerced into saying yes to a sexual activity, I have *agreed*, but I have not *consented*.[41] (Other examples include agreement

[37] Kukla, "That's What She Said," 80–84.

[38] Patwardhan, "Peremptory Blame."

[39] u/tri_nisvx (username), "The No Sex Sulk." Reddit, June 8, 2020, https://www.reddit.com/r/breakingmom/comments/gzf9fe/the_no_sex_sulk/.

[40] My argument does not depend on holding this view of the dynamics of consent. For a recent survey of various views, including a defense of a novel, "scope-shifting" view, see Liberto, *Green Light Ethics*, 60–87.

[41] Some theorists prefer not to use "consent" as a success term, instead distinguishing morally transformative "valid consent" from "invalid consent" (as well as from "no consent at all"). This usage would not change my arguments.

induced by incapacitation, deception, etc.) A sexual activity is consensual if and only if all participants consent to it. Otherwise, it is nonconsensual, or, in other words, the consent of one or more participants has been undermined.

To articulate five jointly sufficient conditions for consent-undermining coercion, I will consider a paradigmatic case. *A* credibly threatens *B*, "I will hit you if you do not let me touch you." *B*, preferring not to be hit, agrees. Clearly, *B* has not consented; *A*'s touch is nonconsensual. In this case, *B* is entitled to have the option of not agreeing and yet not being hit, since being hit would wrong them. But they are confident that this option is unavailable. Their confidence is not accidental. It stems from *A*'s threatening *B*—*A*'s acting at least recklessly, if not knowingly or intentionally. Because *B* prefers to avoid being hit, *B* lets *A* touch them. This decision, importantly, seems eminently reasonable. (Later, I will elaborate on what "reasonable" means.) Hence, *B* lacks meaningful discretion between the options to which they are entitled. Their agreement thereby fails to genuinely exercise authority over their sexual life; *A* still had a duty not to touch *B*.[42] With this illustration, I can now formalize this section's argument.

P4. If someone (*B*) agrees to another person (*A*) doing something (*X*), and the following conditions are met, *B*'s consent to *X* is undermined via coercion:
   a. *Unavailable Option*: *B* has sufficiently high confidence that *A* will do *Y* unless *B* agrees to *X*;
   b. *Moral Baseline*: *Y* would morally wrong *B*;[43]
   c. *Causal Role*: *A*, through words or conduct, intentionally, knowingly, or recklessly caused *B* to have the confidence referred to in Unavailable Option;
   d. *Preferable Compliance*: *B* agrees to *X* because *B* prefers that to facing *Y*;
   e. *Reasonable Compliance*: It is reasonable for *B* to agree to *X* because they prefer to do that rather than to face *Y*.[44]

P5. Many cases of being blamingly sulked into sex meet these conditions.

C2. In many cases of being blamingly sulked into sex, the sulkee's consent to sex is undermined via coercion.

[42] The reasoning guiding this illustration resembles that offered by Wertheimer in *Coercion*, 202–21, 267–86.

[43] For this language of "baselines," see Nozick, "Coercion."

[44] I follow standard conceptions of recklessness: to "recklessly" cause such confidence is to recognize but unjustifiably disregard the risk that one will cause it. See Edwards, "Theories of Criminal Law." A minorly different definition would not affect my arguments.

I have already supported P4 via the earlier analysis of *A* threatening to hit *B*. Moreover, this premise is quite modest. It gives *jointly sufficient* conditions for coercion, not *necessary* conditions. Indeed, many philosophers (including me!) doubt that these conditions are necessary.[45] P4 also need not expose the *best explanation* of why its conditions suffice for coercion; one could reformulate P4 to better "carve at the joints." Such modesty makes P4 well accepted.[46] Crucially, P4 is accepted even by Sarah Conly, Kimberly Kessler Ferzan, and Alan Wertheimer—scholars who doubt that behaviors like sulking can undermine

[45] To begin, P4 does not account for "third-party coercion," whereby a third party, *C*, coerces *B* into submitting to *A*. Moreover, to list some arguments specific to each of P4's conditions: Claudia Card loosens the Unavailable Option and Preferable Compliance conditions, seeming to suggest that even a threatening atmosphere can undermine consent. David Zimmerman modifies the Moral Baseline condition, arguing that *Y* need not wrong *B* for *A* to undermine *B*'s consent. Tom Dougherty removes the Causal Role condition, arguing that *A* need not cause *B*'s confidence to undermine their consent. And Dougherty also argues against a condition similar to the Reasonable Compliance condition, showing that *B*'s consent can be undermined even if noncompliance is reasonable. See respectively Card, "Recognizing Terrorism," 18–19; Zimmerman, "Coercive Wage Offers," 131–38; and Dougherty, "Coerced Consent," 443–51, and "Sexual Misconduct," 333.

[46] Anderson, "Coercion." As Anderson notes, accounts of coercion differ along two dimensions. The first dimension is the extent to which they focus on the coercee's situation or on the coercer's conduct. The second dimension is the extent to which they are "moralized"—requiring prior normative judgments—or "nonmoralized." See also Anderson, "Of Theories of Coercion, Two Axes, and the Importance of the Coercer," 396–404. P4 is closest to a coercee-focused, moralized account. (Note that because P4 does not offer necessary conditions, it is somewhat inaccurate to describe it as a full "account" of coercion.) Nevertheless, I describe P4 as "well accepted" because it aligns with the "standard view" in the contemporary literature on coercion. See Anderson, "Coercion", "Of Theories of Coercion, Two Axes, and the Importance of the Coercer," 396, 411, and "How Did There Come to Be Two Kinds of Coercion?" 24–29. Moreover, as I discuss in the main text below, my opponents accept P4 as sufficient for consent-undermining coercion.

Finally, it is worth noting that my conclusion would still follow from accounts of coercion that are coercer focused and/or nonmoralized. Take Anderson's own coercer-focused, nonmoralized account, discussed in "Of Theories of Coercion, Two Axes, and the Importance of the Coercer," 414–21; "The Enforcement Approach to Coercion," 6–18; and "Conceptualizing Rape as Coerced Sex," 72–85. Despite being explanatorily different from coercee-focused, moralized accounts, his account is extensionally similar, as he mentions in "The Enforcement Approach to Coercion," 10. Moreover, in "Coercion as Enforcement and the Social Organisation of Power Relations," Anderson also extends his account to nonparadigmatic cases of coercion (529–39). For reasons like these, my analysis of coercive sulking does not depend on adopting a coercee-focused, moralized account. That said, thoroughly defending this claim would require too much space here, so I leave this for other work. Thanks to an anonymous reviewer for pressing me to clarify how P4 relates to other accounts of coercion.

consent.[47] Thus, I am content to appeal to P4 without robustly defending it. I now turn to defending P5.

3.2. Consent-Undermining Sulking

Recall my focus on typical cases of sulking into sex—prolonged, pervasive, habitual, blame-laden sulking that gets a close partner to submit. Many such cases meet the Unavailable Option condition. For consider a sulkee who submits to sex. It is likely that either they believe that their partner will keep sulking if they say no, or they believe that their partner will escalate to worse behaviors if they say no. Without one of these two beliefs, the sulkee would likely have refused. Certainly, in some cases, sulkees worry that their partners will escalate if rejected. But frequently, sulkees seem to submit despite not worrying about this. In these cases, it is likely that the sulkees submit because they are confident that their partners will continue sulking if they say no. Thus, many cases of being blamingly sulked into sex meet the Unavailable Option condition.

Of these cases meeting the Unavailable Option condition, many also meet the Moral Baseline condition. Some philosophers, like Conly and Wertheimer, doubt this, skeptical that sulking can wrong a sulkee.[48] But in section 2, I undercut this doubt. There, I argued that blamingly sulking *for* sex often involves wrongful misdirected blame. That argument applies equally to blamingly sulking *into* sex.

Cases that meet the Unavailable Option and Moral Baseline conditions often meet the Causal Role condition. Many sulkers intend to cause sulkees to believe that they must submit for the sulking to stop. Even when sulkers do not intend this, they often realize that their conduct is likely or certain to cause this belief. After all, sulkers often recognize not only that they are sulking for sex but also that sulkees will likely pick up on this. Even truly unaware sulkers are often made aware, for instance, by a sulkee asking, "Will you stop sulking if I say yes?" Hence, in many cases of blamingly sulking into sex, the sulker intentionally, knowingly, or recklessly causes the sulkee to believe that if they do not have sex, the sulking will not stop.

[47] Conly, "Seduction, Rape, and Coercion," 104–10; Ferzan, "Consent and Coercion," 963–65, 968–80, 994–97, and 1005–7; and Wertheimer, *Consent to Sexual Relations*, 165–71 and 177–86. Technically, Conly argues that P4 should require intent (104–5). However, her arguments fail to show that recklessness is insufficient. In fact, they show that negligence would be sufficient. Wertheimer might also modify P4 minorly. In *Coercion*, he suggests that if *B* prefers their agreement to count as consent, then in some cases, it should (277). I will not discuss this kind of objection, because sulkees will rarely want their agreement to count as consent.

[48] Conly, "Seduction, Rape, and Coercion," 114–15; and Wertheimer, *Consent to Sexual Relations*, 183.

Cases that meet these first three conditions frequently meet the Preferable Compliance condition. As one woman named Teresa recounts, "Sometimes I'd just submit, otherwise he'd sulk for three days and be nasty. So it was the lesser of two evils. . . . It was easier to grit your teeth and think of mother England and be done with it."[49] Of course, a sulkee's agreement is not always motivated by a preference to avoid sulking. For example, it can be motivated by a preference not to wrong the sulker—if, say, the sulker deceives the sulkee into seeing sex as obligatory.[50] But as Teresa recounts, sulkees often see sex not as obligatory but as the "lesser evil." Hence, in many cases of being blamingly sulked into sex, the sulkee agrees because they prefer to avoid sulking.

On some views, these first four conditions suffice for consent-undermining coercion.[51] But because Conly's, Wertheimer's, and Ferzan's views require the Reasonable Compliance condition, I will end by showing that cases meeting the first four conditions often meet this fifth condition.

For it to be reasonable for *B* to agree to *X* because they prefer to do that rather than to face *Y*, two conditions are necessary and sufficient. First, agreeing to *X* must be *objectively preferable* to facing *Y* and to pursuing other alternatives. One way of spelling this out is to say that agreeing to *X* must be less harmful than facing *Y* or pursuing other alternatives—not according to *B* but according to a reasonable or ordinary person. After all, if *B* agrees because they miscalculate the costs of *X*, *Y*, and alternatives, we should not paternalistically relieve them of responsibility for that miscalculation.[52] Second, *B* must not have an *easily accessible remedy* for *Y*.[53] Otherwise, the autonomy-constraining threat of *Y* would be counterbalanced.[54]

[49] Murphy, "Tactic #13."

[50] For an argument that such moral deception can undermine consent, though not via coercion, see Patwardhan, "Do I Have To?"

[51] For instance, Dougherty's account of coercion does not require conditions like the Reasonable Compliance condition. See Dougherty, "Sexual Misconduct on a Scale," 324–26, 330–34. For what it is worth, I am similarly skeptical of this condition.

[52] Ferzan talks in terms of "bad choices" and "mistakes" ("Consent and Coercion," 975). Conly talks in terms of "harms great enough to affect [one's] decision procedure" ("Seduction, Rape, and Coercion," 106). I use the term "objective preferability" to unify their terminology, but I am not committed to this specific term. One could reformulate this condition while maintaining its spirit.

[53] Ferzan, "Seduction, Rape, and Coercion," 996–97, 1006; Wertheimer, *Consent to Sexual Relations*, 184, and *Coercion*, 267, 275–76.

[54] Wertheimer has articulated two other necessary conditions for the Reasonable Compliance condition: the harm of *X* must be grave enough to justify third-party intervention; and it must be reasonable to expect *A* to believe that *B* agrees to *X* to avoid *Y*. For discussion, see Wertheimer, *Consent to Sexual Relations*, 184–85, and *Coercion*, 277–78. My

Conly, Ferzan, and Wertheimer doubt that pressures like sulking can meet these two conditions. They believe, roughly, that facing the sulking is objectively preferable to submitting to sex; that there are other alternatives that are objectively preferable; or that the harms of sulking are easily remedied.[55] These beliefs, I will argue, are mistaken.

In many cases of being blamingly sulked into sex, submitting is objectively preferable to facing continued sulking. As discussed earlier, sulking involves unpleasant withdrawal, and misdirected blame involves several serious harms. These harms compound as a relationship gets closer, as the number of affected third parties (e.g., one's children) increases, and as sulking gets longer, more pervasive, and more habitual. For instance, in Tried, the sulkee recounted, "When I tell him no, he fucking pouts about it. His mood is off for hours or even the rest of the day. . . . I wonder how he would feel if he knew how many times I've consented to sex just because I don't want to have to deal with his pouting."[56] In Guilty, the sulkee was ready to end a twelve-year marriage because she was "sick of the guilt!"[57] Yet a third sulkee has shared, "Saying 'no' and holding that 'no' in the face of someone deeply resistant, is *exhausting*."[58]

Undoubtedly, the harms of submitting to sex can also be serious. But sometimes people can reduce some of its harms, e.g., hastening its end by faking an orgasm. More importantly, to claim that submitting can be objectively preferable to facing sulking does not imply that the harms of the former are trivial. It implies only that the harms of the latter can outweigh them.

What if a sulkee submits to sex to avoid brief, compartmentalized, or one-off sulking? They might be irrationally catering to their partner's desires. That said, we should hesitate to draw this conclusion lest we too hastily impute a kind of false consciousness. Moreover, recall that typically, sulking is the opposite: it is prolonged, pervasive, and habitual. In these cases, sulkees who submit

argument can easily extend to these conditions, so for simplicity's sake, I do not discuss them further.

[55] Conly, "Seduction, Rape, and Coercion," 114–15; Ferzan, "Consent and Coercion," 954–56, 971–80, 994–95, 1002–7; and Wertheimer, *Consent to Sexual Relations*, 183. As I discuss in notes 31 and 35, Conly's and Wertheimer's arguments apply to sulking even though they do not explicitly mention it. In contrast, Ferzan explicitly rejects the possibility that submitting to sulking or guilt-tripping could be reasonable (994, 1002–3). Ferzan also makes similar arguments about various verbal pressures: needling, haranguing, cajoling, pestering, badgering, whining, and more (955–56, 972, 974–75, 995, 1006). These arguments do not depend on the pressures being verbal, so they provide additional support for her skepticism about consent-undermining sulking.

[56] See the Reddit post cited above note 39.

[57] See the "Stay at Home Mum" post cited above note 18.

[58] Price, "A Few Words about Sexual Coercion in the Wake of the Aziz Ansari Accusations."

are often choosing rationally. To say otherwise seems baselessly patronizing. Hence, submitting is often objectively preferable to facing continued sulking.

Frequently, submitting is also objectively preferable to pursuing alternatives. Trying to distract a sulker is routinely effortful and ineffective. As observed in section 1, sulkers have strong incentives to prolong their sulking. Similarly, extrication is difficult. Even within nonabusive relationships, sulkees can face myriad obstacles to leaving the relationship or shared space: logistical (e.g., low funds, limited transportation, or childcare needs); psychosocial (e.g., internalized and social sanctions for gender norm violations); religious (e.g., prohibitions against divorce); relational (e.g., the value of the relationship itself); etc. Even if a sulkee does leave, the sulker's blame can keep weighing on them. These difficulties with extrication compound when sulking is prolonged, pervasive, habitual, and in a close relationship. Accordingly, the main alternatives to submission—distraction and extrication—are seldom promising.

Finally, sulking can rarely be remedied easily. Set aside legal remedies, like awards of damages; our focus is the moral sphere. In this sphere, one remedy is improving *resilience* to sulking. But to value interpersonal engagement with someone is to feel a loss when they enduringly, pervasively, and habitually withdraw. Being vulnerable to a partner's blame, moreover, reduces moral complacency.[59] Hence, becoming inured to blame-laden sulking is a costly remedy. Another remedy is a sulker's *atonement*. Could a sulkee convince a sulker to atone? Yes, but recall that challenging a sulker's blame is often quite difficult. Could a sulkee's friends or other third parties encourage a sulker to atone? Yes, but frequently, a sulkee would have little confidence that they could get third parties to intervene; that such intervention would succeed; or that any intervention would not itself have costs, like blowback from bad-mouthing a partner. In sum, sulkees often lack easily accessible remedies for being blamingly sulked into sex.

Thus, in many cases of being blamingly sulked into sex, submitting is objectively preferable to facing continued sulking or to pursuing other alternatives, and such sulking is not easily remedied. In other words, the cases that meet P4's first four conditions frequently meet its final condition, Reasonable Compliance. Contra Conly, Ferzan, and Wertheimer, P5 is true.[60]

[59] Analogously, Krista Thomason argues that "Shame prevents us from ignoring our unflattering features" (Thomason, "Shame, Violence, and Morality," 2).

[60] I have argued that "many" cases of being blamingly sulked into sex meet the Unavailable Option condition, "many" cases that do so also meet the Moral Baseline condition, and so on. Could "many" of "many" (of "many" . . .) cases amount to a few? I doubt it, given how many stories of sulking appear to meet all five conditions. But I welcome evidence suggesting otherwise.

From these premises follows C2: being blamingly sulked into sex often undermines one's consent via coercion. It can be tempting to see sulking as something that little kids do—that is to say, annoying but relatively inconsequential. I hope to have weakened this temptation. Sulking can be quite powerful within adult relationships—powerful enough to induce nonconsensual sex.

4. Implications

> *Initiate*: My husband and I have been married for 16 years.... Over the years, my sex drive has waned because of stress, age, work, children, etc. I try to make an effort to be intimate every week, [but] sometimes I just don't feel like having sex. Rather than trying to "woo" me just a little or even initiate sex when we have quiet time and our kids aren't likely to walk in, my husband sits and sulks until I make the first move. He does this every single time.... I know I could work on my libido, but why is it always up to me? Why do I get the guilt trip?[61]

My argument calls us to reflect seriously on sulking within sexual relationships. But before such reflection, it is useful to clarify the argument's moral and legal import. Additionally, it is useful to see how my argument generalizes (to cases like Initiate) without *over*generalizing. This section's discussion of these points will necessarily be partial: there is much more to say than space here allows.

4.1. Moral and Legal Import

All directed wrongs disrespect someone as a person. But unlike directed wrongs that are nonsexual, sexual wrongs also disrespect someone as an embodied, sexual agent. And unlike other sexual wrongs, nonconsensual sex involves another form of disrespect: a wrongful violation of one's *authority* over one's sexual life. As Hallie Liberto describes, "Any consent-related violation just is ... a breach of their authority within a domain that they are entitled to control."[62]

Violations of sexual authority have at least three interlocking dimensions. They inhibit *sexual autonomy*, the ability to construct and govern one's sexual life; they inhibit *sexual freedom* (from interference and from domination); and they inhibit *trusting, intimate sexual relationships* by disrupting decisions about if, when, and how to engage in sex within various relationships. Of course, sex can also be wrongful for reasons unrelated to authority. Accordingly, sulking into sex can involve many wrongs besides nonconsensual sex. Such wrongs include exploitative sex, unjust sex, stalking, derivatization, inattention or lack

[61] Sugar and Mitchell, "Sulking for Sex."

[62] Liberto, "Coercion, Consent, and the Mechanistic Question," 232.

of attunement, inequality deployed as a form of force, and more.[63] But in the cases that I have discussed, the sulker does not just unfairly take advantage of the sulkee; does not just truncate the sulkee's distinct sexual agency; and so on. The sulker also violates the sulkee's authority. Instead of their close relationship being a space for free, intimate, sexual exploration, the sulker makes it a space where the sulkee capitulates to a sexual life imposed upon them. For this reason, I adopt a "both and" approach. Sulking into sex can involve both nonconsensual sex and wrongs unrelated to consent.

Nonconsensual sex, importantly, is a *degreed* wrong. After all, consent-undermining mechanisms are themselves degreed: victims can be more or less coerced, more or less incapacitated, etc. Relatedly, victims' experiences—which also affect the wrongfulness of nonconsensual sex—vary widely. Even sex itself is widely variable. As is common sense, a heteronormative view of sex as penile-vaginal penetration is arbitrarily narrow. Although literature on consent often neglects the immense diversity of sex, such diversity implies a concomitant variability in the wrongfulness of nonconsensual sex. Finally, nonsexual consent violations, like theft, are degreed wrongs. Why would sexual consent violations differ in this regard? In sum, what unifies the category of "nonconsensual sex" is not the *gravity* of its wrong but the *nature* of it.

C2 does not imply, then, that nonconsensual sex induced by sulking is always as wrongful as other forms of nonconsensual sex. Some instances of the former might be of the utmost gravity. For instance, one sulkee describes her experience as "emotional or psychological abuse," recounting, "My husband sulks and won't speak to me for up to four weeks if I don't respond to his requests for sex."[64] Many other cases of sulking might not be so egregious.

One might be tempted to reserve the category of "nonconsensual sex" for wrongs of the utmost gravity. But let me mention two benefits to conceptualizing the wrong of nonconsensual sex as degreed. For victims of less egregious violations of sexual authority, this conceptualization helps them to understand their experiences *as experiences of violated authority*. Such victims can then see the crucial *continuity* between their own experiences and the experiences of

[63] On exploitative sex, see Anderson, "Sex under Pressure," 368; and Yap, "Conceptualizing Consent," 56–60. On unjust sex, see Cahill, "Unjust Sex vs. Rape," 754–57. On stalking, see Patwardhan, "Stalking by Withdrawing." On derivatization, see Cahill, *Overcoming Objectification*, 32, 138–39. On inattention or lack of attunement, see Anderson, "A Phenomenological Approach to Sexual Consent," 2, 14–21. On inequality deployed as a form of force, see MacKinnon, "Rape Redefined," 469–77.

[64] "Emotional Abuse in Sulking Silence when Sexual Demands Go Begging." For another sulkee's discussion of abusive sulking, see the *Netmums Forum* post by user Jo B(1113) cited above note 12.

victims who face more egregious violations of sexual authority—despite the differences between those experiences. Those who are not victims also benefit from seeing this continuity. Many people recognize and morally attend to sex that is blatantly nonconsensual, like sex induced by threats of fatal violence. But we should dedicate a similar kind of moral attention to sex that is less blatantly nonconsensual, like coercive sulking into sex. Seeing the continuity between these cases helps us to do exactly this.

Conceptualizing the wrong of nonconsensual sex as degreed has a second benefit. By recognizing that my authority over my sexual life can be violated in a vast spectrum of ways, I can more easily recognize that my authority itself is vast. For example, I can more easily recognize that I have the authority to demand a sexual life free from coercive sulking—and from all manner of coercive pressures, no matter their severity. In this way, seeing the wrong of nonconsensual sex as degreed serves to empower us. These two benefits, along with the various considerations above, make it natural to see nonconsensual sex as a wrong with variable gravity.[65]

The discussion above also reveals why it is important to talk about the wrongfulness of sulking in terms of consent. Some theorists might prefer otherwise. For instance, Jonathan Jenkins Ichikawa observes, if consent is "whatever it is that makes sex *qua* sex morally permissible," then it is a normatively useless concept; it cannot explain anything about the ethics of sex.[66] We could focus instead on the equality of sexual interactions, as Catharine MacKinnon suggests.[67] Or maybe the problem is that consent has an extremely low bar; we cannot theorize about subtle or minor sexual pressures in terms of consent.[68] Or perhaps the concept of consent is not flawed, but our deployment of it is. Audrey Yap, for example, argues that people problematically take sexual desire to suffice for consent.[69] And Quill Kukla (writing as Rebecca Kukla) argues that discussions of consent frequently reinforce the narrative that ethical sex simply requires acquiescence—usually, a woman acquiescing to a man initiating.[70] Arguments like these might encourage theorizing about sulking in terms other than consent.

[65] For lengthier discussions of the degreed wrong of nonconsensual sex, see Liberto, *Green Light Ethics*, 244–50; Dougherty, "Sexual Misconduct on a Scale," 337–43; and Boonin, *When Yes Means No*, 145–69.

[66] Ichikawa, "Presupposition and Consent," 23–24, 23n36.

[67] MacKinnon, "Rape Redefined," 431, 436, 439–51, 462–65, 469–70, and 476.

[68] MacKinnon, "Rape Redefined," 440, 443–50; and Yap, "Conceptualizing Consent," 56–60.

[69] Yap, "Conceptualizing Consent," 51–56.

[70] Kukla, "That's What She Said," 75.

But as per the discussion above, the language of consent crucially illuminates how sulking into sex can violate one's *authority* over one's sexual life. The ethical considerations of authority that underlie consent are neither comprehensive—encompassing every consideration relevant to moral permissibility—nor reducible—to considerations like equality. This is why I adopt a "both and" understanding of the wrongfulness of sulking into sex. Moreover, C2 belies the claim that we cannot theorize about sulking in terms of consent, that the bar for consent is too low. The sufficient conditions for coercion that apply to paradigmatic sexual pressures apply to sulking all the same. These pressures are *continuous*. Finally, it is indisputable that our deployment of the language of consent is often pernicious. We can avoid some of these dangers by being more careful, e.g., never taking desire to suffice for consent. But more importantly, the language of consent can also be *empowering*, as discussed above. C2 shows that a sexual life free from coercive sulking is not just valuable but something that we have the authority to demand. For these reasons, it is important to theorize about sulking in terms of consent.[71]

I turn now to C2's legal implications. Criminalizing nonconsensual sex raises complicated questions regarding, among other things, the import of interpersonal privacy; the costs of criminal legal bureaucracy; the requirements for culpability; and the effectiveness of criminal punishment as a remedy for sexual wrongs. For this reason, the fact that a sexual interaction is nonconsensual does not entail that it should be considered a crime—a specific instance of the general principle that moral wrongdoing does not entail criminal wrongdoing. Accordingly, C2 does not imply—nor do I believe—that nonconsensual sex induced by sulking should always be criminalized.

This position might initially seem controversial. But let me offer just one schematic, supporting example. Consider sulking into sex within a longstanding relationship. Suppose that both partners, having recognized the wrong, are reconstructing a healthier sexual life. In such cases, criminal punishment is sometimes helpful, sometimes not. It can be a galvanizing tool, spurring a wrongdoer to take reparation seriously. But it can also meddlesomely interfere with the victim standing up for themselves and with the victim and wrongdoer working things out together. In sum, criminal accountability is not the only form of accountability for nonconsensual sex, nor is it always appropriate. Of course, there is still much more to say (as with every point in this section).[72]

[71] I am grateful to an anonymous reviewer for urging me to engage more explicitly with feminist criticisms of consent.

[72] For more discussions of nonconsensual sex in the moral versus criminal spheres, see, for example, Patwardhan, "Meddlesome Blame for Nonconsensual Sex"; and Wertheimer, *Consent to Sexual Relations*, 2–3, 5–6. For discussion of the meddlesomeness of criminal

Just as C2 avoids implying that sulking into sex should always be criminalized, so too does it avoid implying that sulkers are always culpable. Some sulkers are clearly culpable, e.g., those who knowingly levy misdirected blame. Other sulkers may not be, e.g., those whose pernicious socialization makes them justifiably ignorant that their blame is misdirected. Nevertheless, nonculpable sulkers still ought to rectify the wrong done.

4.2. Extending the Argument

My argument does not entail that all sexual pressures undermine consent. For example, some pressures—like economic incentives in the context of ethical sex work—may not be wrongful. If so, they do not satisfy P4's Moral Baseline condition. That said, my argument also does not entail that sexual pressures like these are *not* consent undermining. This is because P4 articulates jointly sufficient conditions for consent-undermining coercion, not necessary conditions. In this way, I avoid endorsing Robin Morgan's expansive view that consent is undermined whenever not initiated out of one's own affection and desire.[73] So too do I avoid ruling out this view—despite being skeptical of its implication that all pressured sex is similarly nonconsensual. In other words, I take pressured sex to be morally heterogeneous, and my argument leaves room for this.

Although my argument does not overgeneralize, it does generalize. I have focused on blame targeted at sexual *refusal*. But P4 can hold even when blame is targeted at sexual *noninitiation*. Recall Initiate: "my husband sits and sulks until I make the first move."[74] I have also focused on *occurrent* sulking. But in relationships involving habitual sulking, P4 can be met *before the sulking starts*. Consider this story: "At 71, I have no desire to have sex. However, my husband, 80, is still keen, and if I turn him down, he sulks.... For ages now, I have gone along with it."[75] Here, the wife suggests that she "goes along with it" so that her husband does not even start sulking.

Blaming behaviors other than sulking, e.g., incessant criticism, can also satisfy P4. One person recounts their complicated experience: "I said yes ... because I didn't want him to be mad at me. Or yell at me. And I wasn't sure I didn't want it. I was already there, so I just let it happen."[76] Indeed, paradigmatic cases of coercion, like cases of sex to avoid assault, often involve threatening

prosecution of relational wrongs more generally, see, for example, Mendlow, "The Moral Ambiguity of Public Prosecution." I am grateful to an anonymous reviewer for asking me to say a bit more about the legal implications of C2.

[73] Morgan, "Theory and Practice," 165.

[74] Sugar and Mitchell, "Sulking for Sex."

[75] Parker and Parker, "Steph and Dom Solve Your Sex, Love, and Life Troubles."

[76] Bennett and Jones, "45 Stories of Sex and Consent on Campus."

both misdirected blame *and* violence. The coerciveness of such conduct is thereby overdetermined. Furthermore, if someone habitually uses misdirected blame to coerce their partner into sex, P4 can be met even when their conduct does *not* involve blame. For they might still recklessly cause their partner to believe that any sexual refusal will be met with wrongful blame. P4 applies even to pressure tactics that *never* involve blame, as long as they involve a different kind of threat to wrong the coercee.

Finally, P4 can hold for nonsexual interactions too. Say that to end a days-long sulk, my partner lets me paint their office in my favorite color. Clearly, their compliance is not consent. Of course, painting their room is less wrong than having nonconsensual sex—authority over wall color is less important than sexual authority—but what I did is still nonconsensual. In other words, sulking can be a general tactic of coercive control.[77]

Thus, it is vital that we recognize that sex to avoid blame-laden sulking is often nonconsensual. This conclusion has numerous moral and legal implications, not only for sulking but also for myriad other behaviors. Perhaps most importantly, this conclusion reveals the continuity between diverse forms of nonconsensual sex and empowers us to demand a sexual life free from all of them.

5. CONCLUSION

> *Recognition*: As the #MeToo movement began to take form ... I started to question my actions.... I started to see that while I believed I had always been respectful and obtained consent, my sex life involved many incidences of pressuring women into sexual acts until they relented.[78]

[77] Contrast Ferzan's verdict about a different nonsexual case: a teen incessantly sulking (or using similar pressure tactics) to get their parent to buy them ice cream ("Consent and Coercion," 993–96). Her focus on a child-parent relationship in this case muddies our intuitions for multiple reasons. For one, Ferzan observes, "good parents don't give in to their children's whims" (994). In other words, it is objectively preferable for the parent to withstand the sulking; this is what good parenting requires. So this case does not meet P4. But partners are not each other's parents. For this reason and others, conclusions about sulking within child-parent relationships are not straightforwardly parallel to conclusions about sulking within partner-partner relationships. (For what it is worth, unlike Ferzan, I am happy to hold that teens can sometimes undermine the consent of their parents via pressures like sulking. Indeed, holding this position seems to be part and parcel of recognizing that sometimes teens should be treated as full moral agents. Moreover, this position need not lead us to exaggerate the wrong of child-parent coercion or to deny that good parents are resilient to pressure. But this is a complicated matter better left for discussion elsewhere.) Thanks to an anonymous referee for asking me to elaborate on child-parent cases.

[78] Bennett and Jones, "45 Stories of Sex and Consent on Campus."

This paper started with a question: What could explain the wrongfulness of sulking for and into sex, without overgeneralizing? I have now given an answer that cuts against the existing literature. Sulking at someone for sex often involves *wrongfully blaming* them; sulking someone into sex often *coercively undermines their consent*. Such violations of sexual authority, importantly, should be morally but not always criminally sanctioned. These arguments avoid implying that all sexual pressures are the same. Nevertheless, they usefully extend to subtle sexual pressures besides sulking, overtly aggressive sexual pressures, and even nonsexual pressures.

I hope that I have demonstrated to philosophers that we should attend to real stories of how a particular sexual pressure unfolds within close relationships. Such attention helps us to identify the key moral features of the relevant pressure, especially when that pressure is subtle. I also hope that I have helped both sulkees and sulkers. Do the former understand their experiences better and feel more empowered to demand better treatment? Are the latter more equipped to follow the path of change that is described in Recognition? If so, I would be glad. Ultimately, what I hope to have illuminated is the following. Sulking is a complicated, seemingly paradoxical behavior of proximate withdrawal. In intimate relationships, sulking and blame form a fraught, potent pair. Sexual coercion therefore need not involve blatant threats of violence. Often, it operates via simmering absence, a withdrawal that pulls you in its wake.[79]

Macalester College
spatward@macalester.edu

References

Anderson, Ellie. "A Phenomenological Approach to Sexual Consent." *Feminist Philosophy Quarterly* 8, no. 2 (July 2022): 1–24.

Anderson, Scott. "Coercion." *Stanford Encyclopedia of Philosophy* (Spring 2023). https://plato.stanford.edu/archives/spr2023/entries/coercion/.

[79] For feedback on the ideas and drafts that developed into this paper, thanks to two anonymous reviewers, as well as Sam Asarnow, Sonu Bedi, Sarah Buss, April Conway, Emmalon Davis, Kristie Dotson, Tom Dougherty, Guus Duindam, Paul Garofalo, Lisa Gourd, Gillian Gray, Scott Hershovitz, Josh Hunt, Renée Jorgensen, Gabrielle Kerbel, Dan Lowe, Ishani Maitra, Cameron McCulloch, Gabriel Mendlow, Sarah Moss, Marcela Prieto Rudolphy, Laura Ruetsche, Tad Schmaltz, Mark Schroeder, Laura Soter, Angela Sun, Eric Swanson, Brian Talbot, Jamie Tappenden, Adam Waggoner, Brian Weatherson, Katie Wong, and audiences at the University of Michigan, the Minorities and Philosophy 2020 Summer Flash Talk Series, and Washington and Lee University.

———. “Coercion as Enforcement and the Social Organisation of Power Relations: Coercion in Specific Contexts of Social Power.” *Jurisprudence* 7, no. 3 (December 2016): 525–39.

———. “Conceptualizing Rape as Coerced Sex.” *Ethics* 127, no. 1 (October 2016): 50–87.

———. “The Enforcement Approach to Coercion.” *Journal of Ethics and Social Philosophy* 5, no. 1 (October 2010): 1–31.

———. “How Did There Come to Be Two Kinds of Coercion?” In *Coercion and the State*, edited by David Reidy and Walter Riker, 17–29. New York: Kluwer/Springer, 2008.

———. “Of Theories of Coercion, Two Axes, and the Importance of the Coercer.” *Journal of Moral Philosophy* 5, no. 3 (January 2008): 394–422.

———. “On Sexual Obligation and Sexual Autonomy.” *Hypatia* 28, no. 1 (Winter 2013): 122–41.

———. “Sex under Pressure: Jerks, Boorish Behavior, and Gender Hierarchy.” *Res Publica* 11, no. 4 (December 2005): 349–69.

Archard, David. “The Wrong of Rape.” *Philosophical Quarterly* 57, no. 228 (July 2007): 374–93.

Barbee, Anita, and Michael Cunningham. “An Experimental Approach to Social Support Communications: Interactive Coping in Close Relationships.” *Annals of the International Communication Association* 18, no. 1 (1995): 381–413.

Bennett, Christopher. “The Varieties of Retributive Experience.” *Philosophical Quarterly* 52, no. 207 (April 2002): 145–63.

Bennett, Jessica, and Daniel Jones. “45 Stories of Sex and Consent on Campus.” *New York Times*, May 10, 2018. https://www.nytimes.com/interactive/2018/05/10/style/sexual-consent-college-campus.html.

Boonin, David. *When Yes Means No: Problems of Sexual Consent*. Unpublished manuscript.

Cahill, Ann J. *Overcoming Objectification*. New York: Routledge, 2011.

———. *Rethinking Rape*. Ithaca, NY: Cornell University, 2001.

———. “Unjust Sex vs. Rape.” *Hypatia* 31, no. 4 (Fall 2016): 746–61.

Card, Claudia. “Recognizing Terrorism.” *Journal of Ethics* 11, no. 1 (March 2007): 1–29.

Carlsson, Andreas Brekke. “Blameworthiness as Deserved Guilt.” *Journal of Ethics* 21, no. 1 (March 2017): 89–115.

Conly, Sarah. “Seduction, Rape, and Coercion.” *Ethics* 115, no. 1 (October 2004): 96–121.

Dougherty, Tom. “Coerced Consent with an Unknown Future.” *Philosophy and Phenomenological Research* 103, no. 2 (September 2021): 441–61.

———. “Sexual Misconduct on a Scale: Gravity, Coercion, and Consent.” *Ethics* 131, no. 2 (January 2021): 319–44.

Edwards, James. “Theories of Criminal Law.” *Stanford Encyclopedia of Philosophy* (Fall 2021). https://plato.stanford.edu/archives/fall2021/entries/criminal-law/.

“Emotional Abuse in Sulking Silence when Sexual Demands Go Begging.” *News-Mail* (Bundaberg, Australia), March 2, 2009. https://infoweb.newsbank.com/apps/news/openurl?ctx_ver=z39.88-2004&rft_id=info%3Asid/infoweb.newsbank.com&svc_dat=WORLDNEWS&req_dat=6745D0C7EC9246F49DAE70DA1EA3845F&rft_val_format=info%3Aofi/fmt%3Akev%3Amtx%3Actx&rft_dat=document_id%3Anews%252F12797FAEDA676640.

Ferzan, Kimberly Kessler. “Consent and Coercion.” *Arizona State Law Journal* 50, no. 4 (Winter 2018): 951–1008.

Fricker, Miranda. “What’s the Point of Blame? A Paradigm-Based Explanation.” *Nous* 50, no. 1 (March 2016): 165–83.

Hieronymi, Pamela. “The Force and Fairness of Blame.” *Philosophical Perspectives* 18, no. 1 (2004): 115–48.

Hull, Richard T. “Have We a Duty to Give Sexual Pleasure to Others? A Reply to Arthur M. Wheeler.” Commentary presented at the Tri-state Philosophical Association Meeting, Mercyhurst College, Erie, PA, October 1985.

Ichikawa, Jonathan Jenkins. “Presupposition and Consent.” *Feminist Philosophy Quarterly* 6, no. 4 (December 2020): 1–32.

Kukla, Rebecca. “That’s What She Said: The Language of Sexual Negotiation.” *Ethics* 129, no. 1 (October 2018): 70–97.

Liberto, Hallie. “Coercion, Consent, and the Mechanistic Question.” *Ethics* 131, no. 2 (January 2021): 210–45.

———. *Green Light Ethics*. Oxford: Oxford University Press, 2022.

———. “The Problem with Sexual Promises.” *Ethics* 127, no. 2 (January 2017): 383–414.

———. “Threats, Warnings, and Relationship Ultimatums.” In *The Routledge Handbook of Love in Philosophy*, edited by Adrienne M. Martin, 128–37. New York: Routledge, 2019.

MacKinnon, Catharine. “Rape Redefined.” *Harvard Law and Policy Review* 10, no. 2 (Summer 2016): 431–77.

McDermott, Roe. “My Girlfriend Sulks If We Don’t Have Sex and It’s Bringing Back Painful Memories.” *Irish Times*, August 9, 2019. https://www.irishtimes.com/life-and-style/health-family/my-girlfriend-sulks-if-we-don-t-have-sex-and-it-s-bringing-back-painful-memories-1.3971846.

———. “My Partner Wants Sex Every Night and Sulks if I Don’t Agree.” *Irish*

Times, March 25, 2018. https://www.irishtimes.com/life-and-style/health-family/my-partner-wants-sex-every-night-and-sulks-if-i-don-t-agree-1.3438005.

McGeer, Victoria. "Civilizing Blame." In *Blame: Its Nature and Norms,* edited by D. Justin Coates and Neal A. Tognazzini, 162–88. Oxford: Oxford University Press, 2013.

Mendlow, Gabriel. "The Moral Ambiguity of Public Prosecution." *Yale Law Journal* 130, no. 5 (March 2021): 1146–87.

Miceli, Maria. "How to Make Someone Feel Guilty: Strategies of Guilt Inducement and Their Goals." *Journal for the Theory of Social Behaviour* 22, no. 1 (March 1992): 81–104.

Morgan, Robin. "Theory and Practice: Pornography and Rape." In *Going Too Far: The Personal Chronicle of a Feminist,* 163–69. New York: Random House, 1977.

Murphy, Clare. "Tactic #13: Intimate Partner Sexual Abuse." SpeakOutLoud, n.d. Accessed April 19, 2020. https://speakoutloud.net/intimate-partner-abuse/sexual-abuse.

Nozick, Robert. "Coercion." In *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel,* edited by Sidney Morgenbesser, Patrick Suppes, and Morton White, 440–72. New York: St. Martin's Press, 1969.

Parker, Steph, and Dom Parker. "Steph and Dom Solve Your Sex, Love, and Life Troubles: After 50 Years, He Still Wants Loads of Sex—But I Don't!" *Daily Mail,* September 6, 2020. https://www.dailymail.co.uk/femail/article-8703679/Steph-Dom-50-years-wants-loads-sex-dont.html.

Patwardhan, Sumeet. "Do I Have To? Moral Ignorance and Consent." Unpublished manuscript.

———. "Meddlesome Blame for Nonconsensual Sex." Unpublished manuscript.

———. "Peremptory Blame." Unpublished manuscript.

———. "Stalking by Withdrawing." Unpublished manuscript.

Price, Devon. "A Few Words about Sexual Coercion in the Wake of the Aziz Ansari Accusations." *Medium,* January 14, 2018. https://devonprice.medium.com/a-few-words-about-sexual-coercion-in-the-wake-of-the-aziz-ansari-accusations-7db015c1cde5.

Radzik, Linda. *Making Amends: Atonement in Morality, Law, and Politics.* Oxford: Oxford University Press, 2009.

Smith, Sharon, Xinjian Zhang, Kathleen C. Basile, Melissa T. Merrick, Jing Wang, Marcie-jo Kresnow, and Jieru Chen. "The National Intimate Partner and Sexual Violence Survey (NISVS): 2015 Data Brief (Updated Release)." National Center for Injury Prevention and Control, Centers for Disease

Control and Prevention, 2018.

Soble, Alan. *Sexual Investigations*. New York: New York University Press, 1996.

Srinivasan, Amia. "Does Anyone Have the Right to Sex?" *London Review of Books* 40, no. 6 (March 2018). https://www.lrb.co.uk/the-paper/v40/n06/amia-srinivasan/does-anyone-have-the-right-to-sex.

Stemple, Lara, Andrew Flores, and Ilan H. Meyer. "Sexual Victimization Perpetrated by Women: Federal Data Reveal Surprising Prevalence." *Aggression and Violent Behavior* 34 (May 2017): 302–11.

Sugar, Marcy, and Kathy Mitchell. "Sulking for Sex." Advice column, *Annie's Mailbox*, November 16, 2012. https://www.creators.com/read/annies-mailbox/11/12/sulking-for-sex.

Thomason, Krista. "Shame, Violence, and Morality." *Philosophy and Phenomenological Research* 91, no. 1 (July 2015): 1–24.

Tognazzini, Neal, and D. Justin Coates. "Blame." *Stanford Encyclopedia of Philosophy* (Summer 2021). https://plato.stanford.edu/archives/sum2021/entries/blame/.

Wertheimer, Alan. *Coercion*. Princeton: Princeton University Press, 1987.

———. *Consent to Sexual Relations*. Cambridge: Cambridge University Press, 2003.

Yap, Audrey. "Conceptualizing Consent: Hermeneutical Injustice and Epistemic Resources." In *Overcoming Epistemic Injustice: Social and Psychological Perspectives*, edited by Benjamin R. Sherman and Stacey Goguen, 49–62. London: Rowman & Littlefield International, 2019.

Zimmerman, David. "Coercive Wage Offers." *Philosophy and Public Affairs* 10, no. 2 (Spring 1981): 121–45.

Journal of Ethics and Social Philosophy
vol. 28, no. 3 · october 2024

https://doi.org/10.26556/jesp.v28i3.3068

ATTRACTION, AVERSION, AND MEANING IN LIFE

Alisabeth Ayars

The state that philosophers call "desire" comes in two kinds: attraction and aversion. When we are attracted to something, we are "pulled toward" it: we regard it in a positive way. (Think of the desire for a delicious meal or the desire to view a great work of art.) When we are averse to something, we are "pushed away" from it: we regard it in a negative way. (Think of the desire not to be rejected or not to be covered in spiders.)

Writers in the tradition routinely marked the distinction, often writing as if "desire" and "aversion" (or "love" and "hatred") were a pair of distinct attitudes that together supply the fuel for activity fueled by passion.[1] But contemporary theories of desire have paid scant attention to the distinction. Some philosophers are skeptical that the distinction exists at all. Is there really a difference between, say, being attracted to fame and being averse to ordinary anonymity?[2] Descartes did not think so:

> I know very well that in the schools, that passion which tends to the seeking after good, which only is called desire, is opposed to that which tends to the avoiding of evil, which is called aversion. But seeing there is no good, the privation whereof is not an evil, nor any evil taken in the notion of a positive thing the privation whereof is not good. For example, that in seeking after riches, a man necessarily eschews poverty; in avoiding diseases, he seeks after health; and so of the rest.[3]

[1] See, for example, Hume, *Treatise of Human Nature*, 2.3.3.3.

[2] According to Sumner, for instance, what I am calling "attraction" and "aversion" are really just two ways of representing the same attitude. A negative desire that [It does not rain this afternoon], says Sumner, can be represented either as an "aversion" to [It rains this afternoon] or an "attraction" to [The weather is dry this afternoon]. But "all three of these alternatives come to the same thing: that is, your positive desire is satisfied, your negative desire is satisfied, and your aversion is frustrated by exactly the same state of affairs (a rain-free afternoon).... Nothing seems to be gained by introducing the negative element" ("The Worst Things in Life," 428–29). Kagan has also rejected the relevance of the distinction for theories of well-being: Kagan, "An Introduction to Ill-Being," 270–71.

[3] Descartes, *Passions of the Soul*, pa a.87.

Moreover, even if the distinction is real, it is not obvious why we should care about it. It is arguably irrelevant to the empirical explanation of action, since for that purpose an undifferentiated notion of gradable desire (or preference) seems to suffice. Regardless of whether one is "attracted" to fame or "averse" to anonymity, one prefers fame to anonymity, and this preference may suffice to explain why one pursues fame as one does. And for the same reason, the contrast may be irrelevant to the normative theory of rational choice. Subjective utility and its variants are defined in terms of an undifferentiated notion of preference, so if the theory of rational choice tells us to maximize utility, it may not care whether the preferences it takes as input amount to desires or aversions.

I argue that one reason to think there is a difference between attraction and aversion, and to care about the difference, is that attractions and aversions contribute in radically different ways to our well-being. Attractions play an essential role in the good life; in particular, they are critical to the experience of meaning in life. By way of preview, consider a predominantly aversion-driven life—the life of, say, a college professor who is motivated to perform well primarily by an aversion to failure and indictment rather than any positive attraction to the elements of her job. She shows up to teach only to avoid getting fired; grades her students' work only to avoid their anger; does her research only because she fears insignificance; and so on. Every action is taken only to avoid something worse. Notably, her life may have a high level of desire-satisfaction overall; we may suppose her desires are exhausted by various aversions, and that her aversions are all satisfied in the end.[4] But clearly, something is missing. Such an aversion-driven life feels grey and meaningless at best (filled with anxiety and desperation at worst). It is natural for her to wonder even as she is making progress in fending off the objects of her aversions: "What is the point of all this? Once I have averted the evils of failure and indictment, then I will be left with . . . what?"

Contrast this life with the life of the professor who is genuinely attracted to aspects of her work; the professor who does her job because she is pulled forward by the appealing prospect of a job well done. This professor's life may not be perfect, but it will not strike her as empty in the same way. For someone's life to feel meaningful to her, I argue, she must be genuinely attracted to what she

[4] Like Pallies does in "Attraction, Aversion, and Asymmetrical Desires," I call aversion "satisfied" if the state of affairs to which the relevant person is averse does not obtain. So an aversion to being covered in spiders is satisfied insofar as one is *not* covered in spiders and frustrated insofar as one is. One could sensibly adopt the opposite terminological convention of calling an aversion "satisfied" if the state of affairs one is averse to does obtain, as Kelley does in "Well-Being and Alienation" and Heathwood does in "Ill-Being for Desire Satisfactionists."

is pursuing, not merely averse to the alternatives (or lacking in affective desire altogether). If that is right, there must be a real difference between attraction and aversion.

Moreover, the distinction must matter for philosophy. Our two professors differ in well-being because the desires that move them differ in "quality." The theory of well-being thus needs the attraction/aversion contrast even if other parts of philosophy and psychology do not.

But then we want a theory of the distinction—some account of how being attracted to *p* differs from being averse to not-*p*—that puts us in a position to understand the distinctive connection between attraction and this aspect of well-being. What is it about the nature of attraction that explains why attraction-driven activity is valuable? I sketch a theory that illuminates the contribution of attraction-motivated activity to felt meaning.

1. Attraction, Aversion, and Well-Being

It is not hard to glom onto the distinction between attraction and aversion using examples. Consider two ways of being at a party. You see someone across the room, are drawn to them, and approach them; alternatively, afraid of seeming antisocial, you approach them. In both cases, you walk across the room because you want to talk to the person on the other side. But in the first case, you do so because you are attracted to talking to them (and to what might happen if you do), whereas in the second case, you do so because you are averse to not talking to them (and to what might happen if you do not). As Sinhababu observes, attraction and aversion are associated with different emotional syndromes:

> Some desires, like the desire for a delicious meal, give us a delighted happy feeling when we find that we can satisfy them and an unpleasant feeling of disappointment when we discover that we cannot. Others, like the desire not to miss one's flight, give us the pleasure of relief when we find that we can satisfy them and an unpleasant feeling of anxiety or dread when we discover that we cannot. This gives us reason to divide the category of desire into two subcategories, positive desire and aversion.[5]

Of course, not every desire is easily sorted into one of these two bins. Many real desires are mixtures of attractions and aversions. For instance, my motivation to do a good job teaching my classes this semester combines my positive regard for some aspects of the job (benefiting my students, doing a job I can be

[5] Sinhababu, "The Humean Theory of Motivation Reformulated and Defended," 490.

proud of) and my negative regard for others (harming my students, doing a job I would be ashamed of). Just as the net force acting on an object is the vector sum of all the forces acting on it, which may point in opposite directions, the total strength of a preference for *p* is the sum of one's attractions and aversions to features of *p* and the alternatives (along with nonaffective sources of desire if such there be). And in normal cases there will be forces of both sorts.

For the purpose of explaining action, it may be that all that matters is the strength of one's preference for performing the action over the alternatives, regardless of how the aversions and attractions combine to produce this preference. Indeed, the action in both versions of the party scenarios (walking across the room) can be explained by the existence of a preference to speak to the stranger on the other side and a relevant means-end belief, without invoking the presence of an attraction or aversion specifically. Still, there seems to be a difference in the kind of desire that motivates in each scenario.

But there is a reason for insisting that the distinction is real that goes beyond the fact that it certainly seems real on reflection. To see this, we turn our attention to an area of philosophy in which it clearly makes a difference: the philosophy of well-being. The stark contrast between an aversion-driven life and an attraction-driven life indicates that attraction and aversion contribute in radically different ways to well-being.[6] But how do we characterize the contribution that attractions make to well-being in positive terms?

Two ideas may come to mind. Perhaps there is something valuable about attraction *satisfaction*, compared to aversion satisfaction. Or perhaps the attraction-driven life normally contains more pleasure than the aversion-driven one. I will argue that neither of these proposals fully captures the positive contribution of attractions to well-being.

According to the first proposal, having attractions and satisfying them is intrinsically good for us in a way that satisfying aversions is not. This thesis—a modification of the desire-satisfaction theory of well-being—has recently been developed by Daniel Pallies, who argues for one of the conclusions I will be defending: that attraction and aversion must be psychologically real given their

[6] In making this claim, I am adding to a growing chorus of philosophers who argue that the difference between attraction and aversion is real and matters for the philosophy of well-being. See Pallies, "Attraction, Aversion, and Asymmetrical Desires"; Heathwood, "Ill-Being for Desire Satisfactionists"; Kelley, "Well-Being and Alienation"; and Mathison, "Asymmetries and Ill-Being." However, these philosophers have focused primarily on the relevance of the distinction for the desire-satisfaction theory. I aim to show that the distinction matters for theories of well-being that emphasize meaning in life as a dimension of prudential value.

different contributions to well-being.[7] According to Pallies, while satisfying attractions is intrinsically good for us, having aversions and satisfying them is merely not bad for us (though failing to satisfy them is positively bad).

Pallies's thesis can be spelled out as follows. Suppose you start off indifferent to whether *p* in a world in which *p* is true. If you then come to be attracted to *p*, your satisfied attraction adds to your well-being relative to this baseline. For example, if you are attracted to being famous and achieve fame, this is better than being indifferent to fame and nonetheless achieving it. In contrast, a satisfied aversion to *p* adds nothing to your well-being relative to a baseline of indifference to *p*. If you are averse to being covered in spiders, and you are not covered in spiders, then this is no better for you than if you were indifferent to being covered in spiders (and not covered in spiders). In other words, satisfying an aversion cannot raise your well-being above 0; it can only keep you out of the negative range, whereas satisfying an attraction can take you into positive territory, assuming that indifference constitutes neutrality.[8]

Pallies's proposal can explain why the aversion-driven life is low in well-being: the person's desires are all aversions, so their satisfaction does not add positive well-being to the life. Pallies's proposal is moreover plausible (*modulo* the usual reservations about any desire-satisfaction theory of well-being), and if it is correct, it provides an excellent reason for believing in the reality of the distinction and seeking an account of what it comes to. But it does not tell the whole story regarding the contribution of attractions to well-being. Pallies's theory is a desire-satisfaction theory; it explains why *satisfied* attractions contribute distinctively to well-being. But attractions contribute to well-being in ways that do not depend on whether we secure the object of our attraction, whereas aversions (satisfied or otherwise) cannot play this role. Pursuing attractions (but not aversions) contributes to well-being in a way that is not reducible to the value of desire satisfaction.

To see this, consider normally attraction-driven pursuits like preparing a delicious meal, working on a novel, solving a deep philosophical puzzle, or bringing about an attractive moral ideal. (To be clear, such pursuits could be motivated entirely by aversion to the absence of the good, but let us imagine attraction-driven versions of them.) The well-being contributed by these

[7] Pallies, "Attraction, Aversion, and Asymmetrical Desires."

[8] Interest in the distinction between attraction and aversion for the desire-satisfaction theory arose in part from the need to develop an account of ill-being, as emphasized by Kagan in "An Introduction to Ill-Being," 263. See Heathwood, "Ill-Being for Desire Satisfactionists"; Kelley, "Well-Being and Alienation"; and Mathison, "Asymmetries and Ill-Being" for aversion-based accounts of ill-being within the context of the desire-satisfaction theory, developed in response to Kagan's challenge.

projects does not accrue to us only after the attraction is satisfied. Rather, the pursuit of the goal one finds attractive is already intrinsically good for us. One benefits simply from being "pulled forward" by an attractive vision—the creation of the delicious meal, solving the deep puzzle—even before the attractive goal is realized—indeed, even if it is never realized. It is true that such pursuits often involve episodes of local attraction-satisfaction, as one makes progress. But the value of the pursuit is not reducible to these episodes of local satisfaction. Attraction-driven pursuits contribute to our well-being even when we are in between such episodes.[9]

In contrast, pursuits motivated by aversions do not intrinsically contribute to well-being in this way. We do not get a positive welfare boost from aversion-driven pursuits like making a divorce as painless as possible, ensuring that no shame is ever brought to our family, or working to solve a persistent health problem. These things feel like grim chores. Of course, since it is often better for us if an aversion is satisfied rather than not satisfied, it is instrumentally valuable to pursue its satisfaction. But the pursuit is not intrinsically good for us in a sense in which attraction-motivated activity palpably is.

It should be acknowledged that not all attraction-motivated pursuits make a net positive contribution to well-being. In some cases, it would be better for us on the whole if we could rid ourselves of an attraction—e.g., if the attraction is too obsessive or if it has no hope of being satisfied. (Someone who is hopelessly pursuing the dream of becoming a famous athlete but perceives no progress toward the goal and does not find training rewarding would probably be better off without the attraction and the activity it motivates.) The claim is rather that under certain conditions—when the attraction is not too obsessive, when we make consistent progress, when we experience episodes of hope, and so on—attraction-motivated activity benefits us in a way that is not reducible to the benefit that would accrue from satisfying the attraction. Pallies's modified desire-satisfaction theory of well-being cannot fully capture this distinctive contribution of attractions to well-being.

Another time-honored strategy for explaining how and why attraction contributes distinctively to well-being points to an alleged connection between attraction and pleasure or enjoyment. It may well be true in general that we take pleasure in the pursuit (and attainment of) ends to which we are attracted and that we take less pleasure in avoiding what we find aversive. But attraction contributes to well-being in ways that this observation cannot explain.

[9] Indeed, in some cases, securing the aim can be in tension with the relevant good, if there is nothing left to pursue or maintain. Thanks to an anonymous reviewer for pointing this out.

To see this, observe that while attraction-driven pursuits are often pleasurable, they can also be associated with significant stress and even boredom. It is often clear we could obtain more pleasure by doing something else. A student pianist's long hours of repetitive practice may bring him some pleasure as he notices the progress he is making, but he could certainly accrue more pleasure in the same amount of time by going on vacation. Yet the pursuit of the end he finds attractive is intrinsically valuable for him *in a way* that the idle pleasure of the vacation is not, even during the stretches in which he accrues little pleasure from the pursuit. The same can be said of stressful or difficult pursuits like climbing a mountain, understanding a complicated piece of philosophy, or pursuing success or fame. These activities are not always pleasurable in the moment (and are sometimes positively unpleasant); but they are valuable whenever they involve being drawn forward by the attraction.

Another way to develop the point is to note that an attraction-driven life that lacks much pleasure need not be empty or tedious in the way the aversion-driven life is. Consider the stereotypical "tortured artist" who is intensely attracted to creating a great work of art but is constitutionally melancholic. While it would be better if she enjoyed her pursuit, her life is not empty. She is drawn forward by an attractive vision and hence has something the aversion-driven professor does not. The question that arises in the case of aversions—"What is the point of all this? Once I have avoided the greater evil, then I will be left with … what?"—does not arise for her: her purpose is to bring about something of positive value, which goes beyond (let us assume) the privation of evil; and insofar as she is moved by this purpose, her striving will seem to her to have a point.[10]

Of course, it is hard to see how an attraction-driven pursuit that involves no pleasure at all could be good for a person. Suppose someone spends hours and hours training to be an athlete but never gets better; suppose she does not intrinsically enjoy the training and does not indulge in any pleasurable fantasies. At some point, this pursuit no longer adds to her well-being. And this is not just because the positive contribution is outweighed by the negative—rather, it stops generating positive value at all. Does this suggest that attractions must generate some pleasure to be a source of value—at least, the pleasure derived

[10] What is more, an aversion-driven life need not be lacking in pleasure. The pursuit of aversions can be associated with a kind of pleasure: the pleasure of relief. See Sinhababu, "The Humean Theory of Motivation Reformulated and Defended," 490. Fearful of failure, each sign that I am likely to succeed brings me a pleasurable episode of relief. But of course, the addition of many such episodes of pleasurable relief does not eliminate the grimness of a life driven entirely by aversions.

from the moments in which the aim and one's efforts come together in a single consciousness or in which one vividly experiences the appeal of one's goal?[11]

Even if such episodes of positive affect are necessary for the pursuit of an attraction to have value (*pace* the moral of the tortured artist case), this would not entail that the value of the pursuit is reducible to these episodes. As just noted, sometimes many hours of arduous training, often unpleasant, are required before the activity becomes enjoyable; but this sort of disciplined activity, driven by attraction, can be good for a person in a distinctive way well before she finds it pleasant.

Moreover, we can explain why such moments may be an essential component of valuable attraction-motivated pursuits without invoking pleasure. We might say: episodes of positive affect play an important epistemic role. Arguably, during such moments, one is vividly aware in a quasi-perceptual way of the goodness of the object of attraction. (I say *quasi*-perceptual since the object of attraction is normally an as-yet-nonexistent state of affairs, in which case there is no question of literally perceiving its goodness. Yet the state is perception-like in being a phenomenologically vivid presentational state that is distinct from any judgment or belief we might form about its content but that nonetheless normally informs a belief that matches it in content.) When we see a surface as red, we normally take it to be red if the question arises; likewise, when we experience positive affect toward a prospect, we experience it as good and so normally take it to be good if the question arises. These episodes are thus a source of confidence if we come to question the positive value of what we do, even if they are not themselves a source of value.

These considerations show that attractions contribute to well-being in a distinctive way that is not fully captured by Pallies's modified desire-satisfaction theory of well-being or by the fact that activity motivated by attraction is (sometimes) enjoyable. This gives us strong preliminary reason to believe in the reality and importance of the distinction—that the "seeking of the good," whatever this amounts to, must be distinct from the "avoiding of evil"—and to seek a new account of the distinctive value of an attraction-driven life.

2. Attraction and Meaning

Let us start with a clearer characterization of what it is exactly that is good for us when we are pulled forward by attraction. This much is apparent from the preceding discussion: it is not simply *having* attractions that is valuable; the attraction must motivate activity. The value is realized by doing things aimed at

[11] Thanks to Daniel Pallies for pressing me on this point.

furthering an attractive vision. Someone who is attracted to a goal but cannot (or will not) do anything about it and so sits back and waits to see what happens does not attain the prudential benefit.

What is distinctively good for us in the cases we have been discussing is *attraction-motivated activity*: activity that is motivated by our experience of the appeal of something, which is accompanied by episodes of hope that the attraction will be satisfied, perceived progress, and so on. The prudential value of attraction consists in being *drawn forward* by an attractive prospect.

There may be limiting cases in which an attraction contributes to well-being without motivating activity. Suppose that someone is attracted to a particular country's winning a war or the success of a certain sports team. Because the outcome is outside of his control, he can do nothing to further the attraction. But he follows the events closely in the newspapers, is conscious of progress toward the goal, and experiences episodes of hope. This sort of engagement may be meaningful in a sense, but this is because it inherits many of the features of attraction-motivated activity by virtue of his identification with the country or sports team. The attraction still involves a kind of "forward motion," experienced vicariously through the efforts of the country or team, and is hence very different from the case of an idle attraction that affords no opportunity for progress. The good is not just that the Yankees win; it is that one experiences a Yankee win, and the typical fan is active to some extent in pursuit of that goal by watching the game, attending closely to it, etc.[12]

To be clear, it is not strictly necessary that the good lies in the future for attraction-motivated activity to have this distinctive value. The prudential benefit of pursuing an attraction can be attained, for example, by maintaining an existing state of affairs rather than pursuing one that is yet to be, as when one works to maintain a valuable relationship. Such activity still aims at the good, by preserving it. Nevertheless, one must see oneself as *involved* in the good in some way, whether it be its acquisition, promotion, or sustenance.

So far we have been speaking in general terms about the prudential "value" of activity fueled by attraction. My more specific hypothesis is that the contribution of this activity to well-being is best captured using the language of

[12] We might elaborate the link between valuable attraction and activity as follows. As human beings, we are condemned to act; action and choice are unavoidable aspects of the human condition. The value of attraction lies in what it makes possible regarding the teleological structure of action—what it permits regarding the purpose or reason for which we act. Assuming that attraction involves seeing its object as good in some way, attraction allows us to act for the sake of the perceived good, not merely the lesser evil; and this goes some way in resolving the distinctive malaise associated with one's activity feeling "pointless" in some hard-to-pin-down sense. Exactly what this amounts to remains to be seen, but it seems to capture the essential idea.

meaning. In characterizing the predominantly aversion-driven life, we naturally reach for this language. A life spent in pursuit of the lesser evil is not just lacking in pleasure; it feels "meaningless," "gray," and "empty." The experience of meaning in life requires activity fueled by a passion that leads us to seek the good, as it were, not merely the avoidance of evil.

Many philosophers have argued that felt meaning, as distinct from pleasure, is an important component of the good life. A life of idle amusement may be enjoyable, but a person living such a life may reasonably feel unfulfilled. "Subjective meaning" refers to the state of mind that is conspicuously missing in such cases. Subjective meaning is a subjective good (or the subjective component of a hybrid good)—a good that is (at least partly) realized in conscious experience. As Wolf has observed, when we complain of lacking meaning, we are often expressing dissatisfaction with the subjective character of our lives:

> When thinking about one's own life … a person's worry or complaint that his life lacks meaning is apt to be an expression of dissatisfaction with the subjective quality of that life. Some subjective good is felt to be missing. One's life feels empty.[13]

Pleasure is not sufficient for subjective meaning, nor is it necessary. A person's life may feel fulfilling even if it lacks much pleasure (as in the case of the tortured artist). The experience of meaning is a subjective good, but it is not good in virtue of its hedonic quality.

Among those philosophers who believe that subjective meaning is a component of the good life, many claim that what is required for subjective meaning is being sufficiently absorbed in, gripped by, or passionate about one's projects. For example, Taylor emphasizes passionate desire, noting that if the gods were to inject Sisyphus with some substance that gave rise to an obsession to roll stones, his life would become meaningful in the only possible sense—it would be meaningful for him.[14] Wolf emphasizes "active engagement" in projects of worth.[15] And Kauppinen emphasizes goal-directed activity by seeing meaning as a function of the structure of the agent's goal-directed activities: "Life is ideally meaningful when challenging efforts lead to lasting successes."[16]

But I claim that passionate involvement in projects is not enough for subjective meaning, since passionate involvement can be fueled entirely by aversion, and when it is, it does not bring fulfillment. Someone might have an intense

[13] Wolf, *Meaning in Life and Why It Matters*, 11.

[14] Taylor, *Good and Evil*.

[15] Wolf, "Happiness and Meaning."

[16] Kauppinen, "Meaningfulness and Time," 346.

aversion to failure and insignificance that motivates intense engagement in writing a book. Night after night, she works on the book, editing and rewriting, fearful that her efforts will come to nothing. Though she is absorbed in and passionate about her project and not at all "bored" in the traditional sense, there is still a grayness to her pursuit: she could reasonably complain that her work feels meaningless. ("And when I'm done, what then? I will have avoided the evils of failure and insignificance only to find myself with … what?")

Of course, as this example shows, attractions are not necessary for being busy; an aversion-driven person might be thoroughly busy avoiding what she is averse to, like someone constantly running from a tiger. Her life is thus not "empty" or "boring" in one sense. But it can still *feel* empty and unfulfilling. This kind of emptiness is associated with existential unease or deep boredom—the unease we voice by asking, "What's the point of all of this? Why not surrender to the tiger and get it over with?" This is the kind of questioning that an aversion-driven life prompts.

This deep boredom is nicely expressed by Maria von Herbert in a letter to Kant, where she describes an unbearable emptiness resulting from a lack of attraction:

> I feel that a vast emptiness extends inside me, and all around me—so that I almost find myself to be superfluous, unnecessary. Nothing attracts me. I'm tormented by a boredom that makes life intolerable. Don't think me arrogant for saying this, but the demands of morality are too easy for me. I would eagerly do twice as much as they command. They only get their prestige from the attractiveness of sin, and it costs me almost no effort to resist that.[17]

Without attraction, von Herbert felt "superfluous," "unnecessary," and deeply bored. This is another way of saying that her life felt meaningless; and it was meaningless in virtue of the fact that nothing *attracted* her. The problem could not be fixed by giving her new aversions and the opportunity to satisfy them, even if they were to generate passionate involvement in the avoidance of the bad.

It is important to stress that attraction-driven activity does not just forestall existential boredom and so *prevent* a bad. That is consistent with its being of no positive welfare value in itself. The claim is that attraction-driven activity blocks existential boredom by replacing it with its opposite: positive engagement with one's life and its content, a kind of positive motivational interest. This kind

[17] As quoted in Langton, "Duty and Desolation," 493.

of motivational interest is not always pleasurable but can generate subjective meaning even when it is not pleasurable.

But why exactly should pursuing attractions be associated with meaning in a way that pursuing aversions is not? The relation between attraction and felt meaning becomes clearer if we take seriously the idea that attraction is a passion that tends toward the "seeking after good" and aversion only toward the "avoiding of evil," *and* we provisionally assume (*contra* Descartes) that the two are not equivalent. One way to spell out this idea is to say that attraction and aversion have positive and negative normative content, respectively. When we are attracted to something, we see it as good, and when we are averse to something, we see it as bad. The experience of meaning in life seems to have something to do with connecting to positive value. If attractions represent their objects as positively good, it is not mysterious why pursuing attractions brings meaning: in pursuing attractions we are drawn forward by the perceived goodness of our end, something that gives us a reason to be glad to be alive. The goodness we see beckons us, pulling us forward and imbuing our activity with a positive point.

Since aversions do not represent their object as positively good, rather only the alternatives as bad, aversion-motivated activity does not make us feel connected to any positive value. In a case of pure aversion, the best-case scenario is that we succeed and preserve a situation that we take to be the "zero point"—a state of affairs about which nothing positively good can be said. The "emptiness" of such a life is the felt detachment from positive value: one sees the world as devoid of opportunities to involve oneself with goodness.

Consider that when we see ourselves as pursuing a prudential good, we see ourselves as creating or sustaining value that redeems our existence to some extent. A purely aversion-driven life prompts a certain kind of questioning: we wonder what the "point" of it all is. ("And once I have avoided the evils of failure and misery, then I will find myself with . . . what?") What we seek in this questioning is a reason to exist or to be glad that one exists and will exist. But so long as only personal or prudential value is on the scene, only positive goods can do this. A life devoted entirely to preventing prudential bads involves nothing that would constitute a positive reason to carry on or to be glad that one exists. The avoidance of various forms of badness is not something that makes life worth living; it is at best neutral. (We can avoid them simply by ceasing to exist; hence, they give us no reason to carry on.) This is why von Herbert's existence struck her as "superfluous": bereft of attraction, she had nothing that constituted a reason to go on.

Of course, not all attractions are directed at prudential goods, and not all aversions are directed at prudential bads. We can be attracted to world peace or averse to general ill-being. The story I just told is not straightforwardly applicable in such cases, since the elimination of an objective bad may be something that does give us reason to go on, since our existence is necessary to eliminate it. I will have more to say about this shortly. First, let me address a different objection that may arise.

The preceding argument seems to imply that pursuing attractions is *always* meaningful. If seeking after the perceived good is what brings us meaning, and pursuing attractions always involves pursuing a perceived good, then pursuing attractions should always generate some meaning. But this may seem implausibly strong. When someone is attracted to the prospect of eating a delicious ice cream cone, and this motivates her to seek one out, she achieves nothing, it seems, that merits the name "subjective meaning" or "fulfillment."[18]

But why not think the pursuit of the ice cream brings a *little* meaning, however trivial? Of course, a person who worries or complains that his life lacks meaning will not be relieved of the concern simply by pursuing ice cream. But this is because meaning comes in degrees, and this pursuit is hardly enough to take a life from "meaningless" to "meaningful." Still, a person who is genuinely attracted to ice cream has *something* that is lacking in a purely aversion-driven life. Someone who transitions from a deep depression in which absolutely nothing attracts him to a state in which he once again can appreciate the goodness of ice cream experiences a small gain in subjective meaning: his life is a bit brighter than it was before. Small attractions, we might say, can form the building blocks of meaning. Moreover, they can burgeon into larger attractions—e.g., becoming a connoisseur of ice cream or setting up an ice cream shop—that bring more substantial gains in meaning.[19]

But there is a more serious objection to the proposal that only attraction-motivated activity can contribute to meaning, related to the point about objectivity just discussed. When we think of lives that must feel meaningful to the people living them, we often think of people who are fighting against some real evil in the world. Activism—aimed at, say, eliminating poverty or animal abuse—seems to be a paradigmatic sort of meaning-generating activity. Unlike the person who is averse to failure and insignificance, the activist seeks to eliminate an objective bad. He thus has a reason to go on and even to be glad that he exists, since his existence may help eliminate the evil. But if paradigmatic activism is aimed at eliminating a bad in the world, and paradigmatic activism is aversion driven, this suggests that we can get meaning from aversion-driven projects after all.

[18] Thanks to an anonymous reviewer for pressing this objection.

[19] If one is unconvinced by this example, one could interpret the main thesis of the paper as the claim that attraction-driven activity is necessary though insufficient for subjective meaning.

The first thing to note is that paradigmatic activism is not motivated solely by aversion. Granted, activists are normally strongly averse to the bad they are trying to eliminate. But real activists—even if their activism is fundamentally aimed at eliminating a bad—are normally at least partly motivated by attraction. They are attracted to things like making a difference, the rewarding social relationships that activism affords, and the positive goods that eliminating the bad will enable. Antisegregationists, for instance, were motivated at least in part by a positive attraction to a world without racism; Martin Luther King Jr.'s "I Have a Dream" speech expresses this attractive vision. One way to see this is to note that the desire that motivates the paradigmatic activist would not be satisfied if the world were simply to painlessly end. That would secure the end of all injustice (and other bad things), but it would not secure the positive good she is really after.

So, paradigmatic activists *do* see their efforts as bringing about attractive ends, even if their primary focus is eliminating a bad. And these perceptions are appropriate: making a difference is positively good and hence a worthy object of attraction, even if the mere absence of suffering and injustice is not.

My account does, however, entail that someone whose activism is *purely* aversion driven does not acquire subjective meaning from it. And one might continue to insist that this is implausible, insofar as objective evils are on the scene. But, I argue, once we properly imagine the perspective of a purely aversion-driven activist, we have no trouble seeing the sense—or at least *a* sense—in which his activity must feel "meaningless" to him.

The perspective of a purely aversion-driven activist is hard to imaginatively occupy. We must imagine him as a grim character. He is not attracted to making a difference, to communion with fellow activists, or to the better world of peace and justice that will be realized if his activism is successful. This sort of activist is rare, if he exists at all. But we can imagine what his life would be like if he existed. The activist may believe his activism is worthwhile; he may even believe intellectually that his life is "meaningful" in virtue of the difference he is making. But he sees no positive value in his efforts; he is not moved by love of good, only hatred of evil. And this is, in an obvious sense, grim. We would have no trouble understanding him if he complained, "I know I'm making a difference, but my activism does not feel meaningful; to me it feels like a grim chore."

So, when we properly imagine the activist as only aversion driven, the sense that his life must feel meaningful begins to evaporate. Yet as noted earlier, the purely aversion-driven activist is unlike the person motivated by fear of failure and indictment in one respect: the activist has a desire that gives him a reason to go on, deriving from the necessity (or causal relevance) of his existence to the elimination of the bad. And we might wonder about the significance of this factor. Could it suffice to make the pursuit feel meaningful?

I say no. Though the activist's desire is categorical in this sense, his pursuit prompts a questioning akin to that mentioned earlier, this time not about his own life but rather about the world as a whole: "And once all of the suffering and injustice in the world have been removed, we will be left with ... what?" The best-case scenario is that he succeeds and preserves a *world* that he takes to be at the neutral point, about which nothing positively good can be said. And this gives the sense in which his activity remains "pointless." Of course, it is not literally pointless: its purpose is to make the world better. But it is pointless in the sense that even if he succeeds, he will have brought about nothing *good*, nothing that should make him glad that the world exists at all. He sees the universe and all the activity within it as ultimately superfluous. What we seek in this questioning is something that redeems the universe's existence to some extent, something that should make us glad that it (with us in it) exists at all.

When we pursue what we see as impersonally good, such as a world of peace and justice, we *do* see ourselves as creating or sustaining value that redeems the universe's existence. Peace, justice, art, beauty, and pleasure all constitute pockets of redeeming value in the universe; their existence is "something to be said" for the universe, unlike the existence of pockets of empty space that contain no suffering.[20] An activist partly motivated by attraction thus has an answer to the question: "And once all of the evil is gone, we will be left with ... what?" She can say: "We will be left with a more just world, a more beautiful world." And having an answer to this question is, I contend, irreducibly connected to our experience of a pursuit as meaningful and fulfilling, though there may be nothing more we can say as to why this is the case.

In other words, the deep reason why only attraction-motivated activity contributes to felt meaning is that meaning requires seeing ourselves as bringing about pockets of redeeming value in our own lives or in the universe—value that renders the universe and the activity within it nonsuperfluous—and only attraction-motivated activity affords this.

Another question may be raised at this point. I have claimed that only attraction-motivated activity creates the experience of meaning; the implied contrast is with aversion-motivated activity. Yet we might wonder about a slightly different case: someone who has no attractions but believes and even knows intellectually that her activities are positively good. She is thus unlike the activist who sees himself only as removing evils. She has a positive good in view—say, a world of peace and justice—and so does not take herself as trapped in a grueling cycle of only removing what she sees as bad. Her activity has the correct

[20] Of course, one could be attracted to such pockets of empty space, in which case one does see them as having redeeming value.

teleological structure; but the aim is merely *believed* good, rather than seen as good. Is this pursuit subjectively meaningful? If so, this suggests that it is not attraction that is necessary for meaning but merely the belief that one's aim is positively good.

But while evaluative beliefs of this sort may provide an intellectual sense that one's life is "meaningful," they are not enough to make one's activity feel meaningful in the sense I am trying to capture. Someone who believes at an intellectual level that her activity is positively good may still *feel* empty.[21] There is thus a subjective good that goes missing when the only connection to the positive good for which one acts is intellectual: one is "numb" to the value that one believes (and perhaps even knows) to exist. Just as one can believe that a painting is beautiful without seeing it as beautiful, one can believe that a state of affairs is positively good without seeing it as such—and be detached from its goodness in this sense.

When we yearn for meaning in our lives, what we want is for the appeal of something to impact us or strike us—for it to enter our experience. Attraction, as I understand it, is the state that secures this felt connection between our activity and the positive good for which we act; when we are attracted to something we *see* it as good. Attraction is thus often necessary to sustain our confidence in the positive value of what we do. When we see a prospect as good, we normally take it to be good if the question arises, just as when we see a surface as red, we normally take it to be red if the question arises. A person who has only evaluative beliefs lacks this perceptual source of confidence; there may even be a sense in which she cannot fully or completely believe in the goodness of what she does. Someone who manages to be confident that her activity is good despite having no experience of the good may have a pale version of the good that comes with genuine attraction, but it will not be as good as someone who is acquainted with the value she is pursuing.

I have argued that attractions play a central role in the experience of meaning, which is distinct from pleasure or desire satisfaction. Before moving on to a more detailed account of attraction and aversion, a few clarificatory remarks are warranted. First, I do not claim that someone who lacks attraction lives a meaningless life in every sense. My concern here is only with subjective meaning: what makes our lives feel meaningful *to us*. Some philosophers hold that meaning in life is fully subjective.[22] Some hold that meaning in life has subjective and objective components, and one might conceivably hold that meaning

[21] I am in agreement here with Kauppinen who claims that subjective meaning "isn't fundamentally a matter of judgment." Kauppinen, "Meaning and Happiness," 165.

[22] For subjectivist accounts, see Taylor, *Good and Evil*; Calhoun, *Doing Valuable Time*, ch. 2; and Parmer, "Meaning in Life and Becoming More Fulfilled."

can be fully objective.[23] In any case, I aim only to characterize the subjective component; I can remain neutral on whether the subjective component is the whole of meaning or whether it is a component of a hybrid good. It is even consistent with my view to think that people can live objectively meaningful lives despite lacking all attraction if subjective and objective meaning are logically unrelated.

I can even remain neutral on whether living a subjectively meaningful life is always or necessarily good. Someone who finds herself attracted to an end that is in fact worthless (or even positively bad) may experience attraction-motivated activity and so find her life meaningful. Is it positively good for her that she finds it so? Perhaps; just as welcome pleasure is always good for the subject even if it is pleasure in the bad, attraction-motivated activity may always provide a sense of meaningfulness that contributes to the subject's welfare. But I need not insist on that. Even if we say that subjective meaning promotes welfare only when the object is worthy of attraction, it remains the case that attraction is an essential ingredient in a component of well-being.

Finally, my view suggests that attraction-driven activity is not just a component of well-being but a central component. Anyone who thinks that subjective meaning is part of the good life presumably thinks it is a critical component. Attractions play a role in well-being comparable to the role played by pleasure according to hedonistic theories and to the role played by objective goods like knowledge and friendship according to objective list theories. A complete absence of attractions and the opportunity to pursue them can leave us in that state of deep, existential boredom referenced earlier. We may feel as if we are suffocating—that our will cannot get a grip on anything in the right way. No amount of idle amusement or getting wrapped up in satisfying aversions can cure this malaise.

A picture of human psychology that speaks only of desire or preference may be adequate for the empirical explanation of action, but it is not adequate for the philosophy of well-being. Because of the importance of attraction to the philosophy of well-being, contemporary theories of desire should attend to the distinction. We therefore seek an account of attraction and aversion. I will argue—mostly via a process of elimination—that attraction and aversion are distinguished by their normative content.

[23] For accounts that are at least partly objective, see Evers and van Smeden, "Meaning in Life"; Kauppinen, "Meaningfulness and Time"; Kekes, "The Meaning of Life"; Levy, "Downshifting and Meaning in Life"; Metz, *Meaning in Life*, ch. 12; Smuts, "The Good Cause Account of the Meaning of Life"; Wielenberg, *Value and Virtue in a Godless Universe*; and Wolf, *Meaning in Life and Why It Matters.*

3. Toward a Theory of Attraction and Aversion

It is worth dismissing straightaway a tempting but ultimately untenable proposal for distinguishing attraction and aversion. One might think that an attraction is simply a desire for a positive state of affairs—that *p* be the case—whereas an aversion is a desire for a negative state of affairs: that *q* not be the case. Indeed, when we talk of aversions we often talk of wanting things not to happen: not to be rejected, not to be covered in spiders, not to be poor, etc. And when we speak of attractions we speak of wanting things to happen: to be rich, happy, and famous. The problem with this simple proposal is that every desire for *p* to be the case is equivalent to a desire for not-*p* not to be the case. As Schroeder has observed, someone who desires pie can just as well be described as desiring that it not be the case that she lacks pie.[24] Someone who wants to be rich is someone who wants to not *not* be rich.

In other words, attraction and aversion cannot be distinguished by the fact that one has a "positive," the other a "negative," content; for any content *p*, positive or negative, there is a difference between being attracted to *p* and averse to not-*p*. Indeed, to say that attraction and aversion are distinct attitudes is just to say that for any content *p*, positive or negative, there is a difference between being attracted to *p* and averse to not-*p*, even if the underlying states have exactly the same conditions of satisfaction. The same problem befalls dispositional accounts of attraction and aversion: to be *disposed to bring about* that one has pie is no different than being *disposed to avoid* a lack of pie.

Another account might point to the distinctive emotional syndromes with which attraction and aversion are associated, as Sinhababu has.[25] Attractions give us a delighted happy feeling when we find that we can satisfy them and an unpleasant feeling of disappointment when we discover that we cannot. Aversions give us the pleasure of relief when we find that we can satisfy them and an unpleasant feeling of anxiety or dread when we discover that we cannot. Sinhababu's proposal echoes Descartes's observation that the desire someone has when he tends towards some good "is accompanied with love and afterwards with hope and joy," whereas "the same desire, when he tends to the avoiding an evil contrary to this good, is attended with hatred, fear, and sorrow."[26] (Descartes noted that it may nevertheless be "but one passion" that underlies these different syndromes.)

[24] Schroeder, *Three Faces of Desire*, 26.

[25] Sinhababu, "The Humean Theory of Motivation Reformulated and Defended," 490.

[26] Descartes, *Passions of the Soul*, PA a.87.

A useful heuristic for discerning whether a desire is an attraction or an aversion is to consider whether its satisfaction would bring delight or relief. But while it is true that attraction and aversion have distinct emotional profiles, it is implausible that these downstream consequences are the essential difference between attraction and aversion. Something about the nature of attraction as such should explain *why* we feel a delighted happy feeling when the attraction is satisfied. Something about the nature of aversion should explain why we are relieved (but not delighted) when the aversion is satisfied.

One traditional view, going back at least to Hume, holds that attraction and aversion are distinguished by their connections to pleasure and pain: "When we anticipate pain or pleasure from some source, we feel aversion or propensity to that object."[27] But while it is true that attractions often involve the anticipation of pleasure, this is not always the case. One can be attracted to the prospect of tasting a durian fruit or submerging oneself in an ice bath, despite expecting these things to be *wholly* unpleasant.[28] And one can be attracted to a state of affairs that has nothing to do with one's own pleasure or pain at all—e.g., the preservation of an endangered species in a remote time or place.

There is a principled reason to think that attraction does not necessarily involve the anticipation of pleasure: we can become attracted to anything we view as alluring or appealing, yet there is no essential connection between something being alluring or appealing and its being pleasant. There are many ways something can be alluring that are unconnected to its hedonic value—it may be daring, charming, sublime, virtuous, courageous, interesting, transgressive, and so on. The ice bath, for instance, may be attractive because it is purifying, not because it is pleasant. When explaining why we are attracted to something, we often refer to these qualities rather than its pleasantness.

What is more, it is not clear that pleasure can be characterized without invoking attraction. According to a leading view of the nature of pleasure, the attitudinal view, a sensation qualifies as a sensation of pleasure just in case its subject is attracted to feeling it as she is having it. But we cannot analyze attraction in terms of pleasure if pleasure is analyzed in terms of attraction.

Still, attraction may be essentially connected to pleasure in a different way than Hume specified. Pallies has not developed an account of attraction and aversion in detail but suggests that "attraction involves a certain sort of directed

[27] Hume, *Treatise of Human Nature*, 2.3.3.3.

[28] The durian fruit example is drawn from Shaw, "Do Affective Desires Provide Reasons for Action?" 3.

anticipatory pleasure; aversion involves a certain sort of directed anticipatory displeasure."[29]

The problem with Pallies's view is that it is questionable that only attraction is associated with anticipatory pleasure. As Sinhababu has noted, all desires, including aversions, have a hedonic aspect.[30] Aversion satisfaction is associated with the pleasure of relief. And the anticipation of aversion satisfaction is also associated with pleasure: a kind of anticipatory relief. Expecting that I will make my flight, I experience a wave of pleasurable relief, directed at the prospect of being on time.

Of course, attractions may be associated with a distinctive *kind* of anticipatory pleasure: positive delight, in contrast to relief. But this is not a difference in pleasure, only in emotional character. And how do we analyze these two emotional characters? It is natural to think that delight-pleasure is the pleasure that comes with the satisfaction of an attraction, whereas relief-pleasure is the pleasure that comes with the satisfaction of an aversion, in which case attraction and aversion are more fundamental than the two kinds of pleasure. The pleasure-pain analysis, therefore, looks to be unpromising.

What is left? If the difference between being "pulled toward" *p* and "pushed away" from not-*p* cannot be characterized by appeal to dispositions, emotional syndromes, or pleasure and pain, where might it lie?[31]

Let us return to Descartes's assumption that attraction is a passion that tends to the seeking after good, and aversion is a passion that tends to the avoiding of evil. As noted, one way to spell out this idea is to say that attraction and aversion have positive and negative normative content, respectively. When we are attracted to something we see it as good, and when we are averse to something we see it as bad. According to Descartes (following Augustine), this

[29] Pallies, "Attraction, Aversion, and Asymmetrical Desires," 618.

[30] Sinhababu, "The Humean Theory of Motivation Reformulated and Defended," 490.

[31] One theory I have not considered is Schroeder's neuropsychological theory of attraction and aversion. According to Schroeder, attraction and aversion are rooted in the reward system and the punishment system, respectively. If someone desires *p*, she "constitute(s) *P* as a reward" (*Three Faces of Desire*, 131). In contrast, aversions involve constituting not-*p* as punishing. Schroeder's theory is unusual because it *analyzes* attraction and aversion in terms of their neural bases; to be attracted to *p* just is to constitute *p* as rewarding. According to Schroeder, reward signal is a plausible candidate for what folk psychology calls "desire" in the same way the element with atomic number 79 is what folk chemistry calls "gold." Schroeder's theory deserves more discussion than I have the space for here, but I will say a few words. I am happy to grant that Schroeder is correct about the neural bases of attraction and aversion in creatures like us. However, this entails that creatures incapable of learning cannot have attractions or aversions since reward and punishment signal, on Schroeder's view, is tied to learning, but this is too difficult to accept.

is a distinction without a difference, since the privation of goodness is itself an evil; hence anyone who sees *p* as good sees (or should see) not-*p* as bad. But perhaps Descartes's argument is too quick.

Intuitively, there is a difference between the privation of goodness and the presence of evil. It would be positively good for me if I won a million dollars tomorrow; but my failing to win a million dollars is not positively bad—it is just neutral. Similarly, it would be positively bad if I had a toothache right now; but the fact that I do not have a toothache is not positively good—it is just neutral. Folk axiology finds this self-evident, and a theory of goodness can be provided that accommodates these distinctions. Such a theory sees good and bad as existing on a spectrum containing a zero point, rather like the spectrum that runs from intense pleasure to intense pain (of a given sort), which has a natural zero in states that are neither pleasurable nor painful. The neutral state of affairs would be one that is neither good nor bad—e.g., the state of affairs in which absolutely nothing exists or will exist.

So, the proposal is that when we are attracted to something we see it as *noncomparatively good*—i.e., as better than neutral. And when we are averse to something we see it as *noncomparatively bad*—i.e., as worse than neutral. The theory thus distinguishes an attraction to *p* from an aversion to not-*p* by appealing to an axiological distinction between the property of being good and the property of not being bad (or being better than the alternatives). When one is averse to not-*p*, one sees not-*p* as bad, and hence sees *p* as not bad (or better than not-*p*); but one does not see *p* as positively good (unless one is also, independently, attracted to *p*).

This is a qualified version of the "guise of the good" view of desire, though it differs in an important respect from the standard formulation.[32] The guise of the good theory says that all that is desired is seen by the subject as good in some respect or another, and intentional action, or acting for a reason, is action that is seen as good by the agent. But unlike most proponents of this view, I do not say that *desire* as such represents its object as good. My theory says that an attraction to *p* represents *p* as good. An aversion to not-*p*, in contrast, represents not-*p* as bad without representing *p* as good (leaving it open that some desires do neither, as with Radio Man's blank urge to turn on radios).[33] Thus, not all that is desired is seen as good, and not all intentional action is action

[32] Defenders of the guise of the good include Anscombe, *Intention*; Davidson, "How Is Weakness of the Will Possible?"; Quinn, "Putting Rationality in Its Place"; Stampe, "The Authority of Desire"; Scanlon, *What We Owe to Each Other*; Oddie, *Value, Reality, and Desire*; Tenenbaum, *Appearances of the Good*; and Gregory, "Why Do Desires Rationalize Actions?"

[33] Quinn, "Putting Rationality in Its Place."

for the sake of something seen as good. Intentional action can be the product of aversion, in which case one sees the alternatives as bad without seeing the intended action as good.

Of course, one might deny that there is in fact an axiological zero, holding that the only axiological structure is a betterness ordering and that the appearance of a good/bad distinction can be explained away. On this view, the spectrum of value is akin to the spectrum that runs from high notes to low notes, extending infinitely in each direction, which has no zero.[34]

But in the absence of a strong argument to the contrary, it is more natural to think that the spectrum of value admits a zero point. As just noted, some things seem to contain no positive or negative value at all, such as pockets of empty space. We can formulate a further argument in support of this view. Consider that more determinate forms of goodness, like beauty, clearly exist on such a spectrum. There is a spectrum of aesthetic value consisting of the beautiful things toward one end and the ugly things toward the other, with things that are neither beautiful nor ugly in the middle. Plausibly, whenever something is generically good, its goodness is grounded in its being good in some determinate way(s)—such as its being beautiful, just, charming, or pleasant. But if these determinate forms of value exist on a spectrum with a natural zero, then generic value should too: the zero point is given by the zero point of the more determinate value(s).

And finally, the role of attraction in felt meaning supports the view that attraction and aversion have positive and negative normative content. Attraction-motivated activity is central to the good life. But as we saw, this is not because it is pleasant. The better theory is that attraction-motivated activity is valuable because attraction involves an appearance of the good, and such appearances (and activity motivated by them) are irreducibly valuable in virtue of being meaningful. Meaning cannot be attained by pursuing the lesser evil or by pursuing what is favored over the alternatives; it is realized only by being drawn forward by the (noncomparative) good. Thus, to explain the role of attraction in well-being, we must endorse a representational theory of attraction and aversion. And while this may not entail that such properties of absolute goodness and badness exist, it gives us a humanistic reason to take an interest in them.

I have argued that the best account of attraction and aversion differentiates the attitudes by their normative content. To be "pulled towards *p*" is to represent *p* as noncomparatively good in a motivationally efficacious way. It is not just to believe or even to know that *p* is good but to see *p* as good; its positive

[34] For a defense of this view, see Broome, "Goodness Is Reducible to Betterness."

goodness is presented to us in experience in a quasi-perceptual way that cannot be reduced to our believing or even knowing that it is good.[35] The argument for this view is that alternative theories, such as the pleasure-pain theory, are untenable and that, in addition, only the representational view can explain the distinctive nonhedonic contribution of attraction to well-being.

4. Conclusion

I have argued that the role of desire in the good life cannot be fully appreciated without distinguishing between attraction and aversion. The experience of meaning in life requires the pursuit of ends to which we are attracted and cannot be attained simply by pursuing the lesser evil. I have further argued that when we are attracted to something, we see it as (noncomparatively) good, and subjective meaning consists in the experience of being drawn forward by the (perceived) good. Since meaning is central to the good life, and meaning requires attraction, philosophers should take the distinction between attraction and aversion seriously.[36]

University of British Columbia
alisabeth.ayars@ubc.ca

References

Anscombe, G. E. M. *Intention*. Cambridge, MA: Harvard University Press, 2000.

Broome, John. "Goodness Is Reducible to Betterness: The Evil of Death Is the Value of Life." In *The Good and the Economical: Ethical Choices in Economics*

[35] See Stampe, "The Authority of Desire"; and Johnston, "The Authority of Affect." I take the perceptual view to be consistent with an affect-based account of attraction, which claims that attraction essentially involves a disposition to experience positive affect, such as that defended by Smithies and Weiss in "Affective Experience, Desire, and Reasons for Action." The claim would then be that the relevant positive affect consists in a kind of quasi-perceptual experience with normative content.

[36] I am most indebted to Gideon Rosen for extensive comments on this paper and discussion of these ideas. I also thank Daniel Pallies, Sumeet Patwardhan, and two anonymous reviewers for extremely thorough comments, which shaped the paper in significant ways. I thank Scott Sturgeon, Matthew Bedke, and Jonathan Ichikawa for their insights and encouragement. I am also grateful to the attendees of the UBC work-in-progress group where this paper was discussed; the participants at the 2022 Vancouver Summer Philosophy Conference, where this paper was presented; and the attendees of the talk I gave for the University of Birmingham on this paper.

and Management, edited by P. Koslowski and Y. Shionoya, 70–86. Berlin: Springer, 1993.

Calhoun, Cheshire. *Doing Valuable Time: The Present, the Future, and Meaningful Living*. Oxford: Oxford University Press, 2018.

Chang, Ruth. "Can Desires Provide Reasons for Action." In *Reason and Value: Themes from the Moral Philosophy of Joseph Raz,* edited by R. Jay Wallace, Philip Pettit, Samuel Scheffler, and Michael Smith, 56–90. Oxford: Oxford University Press, 2004.

Davidson, Donald. "How Is Weakness of the Will Possible?" In *Moral Concepts,* edited by Joel Feinberg. Oxford: Oxford University Press, 1969.

Descartes, Rene. *The Passions of the Soul*. Translated by Stephen Voss. Indianapolis: Hackett, 1989.

Evers, Daan, and Gerlinde Emma van Smeden. "Meaning in Life: In Defense of the Hybrid View." *Southern Journal of Philosophy* 54, no. 3 (September 2016): 355–71.

Gregory, Alexander. "Why Do Desires Rationalize Actions?" *Ergo* 5, no. 40 (January 2019): 1061–81.

Heathwood, Chris and Philosophy Documentation Center. "Ill-Being for Desire Satisfactionists." *Midwest Studies in Philosophy* 46 (2022): 33–54.

Hume, David. *Treatise of Human Nature*. Edited by L. A. Selby-Bigge. Oxford: Oxford University Press, 1978.

Johnston, Mark. "The Authority of Affect." *Philosophy and Phenomenological Research* 63, no. 1 (July 2001): 181–214.

Kagan, Shelly. "An Introduction to Ill-Being." In *Oxford Studies in Normative Ethics,* vol. 4, edited by Mark Timmons, 261–88. Oxford: Oxford University Press, 2014.

Kauppinen, Antti. "Meaning and Happiness." *Philosophical Topics* 41, no. 1 (Spring 2013): 161–85.

———. "Meaningfulness and Time." *Philosophy and Phenomenological Research* 84, no. 2 (March 2012): 345–77.

Kekes, John. "The Meaning of Life." *Midwest Studies in Philosophy* 24 (2000): 17–34.

Kelley, Anthony Bernard. "Well-Being and Alienation." PhD diss., University of Colorado at Boulder, 2020. https://www.proquest.com/docview/2476324285/abstract/FF2FEC4BCFF04BF4PQ/1.

Langton, Rae. "Maria von Herbert's Challenge to Kant." In *Ethics: Essential Readings in Moral Theory,* edited by George Sher, 377–86. New York: Routledge, 2012.

Levy, Neil. "Downshifting and Meaning in Life." *Ratio* 18, no. 2 (June 2005): 176–89. https://doi.org/10.1111/j.1467-9329.2005.00282.x.

Mathison, Eric. "Asymmetries and Ill-Being." PhD diss., University of Toronto, 2018.

Metz, Thaddeus. *Meaning in Life*. Oxford: Oxford University Press, 2013.

Oddie, Graham. *Value, Reality, and Desire*. Oxford: Clarendon Press, 2005.

Pallies, Daniel. "Attraction, Aversion, and Asymmetrical Desires." *Ethics* 132, no. 3 (April 2022): 598–620.

Parmer, W. Jared. "Meaning in Life and Becoming More Fulfilled." *Journal of Ethics and Social Philosophy* 20, no. 1 (July 19, 2021): 1–29.

Quinn, Warren, ed. "Putting Rationality in Its Place." In *Morality and Action*, 228–55. Cambridge: Cambridge University Press, 1994.

Scanlon, T. M. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press, 2000.

Schroeder, Timothy. *Three Faces of Desire*. Oxford: Oxford University Press, 2004.

Shaw, Ashley. "Do Affective Desires Provide Reasons for Action?" *Ratio* 34, no. 2 (June 2021): 147–57.

Sinhababu, Neil. "The Humean Theory of Motivation Reformulated and Defended." *Philosophical Review* 118, no. 4 (October 2009): 465–500.

Smithies, Declan, and Jeremy Weiss. "Affective Experience, Desire, and Reasons for Action." *Analytic Philosophy* 60, no. 1 (March 2019): 27–54.

Smuts, Aaron. "The Good Cause Account of the Meaning of Life." *Southern Journal of Philosophy* 51, no. 4 (December 2013): 536–62.

Stampe, Dennis W. "The Authority of Desire." *Philosophical Review* 96, no. 3 (July 1987): 335–81.

Strawson, Galen. *Mental Reality*. 2nd ed. Cambridge, MA: MIT Press, 2009.

Sumner, Wayne. "The Worst Things in Life." *Grazer Philosophische Studien* 97, no. 3 (2020): 419–32.

Taylor, Richard. *Good and Evil: A New Direction*. New York: Macmillan, 1970.

Tenenbaum, Sergio. *Appearances of the Good: An Essay on the Nature of Practical Reason*. Cambridge: Cambridge University Press, 2007.

Wielenberg, Erik J. *Value and Virtue in a Godless Universe*. Cambridge: Cambridge University Press, 2005.

Wolf, Susan. "Happiness and Meaning: Two Aspects of the Good Life." *Social Philosophy and Policy* 14, no. 1 (January 1997): 207–25.

———. *Meaning in Life and Why It Matters*. Princeton: Princeton University Press, 2012.

Journal of Ethics and Social Philosophy
VOL. 28, NO. 3 · OCTOBER 2024

https://doi.org/10.26556/jesp.v28i3.3270

ON THE METAPHYSICS OF RELATION-RESPONSE PROPERTIES; OR, WHY YOU SHOULDN'T COLLAPSE RESPONSE-DEPENDENT PROPERTIES INTO THEIR GROUNDS

Spencer M. Smith

> Words are our tools, and, as a minimum, we should use clean tools: we should know what we mean and what we do not, and we must forearm ourselves against the traps that language sets us. . . . And more hopefully, our common stock of words embodies all the distinctions men have found worth drawing, and the connexions they have found worth marking, in the lifetimes of many generations: these surely are likely to be more numerous, more sound, since they have stood up to the long test of the survival of the fittest, and more subtle, at least in all ordinary and reasonably practical matters than any that you or I are likely to think up in our armchairs of an afternoon—the most favoured alternative method.
>
> —J. L. Austin, "A Plea for Excuses"

Austin's cautionary remarks are well taken: words *are* our tools, and we ought indeed to "use clean tools"—particularly when doing philosophy. And while we may reasonably question their details or the extent to which they point toward a viable research program for philosophy, Austin's more hopeful observations about there being important distinctions and connections enshrined in natural language are surely onto something as well. For a family of what I take to be particularly clear confirming instances of the latter observation, consider the following series of predicates:[1]

> "blameworthy," "praiseworthy," "trustworthy," "noteworthy," "buzzworthy," "bingeworthy" . . .

[1] For the sake of readability, I will often proceed as though standalone adjectives such as those listed count as predicates, rather than always including a verb.

> "desirable," "believable," "admirable," "laughable," "memorable," "lovable," "punchable" …
>
> "awe-inspiring," "hope-inspiring," "anxiety-inducing," "fear-inducing," "tear-jerking" …

Each predicate in each of these series appears to denote a property with *relation-response structure*.[2] That is, each predicate appears to denote a particular relational property—namely, the property of standing in a given relation to a given type of response, whether that response be emotional, attitudinal, or behavioral. Each of these lists, of course, goes on.[3]

Moreover, each of the foregoing predicates appears to implicate a particular relation and a particular type of response as figuring in the structure of the property it denotes.[4] There is room to haggle over precisely how to analyze the relation of worthiness or the response type of blame, for instance; but it nevertheless seems clear from the meaning of the English word "blameworthy"

[2] In this paper I assume that, as a general matter, meaningful predicates denote properties, save for troublesome predicates like, e.g., "does not self-instantiate."

[3] It is important to be clear that a predicate's merely having one of these lists' distinctive suffixes—e.g., "-worthy" or "-able"—is not sufficient for it to be a member of the corresponding list. For a thing to be seaworthy, for instance, is presumably not for that thing to be worthy of a certain sort of response picked out (strangely) by "sea." Perhaps certain uses of "seaworthy" *imply*, in corresponding conversational contexts, that the seaworthy item is indeed worthy of a certain type of response, e.g., sailing, floating, etc. But to say that some object is seaworthy is not *in itself* to say, for some response *R*, that that object is worthy of *R*.

[4] Other series of predicates are close kin to the ones I will be focusing on, including:

> "awesome," "fearsome," "loathsome," "irksome," "tiresome," "worrisome" …
>
> "interesting," "irritating," "annoying," "disturbing," "inspiring," "tiring" …

Each of these predicates appears to denote a property with relation-response structure. What distinguishes them from the predicates I will be focusing on is that to the extent that these latter expressions indicate which relation-constituents figure in the relation-response structures of the properties they denote, they appear to do so only with what Quine might have called "studied ambiguity" ("On What There Is," 26). Thus, it is not quite true that the "-some" and "-ing" suffixes, as they appear in the members of our additional series, implicate *particular* relations. It seems better to say that these suffixes serve a generalizing function—namely, the function of allowing a user of the word to implicate the presence of one or another out of a range of possible particular relations without having to specify which. Thus, in saying that a thing is awesome, competent English users have a decent sense of the range of possible particular relations they are implying this thing might bear to the response type of awe: perhaps it is a relation of *engendering* or of *meriting*. (Context, I suppose, can help to narrow this down.)

Everything of importance that I have to say in this paper about relation-response expressions and the properties they denote applies just as well to the members of these additional series. I neglect them only because their generality makes discussion of them messier.

that it is indeed this relation—namely, worthiness—and indeed that type of response—namely, blame—that one must understand if one is to understand the property that "blameworthy" denotes. In light of this, let a *faithful reading* of "blameworthy"—or of the corresponding property name "blameworthiness"—be a reading that has it denote a property with *genuine relation-response structure*—i.e., relation-response structure that is fundamental, or that cannot be "analyzed out"—whose fundamental relation-constituent stands a good chance of being what we standardly mean by "worthy" (in the relevant contexts), and whose fundamental response constituent stands a good chance of being what we standardly mean by "blame" (in the relevant contexts). The notion of a faithful reading generalizes to other predicates of the relevant sort. Moreover, we can talk of relation-response structures themselves or the properties that have them as being faithful to a given predicate or property name.

If you are like me, you may think it a straightforward deliverance of English that we ought to read and use the aforementioned predicates and their corresponding property names faithfully; as Gideon Rosen puts it, our accounts of blameworthiness, trustworthiness, etc. "should respect word structure."[5] But surprisingly, many philosophers appear to use certain such expressions—e.g., "blameworthy"—to denote properties that lack faithful structure. Such philosophers appear instead to use "blameworthy" as a predicate for properties like, e.g., having acted wrongly from ill will—properties that to my mind seem far better fit to serve as *conditions* or *grounds* of blameworthiness rather than as blameworthiness itself.

Upon hearing of such news, you may be disposed to think this a case of mere verbal slippage, that these philosophers were just speaking loosely or carelessly. But if that is the story you wish to run with, it is difficult to know what to think in response, e.g., to Jules Coleman and Alexander Sarch's report that behaving this way with regard to blameworthiness is "standard," or to David Shoemaker's report that theories which strip blameworthiness of faithful relation-response structure in this way are "much more popular" than theories that do not.[6] If these reports are right, respect for word structure seems to be in surprisingly short supply, at least in one major philosophical subliterature.

This paper is, among other things, a plea for respecting word structure when it comes to theorizing putative relation-response properties generally. To some extent this will be an Austinian exercise in terminological hygiene: relation-response expressions figure centrally in a significant number of philosophical

[5] Rosen, "The Alethic Conception of Moral Responsibility," 66.

[6] Coleman and Sarch, "Blameworthiness and Time," 101; and Shoemaker, "Response-Dependent Responsibility," 483.

discussions, so it is all for the best to keep them in good working order. But the project is not merely prophylactic, for I will also spend some time arguing that respect for word structure here can help us to see more clearly what is truly at stake in recent debates concerning the natures of certain value properties.

The paper proceeds as follows: In section 1, I introduce Gideon Rosen's ground-theoretic framework for theorizing blameworthiness, and I offer a generalization of that framework for theorizing putative relation-response properties across the board. This framework will prove useful in the work to come. In section 2, I unpack my contention that many philosophers appear to neglect word structure when analyzing putative relation-response properties, focusing on blameworthiness as my case study. In section 3, I consider two arguments—one recently articulated by Justin D'Arms and Daniel Jacobson, the other by Shoemaker—for the claim that a popular approach to theorizing certain putative relation-response properties requires those who adopt it to deny that such properties have genuine relation-response structure.[7] I show that D'Arms and Jacobson's argument is invalid as it stands, and I argue that at least one natural way of rendering it valid relies upon an account of property individuation that those to whom the argument is directed have good reason to reject. I then show that Shoemaker's argument relies crucially upon an assumption that its targets need not, should not, and do not in all cases accept. Finally, in sections 4 and 5, I argue that whereas recently propounded classification schemes say otherwise, a great deal of the debate between so-called Response-Independence and Response-Dependence theories of certain value properties—properties like, e.g., blameworthiness, trustworthiness, etc.—ought not to be framed as hinging on whether the relation-response structure of such a property is affirmed as genuine. In fact, merely to affirm as much leaves nearly everything of importance in that debate yet to be settled.

1. A Framework for Theorizing Relation-Response Properties

In this section, I draw upon the work of Gideon Rosen to establish a framework for theorizing putative relation-response properties, and I use that framework to distinguish different approaches that one might take to such theorizing.

1.1. Three Question-Schemas

Rosen poses three questions that any comprehensive theory of blameworthiness ought to address:

[7] D'Arms and Jacobson, "The Motivational Theory of Guilt (and Its Implications for Responsibility)"; and Shoemaker, "Response-Dependent Theories of Responsibility."

1. *The Analytic Question*: What is it for something to be blameworthy?
2. *The Grounding Question*: What are the conditions under which something is blameworthy?
3. *The Explanatory Question*: Why are the conditions of being blameworthy as they are?[8]

Rosen's questions get to the heart of the matter and can be adapted for the purposes of theorizing other putative relation-response properties. Here, then, are three question-schemas whose instances any comprehensive theory of a relation-response property, *F*-ness, ought to address:

I. *The Analytic Question-Schema*: What is it for something to be *F*?
II. *The Grounding Question-Schema*: What are the conditions under which something is *F*?
III. *The Explanatory Question-Schema*: Why are the conditions of being *F* as they are?

The Analytic Question-Schema (henceforth "QS-I") asks what it is for something to be *F*, where—as I shall later explain—a true answer constitutes a *metaphysical analysis* or *real definition* of being *F*. The Grounding Question-Schema (henceforth "QS-II") asks not what it is to be *F* but rather what it is in virtue of which *F*-things are *F*. In other words, it asks for an account of the *explanatory ground* or *explanatory grounds* of *F*-ness instantiations.[9] In still other words, QS-II asks for a list of the *F-making* properties there are—i.e., those properties the having of which confers (a degree of) *F*-ness upon their bearers. The Explanatory Question-Schema (henceforth "QS-III") goes a step further. It asks what it is about the *F*-making properties in virtue of which they are *F*-making.

QS-II and QS-III each have to do with a form of noncausal metaphysical determination currently being investigated by philosophers under the name "grounding." QS-I may also have to do with grounding if, following philosophers like Rosen or Fabrice Correia, we construe analysis or real definition ground

[8] See Rosen, "The Alethic Conception of Moral Responsibility," 65–68. I have not reproduced Rosen's questions verbatim, since the questions he considers explicitly concern responsibility rather than blameworthiness. But Rosen proceeds via a series of terminological stipulations to hone in on the topic of blameworthiness, and tasks himself with providing an account of blameworthiness that addresses each of the three questions I have presented. Thus the interpolation.

[9] Because I take it to be relatively unimportant in the context of the present paper, I will for the most part blur the distinction between a thing's being *F* and that thing's having the property *being F*, or *F-ness*.

theoretically.[10] In light of these connections, it behooves us briefly to familiarize ourselves with some basic tools for thinking about grounding bequeathed to us by the literature on it. They will prove useful in drawing out some further features of QS-I–III and in our investigations to come.

1.2. Grounding: Some Basics

Grounding, as I will be thinking of it, is an irreflexive, antisymmetric, transitive, noncausal determination relation between facts. To say that grounding is a noncausal determination relation between facts is to say that when one fact, *A*, grounds another fact, *B*, *A* in some sense *makes B* obtain, but not by way of causing *B* to obtain. In the typical case of grounding thus conceived, a single fact, *A*, is grounded in a plurality of facts, *Γ*, numbering anywhere from one to infinitely many.

Facts in this context are themselves typically conceived as worldly items, in particular, as either so-called true Russellian propositions or Armstrongian states of affairs: the discrete, worldly counterparts to declarative sentences that, in the least controversial instance, consist in certain arrangements of objects and their properties or relations.[11] In what follows, I adopt the standard convention of adjoining brackets to declarative sentences in order to form the names of the facts that correspond to those sentences when true. For example, take the sentence "Blue is a dog." This sentence, when true, corresponds to a fact, namely, [Blue is a dog].

Grounding is also thought to be the relation of noncausal metaphysical explanation, or else the relation that backs such explanation. Thus, when a fact, *A*, is *wholly grounded* in a plurality of facts, *Γ*, *A* is said to obtain *because of* or *in virtue of* the obtaining of the facts comprising *Γ*. In turn, whenever *A* is wholly grounded in *Γ*, *A* is *partly grounded* in each subplurality of facts comprising *Γ* and is thus said to obtain *partly in virtue* of each such subplurality.

1.3. Understanding Question-Schemas I, II, and III

With the foregoing bit of grounding ideology in hand, let us turn to consider more deeply what QS-I–III are asking.

There are different things we might be asking when we ask *what it is* for something to be *F*, for any given predicate we might substitute for "*F*." Following Rosen, I stipulate that QS-I asks for a *metaphysical analysis* or *real definition* of being *F*. (Henceforth, I simplify discussion by supposing that metaphysical

[10] Rosen, "Metaphysical Dependence," 122–26, and "Real Definition," 197–200; Correia, "Real Definitions," 57–59.

[11] See, e.g., Rosen, "Metaphysical Dependence," 114–15; and Audi, "Grounding," 686.

analysis and real definition are the same thing, and I use "analysis" as my term of choice.) To be sure, there are debates to be had about analysis. For instance, Rosen holds that analysandum facts are always wholly grounded in their corresponding analysans facts, whereas Paul Audi—toward whose position I myself am presently more inclined—takes analysandum facts to be identical to their corresponding analysans facts and thus, given the irreflexivity of grounding, not at all grounded in those facts.[12] I do not wish to enter into this debate here. Rather, I mention the disagreement for the purposes of clarifying my understanding of QS-II, toward which I now turn.

QS-II, as I have it, asks for the conditions of being *F*, or the *F*-making properties. We have gone further and explained that QS-II asks for the *grounds* of *F*-ness instantiations (or "*F*-facts," for short). But this can now be seen to be ambiguous: if we suppose with Rosen that the ground of a fact can be that fact's analysans, then some answers to QS-I may double as answers to QS-II. I do not know whether Rosen wants this, but—more importantly for our purposes—*I* do not want this. So I stipulate that QS-II asks after the grounds of *F*-facts where the grounds in question do not stand as analysans to their corresponding *F*-facts.

QS-III, finally, asks why the conditions of being *F* are as they are. In other words, what makes the conditions of *F*-ness be conditions of *F*-ness? When the conditions of *F*-ness are themselves property instantiations, an equivalent question would be: in virtue of what are the *F*-making properties *F*-making? Why are *these* properties—the properties cited in response to QS-II—the *F*-makers? Alternatively and somewhat torturously, we might frame the question in terms of fact forms and ask: When some facts of such-and-such forms get together to ground a fact of some other form, what are the forms of the facts which ground the fact that the former facts ground the latter?[13] Where it is easier to do so, I endeavor to speak in terms of properties rather than of fact forms.

Ultimately, we are left with a grounding structure that can be represented graphically as in figure 1. The arrows represent what may be either whole or partial grounding relations, as the case may be. The three boxes represent facts that correspond to possible answers to QS-I–III, respectively.[14] The bracketed

[12] Rosen, "Metaphysical Dependence," 122–26, and "Real Definition," 197–200; Audi, "Grounding," 686. See also Dorr, "To be *F* is to be *G*," 43, 54, for what is effectively a conditionalized defense of Audi's stance on the point.

[13] By a "fact form," I mean a form that distinct particular facts may share. For instance, [Blue is a dog] and [Thea is a dog] each share the fact form [*x* is a dog].

[14] As I say above, I am inclined to regard analyzable facts as identical to the facts that analyze them. For instance, if we say that to be a bachelor is to be an unmarried eligible male, then I am inclined to say that for all *x*, if *x* is a bachelor or an unmarried eligible male, then [*x* is

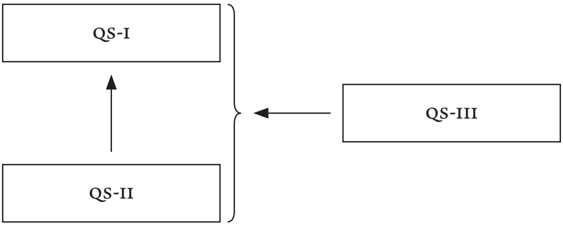


FIGURE 1

contents together represent the higher-order fact that the QS-II fact(s) ground the QS-I fact. Written out, our structure says that (1) the QS-II fact or facts ground the QS-I fact, and (2) the QS-III fact or facts ground the fact [The QS-II fact or facts ground(s) the QS-I fact].[15]

To see how the structure might look when filled, consider a nonnaturalistic version of consequentialism, namely, one that accepts a necessitated version of the standard equivalence—necessarily, an act is right if and only if it maximizes goodness—but denies that to act rightly *just is* to maximize goodness. Nevertheless, the view says that whenever an act is right or is goodness-maximizing, it is right *directly in virtue of* being goodness-maximizing (or "optimal," for short). In other words, facts of the form [*x* acts optimally] are immediate

a bachelor] = [*x* is an unmarried eligible male]—in effect, a single fact has two linguistic or representational garbs, one of which is more perspicuous as to the structure of that fact than the other. Still, for reasons of neatness, I shall often plug in the less perspicuous presentation of an analyzable fact into QS-I boxes. That is, I shall put in an open sentence like "*x* is a bachelor" rather than "*x* is an unmarried eligible male," even though the latter embeds a more proper answer to the "What is it to be a bachelor?" instance of QS-I. On my preferred view of analysis, this is but a minor presentational infelicity, since on that view "*x* is a bachelor" and "*x* is an unmarried eligible male," for a given *x*, designate the same fact. On Rosen's view of analysis, however, it is inaccurate to use "*x* is a bachelor" rather than "*x* is an unmarried eligible male" to designate the fact corresponding to the "What is it to be a bachelor?" instance of QS-I. There is thus a tension between how I shall be portraying grounding structures in this paper and how someone with Rosen's view of analysis would portray such structures. This is unfortunate, but not greatly so: whether we think of QS-I facts in my preferred way or in Rosen's way, we will agree that such facts are to "go above" QS-II facts in the grounding structures we will be looking at; and agreeing about these sorts of structural relations between the facts we shall be considering will generally suffice to ensure that we are on the same page about the relevant claims I shall be making.

[15] One potentially misleading feature of this way of depicting things is that it may be taken as implying that there is always exactly one fact corresponding to each node in the explanatory structure. Such an implication would be false, most clearly in the cases of the QS-II and QS-III nodes: there can be multiple grounds of a given QS-I fact, and there can be multiple grounds of the fact that a given QS-II fact grounds a given QS-I fact. For such cases, we would need many more boxes than just three. But the basic structure we have represented would be preserved, and that is the main thing I want these graphics to assist us in tracking.

whole or partial grounds of corresponding facts of the form [*x* acts rightly]. Thus, we have figure 2:

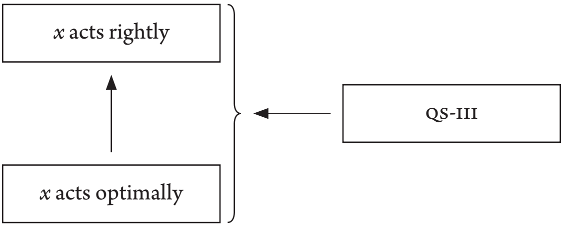


FIGURE 2

We will cover how one might fill in the QS-III box shortly. For the moment, I want to touch upon something I just said, namely, that in our case, facts of the form [*x* acts optimally] are "immediate whole or partial grounds" of corresponding facts of the form [*x* acts rightly]. There is some trouble about how to define immediate grounding, but the notion is sufficiently intuitive that for our purposes it suffices to take it as a working primitive. Following Kit Fine, we may nevertheless gloss the notion by saying that an *immediate ground* of a fact *F* is a ground of *F* whose grounding of *F* "need not be seen to be mediated."[16] In turn, we may then say that a *mere mediate ground* of *F* is a ground of *F* for which this is not so. For example, *A* is an immediate ground of [*A* or *B*] insofar as *A* may be seen to ground [*A* or *B*] without grounding any intermediary item. However, *A* is a mere mediate ground of [[*A* or *B*] or *C*], since *A* may be seen to ground [[*A* or *B*] or *C*], but only by way of first grounding [*A* or *B*].

While I here follow Fine in construing the distinction between immediate and merely mediate grounding in terms of facts, I often prefer to speak in terms of a partly corresponding distinction that holds at the level of properties and may be defined in terms of the fact-theoretic distinction as follows, using subscripted *f*s as variables ranging over facts: for some property, *G*-ness, to be an *immediate ground* of some other property, *F*-ness, is for *F*-ness and *G*-ness to nonvacuously satisfy the condition that necessarily, whenever a fact of the form [*x* is *G*], f_1, grounds a corresponding fact of the form [*x* is *F*], f_2, f_1 is an immediate ground of f_2. On the other hand, for *G*-ness to be a *mere mediate ground* of *F*-ness is for *F*-ness and *G*-ness to nonvacuously satisfy the condition that necessarily, whenever a fact of the form [*x* is *G*], f_1, grounds a corresponding fact of the form [*x* is *F*], f_2, f_1 is a mere mediate ground of f_2.

[16] Fine, "Guide to Ground," 50–51. Fine avoids saying that an immediate ground is one which is not mediated, for—as he demonstrates—such an account is susceptible to counterexamples.

There is another distinction between types of grounds worth bringing out, namely, that between *universal* and *parochial* grounds.[17] Here too I generally prefer to work with such a distinction at the level of properties, construed as follows: for some property, G-ness, to be a *universal ground* of some other property, F-ness—in other words, a *universal* F*-making property*—is for F-ness and G-ness to nonvacuously satisfy the condition that necessarily, whenever any x is F or G, $[x \text{ is } G]$ at least partly grounds $[x \text{ is } F]$. (Thus, a universal ground of F-ness is necessarily equivalent to F-ness: necessarily and for all x, x is F if and only if x is G.) On the other hand, a *parochial ground* of F-ness, G-ness, is a *merely occasional* F*-making property*: possibly some things are F at least partly in virtue of being G, but it is not necessary that everything that is F is F at least partly in virtue of being G.[18]

These distinctions are valuable to have on hand. To see why, consider the different ways we might try to fill in box QS-II in the blameworthiness instance of our explanatory structure for some individual, S, in figure 3:

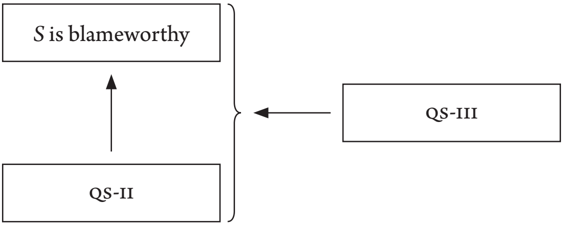


FIGURE 3

To fill in box QS-II here would be to offer an answer to the question "What are the conditions under which S is blameworthy?" In response to this question, we might naturally expect a long and multifarious list of the ever-so-many properties S might have in virtue of the possession of which a person might be

[17] I have not seen the notion of a parochial ground explicitly demarcated elsewhere. Rosen, however, does use the term "universal right-making feature," and the work done by "universal" in this expression of his is the work done by "universal" in mine. See, e.g., Rosen's discussion of Derek Parfit's metanormative views in the former's "Real Definition," 207n24.

[18] Alternatively, we might have appealed to fact forms to construe the distinction between universal and parochial grounds as one obtaining at the level of facts rather than that of properties. As I allude to above in the case of immediate/mere-mediate grounding (by way of saying the property-theoretic distinction "partly" corresponds to the fact-theoretic distinction), these ways of construing the distinction do not correspond perfectly, for the former construal affords us the ability to countenance universal and parochial grounds that have no natural correlates on the latter construal. Still, working with the property-theoretic construal of the distinction makes things easier and suffices for all purposes for which we shall be needing such a distinction.

blameworthy on a given occasion: having lied, having stolen, having killed, etc. Plausibly, each such property would be a mere parochial ground of blameworthiness, since not all who are blameworthy are blameworthy in virtue of, e.g., having lied.

Would the aforementioned blameworthy-making properties be mediate or immediate grounds of blameworthiness? A theorist of blameworthiness could go either way on this, but a common approach to theorizing blameworthiness—in fact, to theorizing normative properties generally—would lead us to say that such properties are mere mediate grounds of blameworthiness. The approach I have in mind would be to say that some property, *G*-ness, is *the unique universal and immediate ground* of blameworthiness, and the various aforementioned parochial grounds of blameworthiness ground blameworthiness only ever by way of grounding *G*-ness. Call this the *Principlist Approach* to theorizing normative properties.

We have already seen a view that conforms to the Principlist Approach, namely, the nonnaturalistic version of consequentialism considered above. That view holds that optimality is a universal and immediate ground of rightness, for it holds that necessarily and for all *x*, *x* is right if and only if *x* is optimal, and right directly in virtue of being optimal. The Principlist Approach to theorizing blameworthiness, then, would be to find some property, *G*-ness, that stands to blameworthiness as optimality stands to rightness and that stands to the many and varied parochial grounds of blameworthiness as optimality stands to the many and varied parochial grounds of rightness.

The major attraction in taking the Principlist Approach to theorizing a given normative property, *F*-ness, is that such an approach, if successful, would seem to simplify the task of answering the *F*-instance of QS-III: "Why are the conditions of *F*-ness as they are?" That is because in taking a Principlist Approach to theorizing *F*-ness, one seeks a partial answer to this question in the form of some unique universal and immediate ground of *F*-ness, *G*-ness, which is such that all other grounds of *F*-ness—the many and varied parochial grounds—are grounds of *F*-ness precisely *because* they are grounds of *G*-ness. To be sure, the discovery of a property like *G*-ness would not leave us with a complete answer to the question of why the conditions of *F*-ness are as they are, for it would remain to be said what it is in virtue of which *G*-ness itself is a condition of *F*-ness. Nevertheless, in discovering *G*-ness, we would thereby discover an explanation as to why *every other* condition of *F*-ness is a condition of *F*-ness. Needless to say, such a discovery would seem to constitute a significant explanatory success.

I have been discussing the reasons for taking a Principlist Approach to theorizing normative properties, but an analogous case can be made for taking a Principlist Approach to theorizing certain putative relation-response properties, normative or not. That is because a great many putative relation-response properties seem never to be possessed fundamentally: no one is ever brutely trustworthy—rather, people are trustworthy in virtue of being, e.g., historically reliable and well-intentioned truth-tellers; no one is ever brutely awe-inspiring—rather, people are awe-inspiring in virtue of being, e.g., extremely skilled in this or that activity; no one is ever brutely lovable—rather, people are lovable in virtue of being, e.g., extremely magnanimous or kind. Indeed, each such relation-response property, *F*-ness, would appear to have many and varied parochial grounds, just like normative properties generally. And to any theorist of *F*-ness, this cries out for explanation: what is it about these many and varied parochial *F*-making properties that makes them *F*-making? The desire for a unifying answer makes a Principlist Approach look attractive.

Just a moment ago, I said that the Principlist Approach to theorizing *F*-ness, if successful, would not by itself supply a complete answer to the question of why the conditions of *F*-ness are as they are, for that approach would not by itself explain why the unique universal and immediate ground of *F*-ness, *G*-ness, is a condition of *F*-ness.[19] Philosophers who have adopted the Principlist Approach to theorizing normative properties have supplied different sorts of answers here, corresponding to the different sorts of answers ground-theorists have offered to the question of how to ground grounding facts generally. We have finally circled back to the question of how to fill in QS-III boxes.

We just witnessed one means of grounding a certain class of grounding facts, namely, facts like [*A* grounds *C*], where *A*'s grounding of *C* is mediated by *A*'s grounding of *B*, which in turn grounds *C*. Here, [*A* grounds *C*] is grounded in at least two facts, namely, [*A* grounds *B*] and [*B* grounds *C*]. Some may

[19] What is more, positing an intermediary grounding property like *G*-ness would create the need for an explanation as to why the many and varied parochial grounds of *F*-ness are grounds of *G*-ness. In other words, though we give an answer as to what it is in virtue of which the many and varied parochial grounds of *F*-ness are such—namely, that they are such because they ground *G*-ness, which itself grounds *F*-ness—we have not yet answered the question of what it is in virtue of which those many and varied parochial grounds of *G*-ness are such. This may seem to undermine any advantage we might have thought we had gained by positing *G*-ness; do not all of our same problems arise anew at this new level we have introduced? Have we not merely shifted the bump in the explanatory rug? No—or at least not if we have found a good candidate to play the role of our universal and immediate ground. That is because a good candidate for the role of universal and immediate ground will be one whose nature makes it very clear why the many and varied parochial grounds of *F*-ness are grounds of *G*-ness. The thought is that it should be easier to see why those grounds of *F*-ness are grounds of *G*-ness than it is to see why they are grounds of *F*-ness. And if it is in turn easier to see why *G*-ness might be a ground of *F*-ness than it is to see why the many and varied parochial grounds of *F*-ness are grounds of *F*-ness, then we have surely made explanatory progress by discovering *G*-ness, since it is an illuminating intermediary.

also wish to say a third fact is required to ensure that these two facts ground [*A* grounds *C*], namely, [It lies in the essence of grounding to be transitive]. If we supplement the picture in this way, we arrive at an instance of a more general approach to grounding grounding facts, which, broadly and basically, is to appeal to essences. More specifically, *essentialists* say that facts about what grounds what—e.g., [*A* grounds *C*]—are themselves at least partly grounded in facts about the essences of one or more of the constituents of those facts, i.e., either the ground*ers*—*A*, in our example—or the ground*eds*—*C*, in our example—or, as we are here supposing, the grounding relation itself.[20] Why does *A* ground *C*? Because *A* grounds *B*, and *B* grounds *C*, and because it lies in the nature of grounding itself that if *A* grounds *B* and *B* grounds *C*, then *A* grounds *C*. Facts about essences, on the other hand—or relevant subpluralities thereof—are frequently supposed by essentialists to be ungrounded.[21]

The essentialist's approach to grounding grounding facts is the most relevant one for our discussion to come, and so I will not consider other approaches to grounding grounding facts—i.e., to filling in a QS-III box in our explanatory structure—save for a brief consideration of another such possibility at the end of section 4.5.

Let us recap. We began this section by introducing Rosen's framework for theorizing blameworthiness. We then considered a generalization of that framework for theorizing putative relation-response properties generally, i.e., an explanatory structure that any comprehensive theory of any putative relation-response property, *F*-ness, ought to guide us in filling out, if only in sketch. We then focused on examining different ways of filling out two nodes of that structure and in the process discussed the Principlist Approach to theorizing normative properties, as well as how and why one might adapt it for the purposes of theorizing putative relation-response properties generally.

We have covered a lot of ground. Let us turn now to our main topics of discussion, keeping our framework and its accompanying distinctions in mind as we go.

[20] Strictly speaking, one might take an essentialist line on the grounds of some grounding facts without taking that line on all.

[21] See, e.g., Rosen, "Ground by Law":

> The essentialist laws are fully satisfying unexplained explainers. If we ask why [*p*] grounds [$p \vee q$], we can answer: "Because it lies in the nature of disjunction that disjunctions are grounded in their true disjuncts." But if we ask why *this* is so, all we can say is: "That's just the nature of disjunction." That's not an answer. It's just a way of saying that when the question is why something has the constitutive essence it has, no answer is possible or necessary. The explanatory buck stops here. (291)

For a similar approach, see Dasgupta, "Metaphysical Rationalism."

2. Disrespect for Word Structure: Widespread? Widely Endorsed?

David Shoemaker reports that what he calls "Response-Independence theories of blameworthiness" are "much more popular" than what he calls "Response-Dependence theories of blameworthiness."[22] The way Shoemaker draws the distinction, Response-Independence theories of a given form of blameworthiness by definition hold that that form of blameworthiness is or is reducible to some property or properties in virtue of whose possession one merits a given form of blame.[23] The sort of properties Shoemaker has in mind are, to use an example he discusses, properties like *having knowingly and voluntarily acted badly from ill will while in control, appropriate historical conditions obtaining*.[24] The Response-Independence theorist of a given form of blameworthiness thus regards that form of blameworthiness as being or as being reducible to a property that lacks faithful relation-response structure, as the foregoing property clearly does. On the other hand, Shoemaker tells us that the much less popular sort of theories—the Response-Dependence theories of (this or that form of) blameworthiness—by definition identify (that form of) blameworthiness with or take it to be reducible to some faithful relation-response property or other. For Shoemaker, that property is *meriting anger* (*of a certain special variety*); for D'Arms and Jacobson, it is *being an appropriate target of guilt*; for Rosen, it is *being an appropriate target of resentment*.[25]

I have thus far stated only the constraints that Shoemaker takes each type of theory to place on possible answers to the blameworthiness instance of QS-I, namely, "What is it for something to be blameworthy?" There are other distinguishing features of Response-Independence and Response-Dependence theories, by Shoemaker's lights. In fact, Shoemaker regards each type of theory as placing constraints on possible answers to the blameworthiness-instances of QS-II and QS-III as well. We will consider these additional constraints in sections 4 and 5.

It is not too difficult to adduce examples of prominent philosophers of blameworthiness speaking as though they endorse the sort of disrespect for word structure that Shoemaker bakes into his definition of Response-Independence theories of blameworthiness. Consider the following examples.

[22] Shoemaker, "Response-Dependent Responsibility," 483.

[23] Shoemaker, "Response-Dependent Responsibility," 498.

[24] Shoemaker, "Response-Dependent Responsibility," 506.

[25] Shoemaker, "Response-Dependent Responsibility," 508; D'Arms and Jacobson, "The Motivational Theory of Guilt (and Its Implications for Responsibility)," 15; and Rosen, "The Alethic Conception of Moral Responsibility, 72–73.

Jules Coleman and Alexander Sarch appear to confirm Shoemaker's judgment as to the popularity of what Shoemaker refers to as Response-Independence theories of blameworthiness, for they tell us that they themselves endorse "the standard view" of blameworthiness according to which it is "a *reason* or a *ground* that explains why blaming . . . would be justified."[26] Thus, for Coleman and Sarch, a person is first blameworthy, and only thereafter (in the order of explanation) are they a justified target of blame. But then blameworthiness must be distinct from the property *being a justified target of blame* because it is prior to it. Thus blameworthiness, for Coleman and Sarch, cannot be *this* relation-response property, namely, being a justified target of blame. But nor do they appear to think it any other genuine relation-response property, for they frequently imply that they take blameworthiness to be or to be reducible to *being culpable for wrongdoing*.[27] Being culpable for wrongdoing may itself appear to be or to partly consist in a genuine relation-response property, namely, culpability. Yet Coleman and Sarch appear to regard culpability as susceptible of analysis in terms of "certain facts about one's agential relationship to the doing or omitting—for example, the fact that it was the product of a defective character, wicked intentions, a bad will, or some other kind of moral failing of the agent."[28] Such an analysis "analyzes out" culpability's relation-response structure and is therefore unfaithful as an analysis of culpability. Since culpability is the only putative relation-response property constitutively involved in the property of being culpable for wrongdoing, to analyze blameworthiness as culpability for wrongdoing when culpability is itself analyzed unfaithfully would be to analyze blameworthiness unfaithfully in turn.

[26] Coleman and Sarch, "Blameworthiness and Time," 101.

[27] Coleman and Sarch imply this by arguing that blameworthiness does not diminish with the mere passage of time entirely on the grounds that culpability for wrongdoing does not diminish with the mere passage of time. One might be inclined to interpret Coleman and Sarch as merely affirming a kind of covariation here between degree of blameworthiness and degree of culpability for wrongdoing, while maintaining that blameworthiness is nevertheless something distinct. But in light of their aforementioned view of blameworthiness, it is more natural to read them as assuming that insofar as culpability for wrongdoing is itself a "ground" or "reason" that explains why blame would be justified, culpability for wrongdoing *just is* blameworthiness. These properties, for them, appear to play the same role.

Alternatively, you may suspect that "being a justified target of blame," in Coleman and Sarch's idiolect, means something distinct from "being a fitting target of blame" or "being an apt target of blame" and then suppose that they regard blameworthiness as being a fitting target of blame, which itself grounds the *distinct* status of being a justified target of blame. But Coleman and Sarch explicitly deny any equivalence between blameworthiness and being a fitting or apt target of blame. Thus, the option of reading them as affirming these other faithful relation-response structures for blameworthiness is not available.

[28] Coleman and Sarch, "Blameworthiness and Time," 103.

T. M. Scanlon, on the other hand, tells us that "to claim that a person is *blameworthy* for an action is to claim that the action shows something about the agent's attitudes that impairs the relations that others can have with him or her."[29] Thus it appears—at least on the basis of this remark and others like them—that, for Scanlon, what it is for an agent, *S*, to be blameworthy for an action, *A*, is for *S*'s *A*-ing to indicate (or perhaps to flow from) *S*'s possession of a relevant set of relation-impairing attitudes. But notice that this description of Scanlonian blameworthiness makes no reference to any sort of response that might be appropriate towards *S* on the basis of *S*'s action or *S*'s relation-impairing attitudes. It certainly seems then that Scanlonian blameworthiness lacks faithful relation-response structure.

And finally there is Michael McKenna, who tells us that "blaming another for something she has done is primarily, albeit not exclusively, a matter of responding in a distinctive fashion to the perceived *morally objectionable quality of an agent's will as manifested in her blameworthy behavior*," where the quality of will McKenna takes to be morally objectionable is the "axiological" property of *being morally ill*.[30] In other words, S_1's blaming of S_2 is primarily a matter of S_1's responding to what S_1 perceives to be S_2's morally ill will. But then he tells us just a page later that "blaming is most fundamentally a response to perceived *blameworthiness*."[31] How can McKenna think that blame is "primarily" a matter of responding to perceived *morally ill will* yet also "fundamentally" a matter of responding to perceived *blameworthiness*? Presumably he can think this only if he thinks that there is no difference between these things. For McKenna, for *S* to be blameworthy for *A*-ing seems just to be for *S*'s *A*-ing to manifest morally ill will. But again, the property *having morally ill will* seems to lack faithful relation-response structure.[32]

I regard it as certain that Coleman and Sarch do in fact endorse an unfaithful analysis of blameworthiness. On the other hand, I regard it as highly probable that Scanlon at least is simply speaking loosely, for he immediately follows up his

[29] Scanlon, *Moral Dimensions*, 128.

[30] McKenna, "Directed Blame and Conversation," 122–23 (emphasis added).

[31] McKenna, "Directed Blame and Conversation," 123 (emphasis added).

[32] While the reading offered in the main text strikes me as faithful to McKenna's words, a nearby alternative reading would have him *identifying* a will's being morally objectionable with that will's being morally ill, rather than taking the latter to explain the former. On that reading, McKenna might better be read as offering a faithful analysis of blameworthiness, provided McKenna also understands the response type of *objection* to constitute a faithful analysans of the response type of *blame*. That there is ambiguity in how best to read McKenna here is not a problem for the case I am making; on the contrary, it further supports the point I am about to make in the main text, namely, that it is often unclear whether authors who speak as though they reject faithful analyses of blameworthiness really do.

aforementioned statement of what it is to claim that somebody is blameworthy by saying, "To *blame* a person is *to judge him or her to be blameworthy* and to take to your relationship with him or her to be modified *in a way that this judgment of impaired relations holds to be appropriate*."[33] We noted above that Scanlon's original description of the content of a claim or judgment of blameworthiness makes no mention of responses, appropriate or otherwise, and we accordingly read him as affirming the identity of blameworthiness with the unfaithful property of having acted from (or in a way that indicates) relation-impairing attitudes. And yet just one sentence later, Scanlon speaks as though a judgment of blameworthiness *does* consist at least partly in a judgment as to the appropriateness of a certain blaming response. Well, does it, or doesn't it? If it does, then perhaps Scanlon does not really regard blameworthiness as having acted from (or in a way that indicates) relation-impairing attitudes; perhaps instead, he regards this latter property as a ground or condition in virtue of the satisfaction of which a person is worthy of certain kinds of response—namely, behavioral or attitudinal modifications of certain sorts—the worthiness of which responses is itself the true bearer of the title "blameworthiness."

Thus, while I offer the foregoing examples primarily as a way of helping you to see more clearly what disrespect for word structure looks like, I offer them secondarily as a way of indicating where I stand with respect to the matter of whether—as Shoemaker and Coleman and Sarch report—such disrespect is widespread and widely endorsed. In short, whether or not they are right that such disrespect is widespread, I hesitate to say that it is widely endorsed. Many philosophers (like Scanlon) seem to be either speaking carelessly or, if not carelessly, using "blameworthiness" in a loose or extended sense, i.e., to refer to what they in fact regard as the (perhaps universal and immediate) *ground* or *condition* of blameworthiness rather than blameworthiness itself. I offer further support for this hypothesis in section 4.

Still, Coleman and Sarch are not speaking loosely or carelessly. As such, I assume they regard themselves as having reasons to identify blameworthiness with or reduce it to an unfaithful property—though so far as I can see, they do not share any such reasons with us.[34] In the next section, I discuss the only

[33] Scanlon, *Moral Dimensions*, 128 (latter two emphases added).

[34] In note 27 above, I mention that Coleman and Sarch deny that blameworthiness and being a fitting target of blame are equivalent. This of course would be sufficient for these properties to be distinct. It may seem then that Coleman and Sarch do offer *some* reason to deny that blameworthiness has faithful relation-response structure, namely, that blameworthiness is inequivalent to one relation-response property that might have otherwise seemed apt to be identified with blameworthiness. But this would be to get the dialectic backward, since Coleman and Sarch *presuppose* that blameworthiness is an unfaithful property in

such reasons I have seen explicitly propounded, namely, an argument recently offered by D'Arms and Jacobson and another by Shoemaker, each of which purports to deduce an unfaithful analysis of prideworthiness from a popular combination of views about putative relation-response properties like it.

3. Reasons to Disrespect?

Consider the following passage from D'Arms and Jacobson:

> If to be prideworthy is to merit pride, and pride is even partly constituted by the thought that something is splendid and mine, then it seems to follow that for something to be prideworthy is just for it to be splendid and mine. But if the prideworthy can be understood via a pride-independent notion of *splendid and mine*, then … pride drops out of the explanation of the prideworthy.[35]

In a footnote attached to the first of these sentences, they add:

> At any rate, this is so if fittingness is tantamount to the truth of the emotion's constitutive thought. Indeed, cognitivism's ability to explain fittingness in this straightforward way is one of its features.

D'Arms and Jacobson's argument, rendered a bit more formally, seems to be this:

1_{DJ} For an object, *x*, to be prideworthy is for *x* to be worthy of pride.
2_{DJ} For an object, *x*, to be worthy of pride is for *x* to be a fitting target of pride.
3_{DJ} Each instance of pride is partly constituted by exactly one thought, and this thought is of the form _____ *is splendid and shiny*, where the blank is to be filled in by the target of that instance of pride.[36]
4_{DJ} For an object, *x*, to be a fitting target of pride is for *x* to be such as to render *x*-targeting instances of the thought that partly constitutes pride true.

Thus,

the ballpark of culpability for wrongdoing *before* setting out to argue for its inequivalence with being a fitting target of blame. It is this presupposition that I am saying Coleman and Sarch seem not to offer reasons for.

[35] D'Arms and Jacobson, "The Motivational Theory of Guilt (and Its Implications for Responsibility)," 12.

[36] D'Arms and Jacobson's toy example of pride's cognitive content is ____ *is splendid and mine*. (They borrow the example from Foot, "Hume on Moral Judgment.") I have replaced that content with ____ *is splendid and shiny*, since this latter content does not involve us in any complications having to do with indexical contents, as the former does.

c1$_{DJ}$ For an object, *x*, to be a fitting target of pride is for *x* to be splendid and shiny. (from 3$_{DJ}$ and 4$_{DJ}$)

Thus,

c2$_{DJ}$ For an object, *x*, to be prideworthy is for *x* to be splendid and shiny. (from 1$_{DJ}$, 2$_{DJ}$, and c1$_{DJ}$)

Let me state three assumptions. First, D'Arms and Jacobson employ the "to be *F* is to be *G*" locution, whereas I employ the "for *x* to be *F* is for *x* to be *G*" locution. I assume this is fine, exegetically speaking. Second, D'Arms and Jacobson need for the argument's locution of choice to impose a kind of transitivity, otherwise the argument has no hope of being valid. I assume this holds of my locution of choice. More specifically, I assume that if it is true that for *x* to be *F* is for *x* to be *G*, and it is also true that for *x* to be *G* is for *x* to be *H*, then it is also true that for *x* to be *F* is for *x* to be *H*. Third and finally, I assume that each statement of the form "for *x* to be *F* is for *x* to be *G*" that we will be considering in this paper is equivalent to a corresponding statement of the form "the property *F*-ness is or is reducible to the property *G*-ness."

Premise 1$_{DJ}$ is a truism. Premise 2$_{DJ}$ is not a truism, but it is a corollary of the popular view that worthiness (of the relevant sort) just is fittingness. In any case, it is not something I wish to question here. Premise 3$_{DJ}$ is an instance of *cognitivism about pride*: the view that pride is partly constituted by a thought with a certain distinctive content. Premise 4$_{DJ}$ is an instance of the *alethic conception of fittingness*: the view that what it is for instances of certain types of (psychological) response to be fitting is for their constitutive thought to be true.

On any natural way of filling in the details, c2$_{DJ}$ conflicts with my core thesis, since prideworthiness clearly lacks faithful relation-response structure if prideworthiness is or is reducible to being splendid and shiny.[37] But that is not the worst of it. Premises 1$_{DJ}$–4$_{DJ}$ collectively amount to a theory of prideworthiness, and analogous theories can and have been offered for other putative relation-response properties. Indeed, packages of views like these are popular.[38] Thus,

[37] I regard as unnatural the way of filling in the details according to which prideworthiness has multiple distinct types of structure *fundamentally*. Still, I would be happy to read my core thesis as ruling out this sort of story and so would be happy to say that c1$_{DJ}$ and c2$_{DJ}$ conflict with my core thesis no matter how naturally or unnaturally you fill in the details.

[38] For a nice sampling of recent theories of blameworthiness that endorse analogous packages of theses, see the discussion in Clarke and Rawling, "True Blame," 3–4. Of course, such an approach to theorizing certain putative relation-response properties cannot straightforwardly be adopted for *all* such properties since in many cases the type of response at issue, not being psychological in kind, will not sensibly be susceptible of a cognitivist construal. But the approach is quite popular for such properties when the type of response at issue is, e.g., a reactive attitude, and it may naturally be thought to apply in the case of putative relation-response properties involving certain other nonreactive attitudes like, say, believability.

if D'Arms and Jacobson's argument is sound, my core thesis conflicts with a popular approach to theorizing a greater number of putative relation-response properties than just prideworthiness.

The trouble for this formulation of D'Arms and Jacobson's argument is that it is invalid: premises 3_{DJ} and 4_{DJ} do not entail $C1_{DJ}$, and $C2_{DJ}$ does not follow without $C1_{DJ}$. Premises 1_{DJ}–4_{DJ} *do* entail that for *x* to be a fitting target of pride—and thus for *x* to be prideworthy—is for *x* to be such as to render *x*-targeting instances of pride's constitutive thought true. In other words, premises 1_{DJ}–4_{DJ} do yield the result that prideworthiness is or is reducible to being such as to render appropriate instances of pride's constitutive thought true. But this claim neither is nor entails the claim that prideworthiness is or is reducible to being splendid and shiny.

One natural way of repairing the argument would be the following. First, suppose something rather natural for a cognitivist about pride to suppose, namely, that pride is necessarily partly constituted by its distinctive thought (and necessarily is not partly constituted by any other thought); let this be premise 3^*_{DJ}; then suppose *intensionalism about property individuation*—the thesis that any two necessarily coextensive properties are identical; and let this be premise 5_{DJ}. It now follows, given what has been said, that prideworthiness is identical to being splendid and shiny, since it now follows that necessarily and for all *x*, *x* is prideworthy if and only if *x* is splendid and shiny.[39]

The trouble for this way of repairing the argument is that intensionalism is implausible as an account of property individuation. In fact, our very own case supplies us with good reason to reject it. That is because it is extremely plausible that on the picture laid out, facts about prideworthiness are always grounded in corresponding facts about what is splendid and shiny, whereas facts about what is splendid and shiny are of course not thus grounded, since grounding is irreflexive. On the pictures of fact and property individuation that I prefer, this alone would suffice to show that the property of being prideworthy and the property of being splendid and shiny are distinct. On more fine-grained conceptions—à la Rosen's—we need to say more: in particular, we need to say that facts about prideworthiness are only ever *partly* grounded in corresponding facts about what is splendid and shiny.[40] But that, I submit, is eminently

[39] Strictly speaking, this follows only if we can validly infer from "necessarily and for all *x*, *x* is prideworthy if and only if *x* is splendid and shiny" to "prideworthiness and being splendid and shiny are necessarily coextensive." I shall assume we can.

[40] Suppose that to be a bachelor is to be an unmarried male. In that case, Rosen would say that for any bachelor, *S*, the fact [*S* is a bachelor] is wholly grounded in [*S* is an unmarried male] ("Metaphysical Dependence," 122–26, and "Real Definition," 199–200). Remarkably, he would also say that under such a supposition, the property of being a bachelor is identical to the property of being an unmarried male ("Metaphysical Dependence," 125n14,

plausible given the conception of prideworthiness that we are supposing. On that conception, what it is for *x* to be prideworthy is for *x* to be such as to render *x*-targeting instances of pride's constitutive thought true. But this makes prideworthiness a higher-order property, i.e., the property of having some other property. Specifically, it is the property of having that property, whatever it is, the possession of which by any *x* renders *x*-targeting instances of pride's constitutive thought true. Thus prideworthiness is not just a higher-order property but a *generalized* higher-order property: it is not the property of having some particular property specified *de re*, such as redness or sharpness, but is rather the property of having *that property, whatever it is*, the possession of which by any *x* renders *x*-targeting instances of pride's constitutive thought true. But this means that facts about prideworthiness must be grounded *both* in a corresponding fact about something's being splendid and shiny *and* in the fact that pride's constitutive thought is that its target is splendid and shiny. The complete grounds of prideworthiness must always include this latter, "bridging" fact.

Thus the toy theory of prideworthiness encapsulated by premises 1_{DJ}–4_{DJ}—i.e., the theory that combines (i) the identification of worthiness (of the relevant sort) with fittingness, (ii) an instance of cognitivism about pride, and (iii) the alethic conception of fittingness—itself tells against intensionalism about property individuation precisely because it commits one to an apparent ground-ordering between necessarily coextensive properties that plausibly entails their distinctness. Thus anybody who accepts that toy theory of prideworthiness ought to reject our amended version of D'Arms and Jacobson's argument. And the commitments of that theory that imply the counterexample to intensionalism are not distinctive to it: analogous theories—of blameworthiness, of trustworthiness, etc.—imply analogous counterexamples. I therefore conclude that D'Arms and Jacobson's argument fails to establish that this popular approach to theorizing putative relation-response properties commits one to theorizing such properties unfaithfully.

D'Arms and Jacobson are not the only ones to argue for this result, however. Let us turn now to consider the following passage from Shoemaker:

and "Real Definition," 202–5, 190n2). This is because for Rosen, the property of being *F* = the property of being *G* if it lies in the nature of *F*-ness that whatever is *F* or *G* is *F* wholly in virtue of being *G*, and this latter condition, according to Rosen, holds if and only if to be *F* is to be *G*. Importantly, Rosen thinks that if we do not have whole grounding here, then we do not have this property identity ("Real Definition," 207n24). In the case of prideworthiness presently conceived, it seems to lie in its nature that anything that is prideworthy or splendid and shiny is prideworthy in virtue of being splendid and shiny. Thus Rosen's account would yield the result that the property of being prideworthy *just is* the property of being splendid and shiny *if* we were here dealing with *whole* grounding. But as I argue in the main text, we are not. And if not, then we are dealing with distinct properties here.

> Resentment is almost universally taken to be what D'Arms and Jacobson call a "cognitively sharpened" emotion, namely, anger plus a judgment, e.g., that the to-be-resented agent culpably wronged you.... But if that is the correct characterization of our paradigm responsibility emotion, then the game has been given away to the response-independent theorist, for resentment *presupposes* the responsibility of the resented agent. If you deliberately step on my foot, and my resentment includes the judgment that you culpably wronged me, then what makes my response apt is just that that constitutive judgment is *true*, and your judgment will be rendered true *by your antecedent responsible blameworthiness*, as that is just what a judgment of culpable wronging amounts to. Cognitive theories of blame beg the question in favor of response-independence.[41]

The argument presented in this passage is certainly enthymematic, and I confess I am not entirely certain how best to fill in its details. Upon first encountering this passage, it seemed to me that Shoemaker was arguing along more or less the same lines as D'Arms and Jacobson, albeit in the case of blameworthiness rather than prideworthiness. If that were right, then what I had to say about D'Arms and Jacobson's argument should apply just as well to Shoemaker's.

But there is another way to read the passage according to which it presents something distinct.[42] On that reading, a more perspicuously rendered formulation of Shoemaker's argument might go roughly as follows:

1_S For an object, *x*, to be a fitting target of blame is for *x* to be such as to render *x*-targeting instances of the thought that partly constitutes blame true.

2_S Each instance of blame is partly constituted by exactly one thought, and this thought is of the form _____ *culpably wronged*, where the blank is to be filled in by the target of that instance of blame.[43]

3_S If 1_S and 2_S, then whenever any object, *x*, is a fitting target of blame, [*x* is a fitting target of blame] is (at least partly) grounded in [*x* culpably wronged].[44]

[41] Shoemaker, "Response-Dependent Responsibility," 314.

[42] I thank an anonymous referee for encouraging roughly this reading of Shoemaker, which upon reflection seems to me superior to the reading I initially had.

[43] As with the example that D'Arms and Jacobson borrowed from Foot above, Shoemaker's example of blame's thought content, namely, ____ *culpably wronged me*, is partly indexical. As before, I opt to simplify my presentation of the argument by removing the indexical element, leaving ____ *culpably wronged*.

[44] Of course, the antecedent of this premise, "If 1_S and 2_S" is strictly speaking ungrammatical (as is that of premise 5_S), given that "1_S" and "2_S" are names of premises and not themselves

4_S For an object, *x*, to culpably wrong is (at least in part) for *x* to be blameworthy.

5_S If 4_S and for some object, *x*, [*x* is a fitting target of blame] is (at least partly) grounded in [*x* culpably wronged], then [*x* is a fitting target of blame] is (at least partly) grounded in [*x* is blameworthy].

6_S If for some object, *x*, [*x* is a fitting target of blame] is (at least partly) grounded in [*x* is blameworthy], then blameworthiness is distinct from *being a fitting target of blame*.

7_S If blameworthiness is distinct from *being a fitting target of blame*, then blameworthiness is response independent.

But,

8_S Some object, *x*, is a fitting target of blame.

Thus,

$C1_S$ For some object, *x*, [*x* is a fitting target of blame] is (at least partly) grounded in [*x* culpably wronged]. (from 1_S, 2_S, 3_S, and 8_S)

Thus,

$C2_S$ For some object, *x*, [*x* is a fitting target of blame] is (at least partly) grounded in [*x* is blameworthy]. (from 4_S, 5_S, and $C1_S$)

Thus,

$C3_S$ Blameworthiness is distinct from *being a fitting target of blame*. (from 6_S and $C2_S$)

Thus,

$C4_S$ Blameworthiness is response independent. (from 7_S and $C3_S$)

I wish briefly to note and justify three small ways my formulation departs from Shoemaker's. First, my formulation is framed in terms of blame, whereas Shoemaker's is framed in terms of resentment. This is a mere simplification, and a harmless one at that.[45] Second, Shoemaker's formulation speaks of aptness,

sentences. This is a mere infelicity of presentation, for I here intend "If 1_S and 2_S" as shorthand for the unwieldy phrase that would result by replacing "1_S" and "2_S," as they appear in it, with the sentences that state the premises themselves.

[45] In the sentences preceding this passage, Shoemaker indicates that rather than considering how things stand if we adopt a cognitivist approach to theorizing blame and an alethic approach to theorizing blame's fittingness, he focuses on resentment out of the convictions that there are many different blaming response types, and resentment is commonly regarded as a paradigmatic such type ("Response-Dependent Responsibility," 313–14). With this we may happily agree, and we could—if we wished—replicate our discussion of Shoemaker's argument for any such type. But this would be tedious, and what is more, our already complex rendering of Shoemaker's argument would become even more complex were we to focus on resentment, for then we would need in turn to speak not of blameworthiness *simpliciter* but of what we might call "resentment blameworthiness." While I regard this degree of presentational rigor as generally desirable and for that reason do

whereas mine speaks of fittingness. But my decision here is informed by Shoemaker's own tendency to treat these things as the same in relevant contexts. Third, Shoemaker speaks not of resentment's constitutive thought but of its constitutive judgment. This difference will not matter.

Let us consider the argument's premises. Premises 1_S and 2_S are familiar: they are respectively just a blame-centric instance of the alethic conception of fittingness and an instance of cognitivism about blame. Premise 3_S is a consequence of the plausible thought, on display in my foregoing criticism of D'Arms and Jacobson's argument, that for any true thought, *t*, that *p*, the fact [*t* is true] will be (at least partly) grounded in [*p*]. Premise 4_S is something I take Shoemaker to be committed to by way of what he commits to when he says that "a judgment of culpable wronging *amounts to*" a judgment of "antecedent responsible blameworthiness."[46] Shoemaker's wording here is a bit particular, but the thought seems to be that for *x* to culpably wrong is (at least in part) for *x* to be blameworthy.

Premise 5_S looks plausible given the worldly conception of facts we are working with. The idea behind it is that if for some *x* to culpably wrong is (at least partly) for *x* to be blameworthy, then if [*x* culpably wrongs] (at least partly) grounds [*x* is a fitting target of blame], so too presumably would [*x* is blameworthy]. Recalling my preferred, slightly more coarse-grained conception of facts and properties, this alone would suffice to show that blameworthiness is distinct from being a fitting target of blame, as premise 6_S says. As noted above, more must be said if we embrace Rosen's more fine-grained conception of fact and property individuation. But I do not wish to challenge premise 6_S and so am content to work with it rather than with a version that more studiously establishes that the grounding of [*x* is a fitting target of blame] by [*x* is blameworthy]—as Shoemaker here conceives of it—meets Rosen's criteria for implying that blameworthiness and being a fitting target of blame are distinct.

I am least confident in attributing premise 7_S to Shoemaker, yet something like 7_S seems to be needed in order to proceed, as Shoemaker appears to, from the implicit result that blameworthiness is distinct from (because prior to) being a fitting target of blame to the claim that blameworthiness is response independent. After all, to derive that blameworthiness is distinct from being a fitting target of blame is not *yet* to derive that blameworthiness cannot be identified with or reduced to some other genuine relation-response property. Presumably, Shoemaker is thinking that being a fitting target of blame is the best or only candidate for a faithful analysis of blameworthiness, and so if it

adopt it in my discussion of Shoemaker's and D'Arms and Jacobson's own views in section 4, the formalization of Shoemaker's argument that we are presently considering is already complex enough without this additional complication. Hence the simplification.

[46] Shoemaker, "Response-Dependent Responsibility," 314 (emphasis added).

cannot work, no other genuine relation-response property deserves the role. Premise 8_s, on the other hand, is clear and needs no defense in this context.

The argument is valid, and on the plausible assumption that if a property is response independent, it lacks genuine relation-response structure, $\text{C}4_s$ implies that blameworthiness lacks genuine relation-response structure.[47] This argument is evidently distinct from D'Arms and Jacobson's, and if Shoemaker is correct about its upshot—namely, that "cognitive theories of blame beg the question in favor of response-independence"—it purports to deliver the result that if we embrace the popular approach to theorizing blameworthiness, which embeds the combination of cognitivism about blame plus an alethic conception of blame's fittingness, we must theorize blameworthiness unfaithfully.

Fortunately, if the foregoing argument is indeed Shoemaker's, then I think his judgment about its upshot is mistaken: the combination of an alethic conception of blame's fittingness (namely, 1_s) plus the particular version of cognitivism about blame that Shoemaker focuses on (namely, 2_s) does *not* require us to say that blameworthiness is distinct from being a fitting target of blame—not these premises by themselves, anyhow. And not even by themselves together with the relatively uncontroversial premises 3_s, 5_s, and 8_s; nor by all of these together with the perhaps more controversial premises 6_s and 7_s. Our formulation of the argument makes this much clear, for according to it, the conclusion that blameworthiness is distinct from being a fitting target of blame (namely, $\text{C}3_s$) relies crucially on $\text{C}2_s$—namely, that [*x* is a fitting target of blame] is (at least partly) grounded in [*x* is blameworthy]—which in turn relies crucially on 4_s, namely, that for an object, *x*, to culpably wrong is in part for *x* to be blameworthy. But 4_s is an independent premise, not delivered by any other of premises 1_s–8_s.

Still, it may be that proponents of the rest of premises 1_s–8_s *ought* to embrace 4_s. Shoemaker himself embraces 4_s or something like it insofar as he wishes to analyze culpability in terms of blameworthiness—a project he regards as part of the broader project of giving a Response-Dependence theory of responsibility.[48] I myself am partial to this project, provided we understand it in the way I propose to understand Response-Dependence theories of properties generally

[47] In section 4, I reveal that I take this assumption to be an analytic truth, given what is generally meant by "response independent."

[48] Of course, as exemplified by the Response-Independence view that Shoemaker here considers, merely analyzing a putative relation-response property (in this case, culpability) in terms of another putative relation-response property (in this case, blameworthiness) will not suffice for giving a faithful theory of the former, since the view at hand proceeds to say that the analysans here is itself to be understood as a response-independent property. To embrace a faithful theory of a putative relation-response property, *F*-ness, one cannot simply affirm an analysis of that property in terms of another, nearby-seeming putative relation-response property; rather, one must also say that faithful relation-response structure

in sections 4 and 5, and so myself am attracted to something like premise 4_S. The point I have made thus far is not that 4_S is false but merely that proponents of cognitivism about blame plus an alethic conception of blame's fittingness are not, apparently *contra* Shoemaker, committed *as such* to 4_S or to anything like it.

But what might 4_S-sympathizers like myself say in the face of Shoemaker's argument? Must we embrace Shoemaker's conclusion that cognitivism about blame, an alethic conception of blame's fittingness, 4_S, and the rest together imply that blameworthiness is distinct from being a fitting target of blame? No, for we might instead simply reject the specific version of cognitivism that Shoemaker here apparently assumes is mandatory for cognitivists about blame, namely, premise 2_S. That version of cognitivism commits one to the idea that culpability figures in the content of blame's constitutive thought. But cognitivists about blame who are partial to something like 4_S can reject this. Indeed, they *should* reject this if they wish also to say that blameworthiness is or is reducible to being a fitting target of blame.[49] More specifically, such theorists should not say that blame's constitutive thought involves anything like that its target is blameworthy

is *ineliminable* from the original property's *final* analysis. I discuss how faithfulness relates to Response-Independence and Response-Dependence views further in sections 4 and 5.

[49] Rosen makes the same point when he writes:

> Why not just say that in addition to the thought that *A* was wrong and that *X* showed ill will in doing it, resentment of *X* for *A* involves the thought that *A was X's fault*, or that *X has no excuse*, or (what amounts to the same thing in this context), *X is blameworthy for A*?… This account would be disastrous for the Alethic View given its explanatory ambitions. The fundamental premise of the view is that when *X* is blameworthy for *A*, that is because the thoughts implicit in resentment are true of *X* and *A*. But if one of the thoughts implicit in resentment is just the thought *that X is blameworthy for A* (or some close equivalent), this would yield what amounts to an explanatory circle, according to which *X* is blameworthy for *A* because it's true that *X* is blameworthy for *A*. Of course this is not literally a circle—*p* because *p*—but it's just as bad. Just as *p* cannot explain *p*, it's true that *p* cannot explain *p*. Rather the order of explanation runs the other way: when a proposition *p* is true, *p* is true *in virtue of the fact that p*. (It's true that snow is white *because snow is white*.) Any account of the content of resentment according to which resentment involves thoughts about blameworthiness thus leads to absurdity when combined with the Alethic View. ("The Alethic Conception of Moral Responsibility," 80–81)

It is noteworthy that Rosen and Shoemaker differ in their understandings of what would follow from the conjunction of cognitivism about blame plus the alethic conception of blame's fittingness were blame's constitutive thought to predicate blameworthiness of its target. By Shoemaker's lights (as I have interpreted him), it would follow that blameworthiness is distinct from being a fitting target of blame and is therefore response independent. But Rosen does not go this way. Instead, Rosen holds fast to the claim that to be blameworthy is to be a fitting target of blame and, for that reason, is led to interpret the view at hand as committed to the claim that [*x* is blameworthy] is (at least partly) grounded in [It is true that *x* is blameworthy], which (as I discuss in note 50 below) Rosen finds problematic.

(or that its target is *F*, where to be *F* is at least in part to be blameworthy); for were they to do this, they would be led—by the reasoning on display in the arguments for C1_{S} and C2_{S} above—to say that for some *x*, [*x* is blameworthy] (at least partly) grounds [*x* is a fitting target of blame]. Were they also to hold that blameworthiness is or is reducible to being a fitting target of blame, they would then be forced to say that [*x* is blameworthy] (at least partly) grounds itself. In other words, such theorists would be caught in a circle of grounding, which is bad.

It is for effectively this reason that Rosen—himself a Response-Dependence theorist of blameworthiness who advances a cognitivist, alethic conception of blame and blameworthiness—opts not to imbue blame's constitutive thought's content with anything having to do with responsibility.[50] There are of course different options for doing this. Rosen's own account holds that blame's constitutive thought content is of the form _____ *deserves to suffer for doing A*. Alternatively, one might attempt to repurpose something in the ballpark of Shoemaker's example of a Response-Independence-theoretic conception of blameworthiness, cited earlier in section 2—namely, *having knowingly and voluntarily acted badly from ill will while in control, appropriate historical conditions obtaining*—and say that while this property is not itself identical to blameworthiness or that to which blameworthiness reduces, it is the condition that blame's constitutive thought presents its target as satisfying.

To be clear, I mention these alternative accounts of blame's constitutive thought's content not to affirm or defend either but simply to show that embracing the trio of cognitivism about blame, the alethic conception of blame's fittingness, and premise 4_{S}—namely, that to be responsible is (at least partly) to be blameworthy—does not force one to embrace a Response-Independence theory of blameworthiness. This result would follow only given *a particular version* of cognitivism of blame—namely, one that imbues its constitutive thought with blameworthiness-involving content—which those who embrace this trio of theses can, should, and (in the case of Rosen, at least) sometimes do reject. In other words, the popular approach to theorizing blameworthiness that we

[50] I say "effectively for this reason," for as may be seen in the passage cited in note 49 above, Rosen stops short of accusing the version of this view, which he therein considers of being circular, claiming instead that while that view is not literally committed to a circle, what it is committed to is just as bad, namely, that for some *x*, [*x* is blameworthy] is (at least partly) grounded in [It is true that *x* is blameworthy]. I am not certain why Rosen forgoes completing the circle, as it were, by observing that [It is true that *x* is blameworthy] would itself need to be grounded in [*x* is blameworthy], given the principle—which he himself accepts in the passage above—that facts of the form [It is true that *p*] are generally grounded in corresponding facts of the form [*p*]. In any case, I do think this principle—or at least a relevant analogue of it that holds for the truth of thoughts—is extremely natural, and so I do think the view in question implies circular grounding given extremely natural ground-theoretic assumptions.

have been considering does not by itself require one to analyze blameworthiness unfaithfully, *contra* Shoemaker.

Let us recap the results of this section. We considered two arguments for thinking that a popular approach to theorizing certain putative relation-response properties—namely, an approach that combines cognitivism about the property's response constituent with an alethic conception of the fittingness of responses of that type—requires one to theorize such properties unfaithfully. I argued that neither argument works as advertised. More specifically, I argued that D'Arms and Jacobson's argument is invalid as it stands and that a natural way of repairing it is not viable. I then argued that Shoemaker's argument does not in fact show that cognitivism about blame, together with an alethic conception of blame's fittingness, implies an unfaithful analysis of blameworthiness. Rather, this result follows only given a substantial additional premise (namely, 4_S), as well as a particular version of cognitivism about blame that cognitivists about blame can, should, and (in some cases) do reject.

4. RESPONSE INDEPENDENCE AND RESPONSE DEPENDENCE

In this penultimate section, I turn to the role that faithfulness plays in recent debates over Response-Independence (henceforth, "RI") and Response-Dependence (henceforth, "RD") theories of value properties. In particular, I draw upon our ground-theoretic framework from section 1 to argue that recent ways of drawing the distinction between RI and RD theories of such properties render that distinction partly merely verbal and otherwise unhelpfully arbitrary. Afterward, in section 5, I argue that embracing faithful analyses of putative relation-response properties does not require us to say the controversial things that self-proclaimed RD theorists of such properties typically say. In other words, faithful analyses of these properties come much more cheaply than has been suggested.

I begin in section 4.1 by depicting the grounding structure that I regard as obtaining whenever there obtains a fact involving the instantiation of at least a great many relation-response properties. Then, after a necessary terminological interlude in section 4.2, I show in section 4.3 that the self-proclaimed RD theorists we have been discussing—Shoemaker and D'Arms and Jacobson—accept that this same grounding structure obtains across a number of such kinds of cases. Then, in section 4.4, I show that RI theorists *also* accept this same grounding structure across these cases. In section 4.5, I draw my conclusions from the work done—namely, that the RI theorists under discussion differ from their RD-theoretic counterparts merely over which items in that grounding structure they denote by way of which expressions, and over which items they permit to occupy the QS-III position in the common grounding structure—and

I attempt to reveal what our results have shown about where the heart of the RI/RD debate really lies.

4.1. Common Ground(ing Structure)

In section 1, I observed that many putative relation-response properties seem never to be possessed fundamentally but rather are always possessed *in virtue of* the possession of other properties: no one is ever brutely blameworthy, for instance; rather, they are blameworthy in virtue of, e.g., having lied, stolen, murdered, etc. I further suppose that facts of the form [*A* grounds *B*] are themselves always grounded.

Thus, insofar as I say we ought to endorse only faithful analyses for putative relation-response properties, I am committed to supposing that for any such property, *F*-ness, a true theory of *F*-ness will situate *F*-ness facts in ground-theoretic explanatory structures as in figure 4:

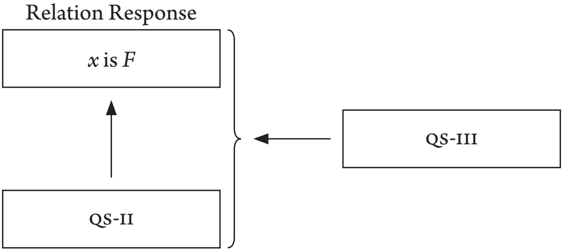


FIGURE 4

Read out, this graphic says: (i) facts of the form [*x* is *F*] are relation-response facts—i.e., facts involving a thing being *F*, where *F*-ness is a genuine relation-response property; (ii) facts of the form [*x* is *F*] are each grounded in some further fact or facts; and (iii) the grounding of each fact of the form [*x* is *F*] in such further fact or facts is itself grounded in some further fact or facts.

Presumptuously, I call this the "Common Grounding Structure," since I will shortly argue that, as they are defined by certain theorists, both RI and RD theories of putative relation-response properties share commitment to instantiations of their respective properties standing in grounding relations that together instantiate the Common Grounding Structure. But first, a necessary terminological interlude.

4.2. A Necessary Terminological Interlude

Where *F*-ness is "a value," D'Arms and Jacobson stipulate that *sentimentalism about* F*-ness* is the thesis that *F*-ness is response dependent, where a value is response dependent just insofar as it "cannot adequately be explained without appeal to the emotions."[51] For example, a response-dependent conception of funniness "identifies [it] with what causes or, more plausibly, what merits amusement."[52]

Notably, D'Arms and Jacobson restrict the scope of "response" as it appears in their use of "response dependent" to emotional responses alone. This has the potential to make for awkwardness insofar as I have intended and continue to intend for "response" to range over responses of all types, whether they be emotional, attitudinal, or behavioral. But I assume D'Arms and Jacobson would be happy to countenance a more expansive definition of "response dependent" corresponding to my more expansive sense of "response."[53] Speaking in that more expansive sense, we can say that what it is for a property to be *response dependent* is for it to be *response involving*, i.e., to have a structure that embeds fundamentally some type of response as a constituent. (In turn, we can say that what it is for a property to be *response independent* is for it to have a structure that is not fundamentally response involving.) By our definitions set out in the introduction, it follows that a property's having genuine relation-response structure suffices for its being response dependent.

The foregoing, I take it, is the standard way of defining these predicates as they apply to properties. What about the labels "RI theory" and "RD theory"? It would seem most natural to say that a theory of *F*-ness is an RI theory of *F*-ness just insofar as that theory says that *F*-ness is response independent, and *mutatis mutandis* for RD theories.

Notably, if we define things this way, it will turn out that I am an RD theorist wherever putative relation-response properties are concerned. That result is fine by me. But it implies—in conjunction with my earlier claim that it is "a straightforward deliverance of English" that we ought to analyze putative relation-response properties faithfully—that I am committed to its being a straightforward deliverance of English that we ought to embrace RD theories

[51] D'Arms and Jacobson, "Whither Sentimentalism?" 250.

[52] D'Arms and Jacobson's use of the term "value" suggests that they have in mind value properties, e.g., goodness, badness, blameworthiness, etc. However, they subsequently opt out of construing their preferred version of sentimentalism as a thesis about properties, opting instead to construe it as a thesis about value concepts (D'Arms and Jacobson, "Whither Sentimentalism?" 254). Still, they apply the language of "response-dependence" and "response-independence" to properties as well as concepts, and so their cited remarks are appropriate to the task to which I am putting them.

[53] Provided of course that we do not then go on to attempt to say that sentimentalism about *F*-ness is the thesis that *F*-ness is response dependent in our more expansive sense of "response dependent." That would be bad, as it would imply that one can be a sentimentalist about properties that have nothing to do with sentiments, e.g., punchability or bingeworthiness.

of the properties designated by English predicates like “blameworthy,” “desirable,” “awe-inspiring,” etc. And this certainly sounds rather less anodyne; after all, surely the RI/RD debate over blameworthiness, say, could not be won simply by observing that “blameworthiness,” as a matter of good English, denotes worthiness of blame. Something seems to have gone wrong.

I answer that a number of things have gone wrong: First, as I speculated in section 2, I take it that a number of philosophers of blameworthiness who seem to make RI-theoretic remarks are simply speaking carelessly or loosely, à la Scanlon. Second, as I argued in section 3, I take it that a number of philosophers of blameworthiness have erroneously supposed that a popular approach to theorizing blameworthiness requires you to collapse blameworthiness into its response-independent ground. Finally, as I will shortly illustrate, I take it that a number of philosophers of blameworthiness have misjudged the implications that do and do not follow from the affirmation of a faithful analysis of blameworthiness. In this vein, I hypothesize that rather than reserve the labels “RD theory of *F*-ness” and “RI theory of *F*-ness” for theories that affirm *F*-ness’s response dependence or response independence respectively, such philosophers overextend these labels to cover the theories that result from conjoining each respective affirmation with the implications they take to follow from it. It is no surprise, then, that the RI/RD debate should appear insusceptible of trivial resolution by appeal to word structure, since quite a number of the major theses at issue in that debate are *not* susceptible of such resolution. That such theses are not thus susceptible is a testament to the fact that they do not follow from what *is* trivial, namely, as I say, that we ought to endorse only faithful analyses of putative relation-response properties.

To make good on these contentious claims, let us return to our main task and consider where self-professed RD theorists stand vis-à-vis the Common Grounding Structure.

4.3. Response-Dependence Theories and the Common Grounding Structure

Shoemaker is a self-proclaimed RD theorist about a certain form of blameworthiness that we may call “*angry-blameworthiness*.”[54] In particular, he endorses a “fitting” or “normative” RD theory of angry-blameworthiness, according to which that property *just is* the property of being a fitting target of a certain form of anger. Thus, this form of angry-blameworthiness, for Shoemaker, clearly has faithful relation-response structure. Moreover, Shoemaker holds that angry-blameworthiness, so understood, is always grounded in what he refers to as “objective features,” such as, e.g., “control, knowledge, voluntariness, quality of will, or

[54] Shoemaker, “Response-Dependent Responsibility” and “Response-Dependent Theories of Responsibility.”

history."[55] In other words, it is always some combination of response-independent features that *make* persons who have them fitting targets of angry-blame. Finally, Shoemaker says that the "fundamental fitting response-dependent feature of [this] theory is really about what makes certain objective features [like, e.g., those just listed] the *anger fitmakers* in the first place," which, for him, is that such features "trigger our [refined] anger sensibilities."[56]

Shoemaker's RD theory of his target form of angry-blameworthiness thus answers each of the angry-blameworthiness instances of QS-I–III. If we abstract out a bit, we are left with the grounding structure in figure 5:

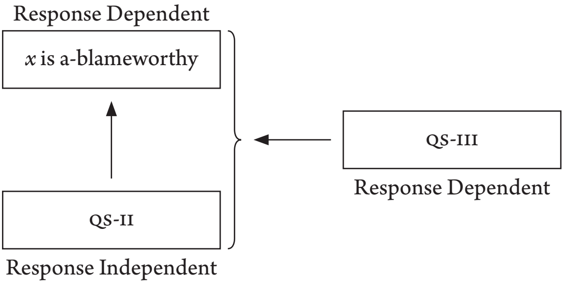


FIGURE 5

This graphic says: (i) facts of the form [*x* is angry-blameworthy] are response dependent (i.e., they are facts involving the instantiation of a response-dependent property); (ii) each such fact is grounded in facts that are response independent (i.e., facts involving the instantiation of a response-independent property); and (iii) the grounding of each fact of the form [*x* is angry-blameworthy] in some such response-independent facts is itself at least partly grounded in some fact or facts concerning a relation (or relations) that the grounds of angry-blameworthiness stand in to the type of response at issue in angry-blameworthiness—a type of response that Shoemaker sometimes calls "*angry-blame*."

It should be clear that Shoemaker's fitting-RD theory of angry-blameworthiness construes it as a genuine relation-response property and situates facts involving the instantiation of that property in a series of grounding relations that together instantiate the Common Grounding Structure.

D'Arms and Jacobson's fitting-RD theories of various putative relation-response properties do the same.[57] To keep things simple, let us focus on their RD theory of *self-blameworthiness*, which says that to be self-blameworthy just is

[55] Shoemaker, "Response-Dependent Responsibility," 509.

[56] Shoemaker, "Response-Dependent Responsibility," 509–11 (bracketed words added).

[57] D'Arms and Jacobson, "Whither Sentimentalism?" and "The Motivational Theory of Guilt (and Its Implications for Responsibility)."

to be a fitting target of guilt. D'Arms and Jacobson's story centers on their own special conception of fittingness. They hold that the emotion involved in self-blame is guilt and that this emotion, like other "natural emotions," is susceptible of an "interpretation" according to which it "appraises" its target as—to use their self-professedly "rough" answer as an example—"having engaged either in some sort of *wrongdoing* or in a *personal betrayal*."[58] Crucially, D'Arms and Jacobson depart from Shoemaker insofar as they warn against reading the properties that figure in such appraisals as being response-independent properties: "Since these emotional appraisals are derived from the emotion holistically, including its motivational element, they must be understood as response dependent—even if their terms have response-independent senses in ordinary language."[59]

D'Arms and Jacobson then propose to understand the fittingness of natural emotions as the correctness of such appraisals. Thus, x is a fitting target of guilt when x is such as to render correct guilt's distinctive appraisal, as yielded by some interpretation. On this picture, then, the properties of having acted wrongly and having engaged in personal betrayal—where, recall, these properties are being conceived as covertly response dependent—are grounds of being self-blameworthy not because they "trigger our refined

[58] D'Arms and Jacobson, "The Motivational Theory of Guilt (and Its Implications for Responsibility)," 18, 23. For an admirably condensed sketch of the details of how D'Arms and Jacobson take the appraisals at issue in fittingness to work, see the following:

> Begin with an *empirical* characterization of the general emotional syndrome: the cluster of feelings, patterns of attention, typical elicitors and palliators, characteristic thoughts, and especially the motivational role occurring in paradigmatic episodes of the emotion kind. In light of this data, give an *interpretation* into language of how someone in the grip of such an emotion appraises its object as specifically good or bad. Appraisals in this sense are not constitutive thoughts or components of emotion, but ways of understanding how the emotion as a whole evaluates its object. Any gloss into language will be imperfect and can at most help to point in the direction of the distinctive way that the emotion appraises its object. *Since these emotional appraisals are derived from the emotion holistically, including its motivational element, they must be understood as response dependent—even if their terms have response-independent senses in ordinary language*. . . .
>
> An empirical characterization of fear favors the suggestion that it should be interpreted as appraising its object as dangerous, for example; this makes sense of how fear engages with its object—as something to be avoided directly and urgently. . . . *What is distinctive about our approach is how it understands the claim that fear is about danger: not as a response-independent thought one must have in order to count as afraid, but rather as an effort to articulate the distinctive emotional appraisal involved in the combination of feelings, goals, and action tendencies of fear*. (18–19, emphasis at the end of each paragraph added)

[59] D'Arms and Jacobson, "The Motivational Theory of Guilt (and Its Implications for Responsibility)," 18.

guilt sensibilities," to use Shoemaker's phrase, but rather because: (i) it lies in the nature of guilt that it is interpretable as appraising its targets either as having acted wrongly or as having engaged in personal betrayal; and (ii) what it is for guilt to be fitting is for its interpreted appraisal of its target to be accurate. In other words, D'Arms and Jacobson may be understood as providing what I earlier (in section 1.3) called an "essentialist" answer to the self-blameworthiness-instance of QS-III.

Thus, if we abstract out a bit, D'Arms and Jacobson's fitting-RD theory of self-blameworthiness situates self-blameworthiness facts in grounding structures of the sort in figure 6:

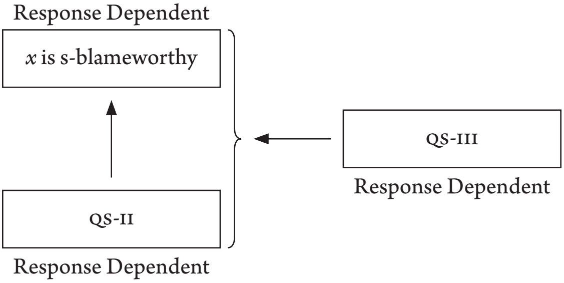


FIGURE 6

This structure differs from that posited by Shoemaker's fitting-RD theory insofar as it embeds a different constraint on permissible occupants of the QS-II position, namely, that they be response dependent. On the other hand, while D'Arms and Jacobson do not adopt Shoemaker's style of answer to the self-blameworthiness-instance of QS-III, they agree with Shoemaker that the answer must refer to some fact or facts about relations borne by the occupants of QS-II to the type of relation at issue in the relevant form of blameworthiness. Differences with Shoemaker aside, it should be clear that D'Arms and Jacobson's fitting-RD theory of self-blameworthiness also construes that property as a genuine relation-response property and situates facts involving the instantiation of that property in a series of grounding relations that together instantiate the Common Grounding Structure.

4.4. Response-Independence Theories and the Common Grounding Structure

What about RI theories of putative relation-response properties, like blameworthiness? How do such theories construe blameworthiness, and where do they situate it in relation to other facts and grounds? To answer this, consider once more what Coleman and Sarch say about blameworthiness, namely, that it is "a *reason* or a *ground* that explains why blaming ... would be justified," which they take to be or to be reducible to some property in the ballpark of culpability for wrongdoing (which, recall, Coleman and Sarch take to be response independent).[60] In other words, blameworthiness is a response-independent ground of a property, like being a justified target of blame. Thus Coleman and Sarch's theory of blameworthiness implies that its instantiations occupy grounding structures of the following sort, where "JTB" abbreviates "justified target of blame" (figure 7):

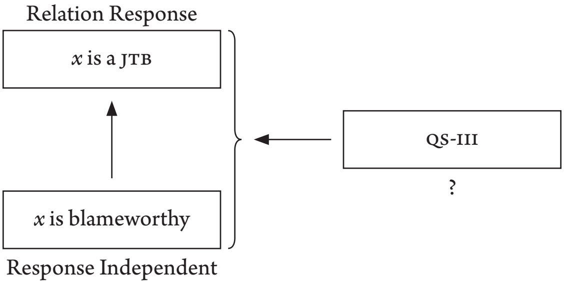


FIGURE 7

This sort of picture largely accords with the schematic definition of RI theories of angry-blameworthiness that Shoemaker offers, namely:

> *Response-Independence about the Blameworthy*: The blameworthy consists in a property (or properties) of agents that makes anger at them appropriate, a property (or properties) whose value-making is ultimately independent of our angry responses. Anger at someone for *X* is appropriate if and only if, *and in virtue of the fact that*, she is antecedently blameworthy (and so accountable) for *X*. What makes her blameworthy is thus ultimately response-independent.[61]

This schema is rife with commitments that can be helpfully captured by surveying what it has to say about the angry-blameworthiness-instances of QS-I–III. Starting with QS-I, Shoemaker's schema tells us that the RI theorist of angry-blameworthiness is bound by definition to saying that what it is for *x* to be angry-blameworthy is for *x* to be *F*, where *x*'s being *F* grounds *x*'s being an appropriate target of anger. This of course is not an answer to the angry-blameworthiness-instance of QS-I but rather a constraint upon possible answers to it.[62] Shoemaker is clear, however, about what sorts of answers he regards as

[60] Coleman and Sarch, "Blameworthiness and Time," 101, 103.

[61] Shoemaker, "Response-Dependent Responsibility," 498.

[62] Shoemaker does, however, appear to imply that for the RI theorist, at least part of what it is to be blameworthy for something is to be accountable for that thing.

typically offered here, saying that "the response-independent theorist says that the response-independent property of the [angry-]blameworthy (that it was a bad action performed with voluntariness, control, knowledge, and so on) is what makes anger appropriate"[63] In other words, angry-blameworthiness is an "objective"—i.e., response-independent—property.

Recall that the angry-blameworthiness-instance of QS-II asks: "What are the conditions under which something is angry-blameworthy?" Shoemaker's schema does not tell us that the RI theorist of angry-blameworthiness is bound by definition to say anything special here. Instead, it tells us that such theorists are bound by definition to give a specific answer to a different instance of QS-II, namely, "What are the conditions under which something is anger-worthy?" The RI theorist of angry-blameworthiness must say that angry-blameworthiness is an apparently universal ground of anger-worthiness (which, for Shoemaker, is identical to being a fitting target of anger).

Finally, the angry-blameworthiness-instance of QS-III asks, "Why are the conditions of angry-blameworthiness as they are?" It is a bit ambiguous what Shoemaker's schema requires the RI theorist of angry-blameworthiness to say here. Initially, we are not offered any information on the RI theorist's response to this question; instead, we are offered information on the RI theorist's response to a different instance of QS-III, namely, "Why are the conditions of *anger-worthiness* (i.e., being a fitting target of anger) as they are?" The RI theorist of angry-blameworthiness, we were told, identifies angry-blameworthiness with the condition (or conditions) of anger-worthiness, and now we are told that the RI theorists also say that angry-blameworthiness's status as an anger-worthy-making property is not explicable by reference to its relation to our angry responses. In other words, facts of the form [[*x* is angry-blameworthy] grounds [*x* is anger-worthy]] are never even partly grounded in a fact of the form [Angry-blameworthiness bears *R* to our angry responses] for any relation, *R*. But then, slightly thereafter, Shoemaker concludes the RI schema by saying that "what makes [an agent] [angry-]blameworthy is thus ultimately response-independent." On the basis of *this* remark, it seems Shoemaker *does* regard the RI theorist of angry-blameworthiness as committed to a constraint on possible answers to the angry-blameworthiness-instance of QS-III, namely, that the answer not appeal to angry-blameworthiness's bearing some relation to our responses. So RI theories of angry-blameworthiness—according to Shoemaker—appear to place the same "response-independent answers only" constraint on two distinct instances of QS-III: one for angry-blameworthiness and one for what angry-blameworthiness grounds, namely, anger-worthiness (i.e., being a fitting target of anger).

[63] Shoemaker, "Response-Dependent Responsibility," 509 (bracketed text added).

If the foregoing remarks prove difficult to track, do not worry. The important takeaway is that, by Shoemaker's lights, the RI theorist of angry-blameworthiness is committed to its instantiations standing in grounding structures of the sort in figure 8:

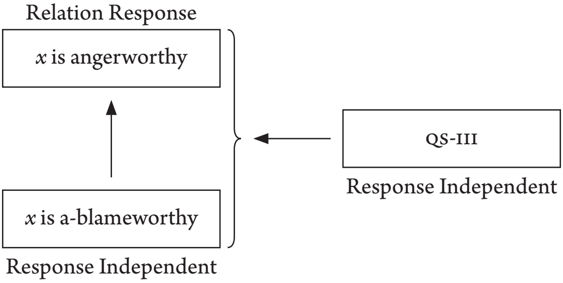


FIGURE 8

This grounding structure is nearly identical to the one we attributed to Coleman and Sarch's RI theory of blameworthiness. The only difference is that an extra constraint has been placed on possible occupants of the QS-III position, namely, that they be response independent. Still, like Coleman and Sarch's theory of blameworthiness, as well as Shoemaker's and D'Arms and Jacobson's fitting-RD theories of their respective forms of blameworthiness, the RI theorist of angry-blameworthiness, as Shoemaker conceives of them, situates facts involving the instantiation of angry-blameworthiness in a series of grounding relations that together instantiate the Common Grounding Structure.

4.5. *Problems Observed*

Hopefully you see what I see: Coleman and Sarch and the rest of the RI theorists of blameworthiness as Shoemaker conceives of them do *not* disagree with RD theorists of blameworthiness that certain relation-response properties with a type of blame as the response constituent—e.g., being a justified target of blame or being a fitting target of anger—are grounded in further properties. They do not even disagree with Shoemaker, a prominent RD theorist of blameworthiness, over roughly what sorts of properties do the grounding, here, namely, "objective" or response-independent properties like having acted wrongly from ill will or suchlike. In this area, the only disagreement between these camps is with respect to how we name the nodes in the grounding structure: RI theorists use relation-response expressions like "blameworthiness" to name objective, response-independent grounds, whereas RD theorists use it to name faithful relation-response properties like being a justified target of blame or being a fitting target of anger. This dispute is therefore merely verbal: it has to do not with worldly facts and their relations but rather with the question of whether or not to respect word structure.

This is not the only difference we have brought out. If we go with Shoemaker, it seems RI theorists also characteristically disagree with RD theorists over possible answers to relevant instances of QS-III: the RI theorist only accepts response-independent answers, whereas the RD theorist demands response-dependent answers.

The first thing to say here is that Shoemaker's claim that there are many RI theorists thus construed seems rather unlikely. That is because it is hard to imagine how one might attempt to ground facts of the form [[*x* is angry-blameworthy] grounds [*x* is a fitting target of anger]] or suchlike without appealing to *any* facts involving relations between angry-blameworthiness and anger. Of course one could say that such grounding facts are ungrounded; that would satisfy the constraint under consideration. But I suspect few would do such a thing. As we noted in section 1, it is much more common to locate the grounds of grounding facts partly in facts about the essences of one or more of the constituents involved in them. (This, for instance, is what Rosen does in answering the blameworthiness instance of the QS-III schema.[64]) Alternatively, one might take the relevant sort of fittingness as a primitive, nonnaturalistic normative relation and endeavor to ground the grounding of fittingness facts partly by appeal to normative bridge-laws. We need not explore this option further except to say that any such approach to grounding our grounding facts would certainly appeal to a relation borne by blameworthiness to anger: relating these items is just what such a bridge-law would be posited to do. Thus the most common approaches to answering the relevant instance(s) of QS-III violate the constraint Shoemaker takes to be constitutive of RI theories. In the absence of alternative approaches that satisfy the response-independence constraint, then, it is hard to imagine who exactly holds the view that Shoemaker thinks is "much more popular."

But—and this is the second thing—even if we grant that there are theorists who eschew response-dependent answers in cases of this sort, why would we ever promote this questionable eschewing to the status of a *defining feature* of being an RI theorist of blameworthiness? To see the problem, consider Coleman and Sarch. They mean to identify blameworthiness with or reduce it to some response-independent property in the ballpark of culpability for wrongdoing. Surely *this* should be the point at which we say that Coleman and Sarch are RI theorists of blameworthiness! But Shoemaker is committed to disagreeing: should Coleman and Sarch happen to proceed to give a response-dependent

[64] Rosen, "The Alethic Conception of Moral Responsibility," 73–74.

answer to the justified-target-of-blame-instance of QS-III—a perfectly natural thing to do, given the popularity of essentialist answers to such questions—this, according to Shoemaker's schema, would somehow render Coleman and Sarch undeserving of the label of "RI theorists of blameworthiness." This, I take it, is a patently absurd way of carving up the conceptual space: when we say that somebody endorses an RI theory of *F*-ness, surely this should mean only that they are committed to the response independence *of F-ness*.

To sum up, it seems that, as a number of theorists define things, the difference between RI and RD approaches to theorizing certain relation-response properties boils down partly to a mere verbal disagreement and partly to a disagreement over how to answer certain instances of QS-III. The merely verbal disagreement is easily won by the RD theorists, since they respect word structure and their opponents do not. The nonverbal disagreement, on the other hand, seems neither here nor there with respect to the joints that seem most apt to be carved by the labels "RI theory of *F*-ness" and "RD theory of *F*-ness." We would do better to reserve these labels precisely for theories of *F*-ness that affirm its response independence or dependence respectively. If we do, we find that the RI/RD disputes over putative relation-response properties entirely reduces to the question of whether to respect word structure. That question, I have claimed, is easily answered.

5. Closing Thoughts: On the Lightness of Faithfulness

Still, you may have lingering doubts. You may worry in particular that to embrace a faithful analysis of a property like blameworthiness is to do something much bolder than I have been suggesting. After all, look at all of the mileage Shoemaker and D'Arms and Jacobson seem to get out of the claim that blameworthiness, say, is response dependent. Recall that for Shoemaker, the "fundamental fitting response-dependent feature of [the normative or fitting-RD theory of angry-blameworthiness] is really about what makes certain objective features the *anger fitmakers* in the first place," namely, that such features "trigger our [refined] anger sensibilities."[65] And recall that D'Arms and Jacobson's special brand of response dependence about self-blameworthiness implies the startling claim that the grounds of self-blameworthiness must themselves be covertly response dependent. These are controversial claims. Must the RD theorist, *qua* RD theorist, accept any of them?

No. To be an RD theorist of *F*-ness, I have argued, ought just to be to affirm the response dependence of *F*-ness. In other words, it ought just to be to give

[65] Shoemaker, "Response-Dependent Responsibility," 509–11.

a response-dependent answer to the *F*-ness instance of QS-I. What you then go on to say about the *F*-ness instances of QS-II and QS-III is your own business. Shoemaker's and D'Arms and Jacobson's respective answers to the blameworthiness instances of QS-III that they each consider are particular to their respective conceptions of fittingness, as is D'Arms and Jacobson's requirement that answers to the blameworthiness-instance of QS-II be covertly response dependent. RD theorists *as such* are not required to conceive of fittingness in these ways and thus are not required to answer questions of these sorts in these ways. In fact, in light of our results from section 3, RD theorists of *F*-ness may choose (in relevant cases) to avail themselves of an alethic conception of fittingness, paired with cognitivism about the type of response involved in *F*-ness, to yield a theory of *F*-ness that offers different answers to the *F*-ness instances of QS-I–III than those offered by D'Arms and Jacobson and Shoemaker, as Rosen does. Or else RD theorists may go in for an entirely different conception of fittingness, yielding entirely different answers to these questions. The point is that there is room to maneuver here, as faithfulness leaves much unsettled. That is a virtue, not a vice: it is part of what makes faithfulness a good starting point for theorizing about relation-response properties.[66]

University of Notre Dame
ssmith62@nd.edu

REFERENCES

Audi, Paul. "Grounding: Toward a Theory of the 'In-Virtue-of' Relation." *Journal of Philosophy* 109, no. 12 (December 2012): 685–711.

[66] This paper is effectively the product of two graduate seminars: Robert Audi's seminar on moral rationalism at University of Notre Dame in the spring of 2022 and Randolph Clarke's seminar on blame at Florida State University in the spring of 2021. I am grateful to Robert and Randy for their helpful comments on and discussion of this paper's earlier drafts. Special thanks are also due to this paper's truly excellent referees, whose insightful and discerning comments—delivered with great kindness and enthusiastic support—inspired a number of extremely helpful amendments. Special thinks are also due to Kevin Jia, to whom I am immensely grateful both for numerous helpful discussions of key parts of this paper and for his interest in and dedication to carefully reading and critically commenting on the paper during Robert's seminar (despite the fact that the paper was several thousand words over the assigned word limit), as well as on subsequently edited bits. Finally, I am grateful to many others either for their comments on the paper itself or for helpful discussion of this or that idea expressed in it, including Marshall Bierson, Joseph Blado, Isabel Canfield, Stephen Kearns, William Kombian, Katelyn O'Dell, Gideon Rosen, David Shoemaker, Tucker Sigourney, Mack Sullivan, Johnny Waldrop, and Andreas Waldstein.

Austin, J. L. "A Plea for Excuses: The Presidential Address." *Proceedings of the Aristotelian Society* 57, no. 1 (June 1956): 1–30.

Clarke, Randolph, and Piers Rawling. "True Blame." *Australasian Journal of Philosophy* 101, no. 3 (2022): 1–14.

Coleman, Jules, and Alexander Sarch. "Blameworthiness and Time." *Legal Theory* 18, no. 2 (June 2012): 101–37.

Correia, Fabrice. "Real Definitions." *Philosophical Issues* 27, no. 1 (October 2017): 52–73.

D'Arms, Justin, and Daniel Jacobson. "The Motivational Theory of Guilt (and Its Implications for Responsibility)." In *Self-Blame and Moral Responsibility*, edited by Andreas Brekke Carlsson, 11–27. Cambridge: Cambridge University Press, 2022.

———. "Whither Sentimentalism? On Fear, the Fearsome, and the Dangerous." In *Ethical Sentimentalism*, edited by Remy Debes and Karsten R. Stueber, 230–49. Cambridge: Cambridge University Press, 2017.

Dasgupta, Shamik. "Metaphysical Rationalism." *Noûs* 50, no. 2 (June 2016): 379–418.

Dorr, Cian. "To Be *F* Is to Be *G*." *Philosophical Perspectives* 30, no. 1 (December 2016): 39–134.

Fine, Kit. "Guide to Ground." In *Metaphysical Grounding*, edited by Fabrice Correia and Benjamin Schnieder, 37–80. Cambridge: Cambridge University Press, 2012.

Foot, Philippa. "Hume on Moral Judgment." In *Virtues and Vices and Other Essays in Moral Philosophy*, 74–80. Oxford: Oxford University Press, 2002.

McKenna, Michael. "Directed Blame and Conversation." In *Blame: Its Nature and Norms*, edited by Justin D. Coates and Neal A. Tognazzini, 120–40. Oxford: Oxford University Press, 2012.

Quine, W. V. O. "On What There Is." *Review of Metaphysics* 2, no. 5 (1948): 21–38.

Rosen, Gideon. "The Alethic Conception of Moral Responsibility." In *The Nature of Moral Responsibility: New Essays*, edited by Randolph Clarke, Michael McKenna, and Angela M. Smith, 65–88. Oxford: Oxford University Press, 2015.

———. "Ground by Law." *Philosophical Issues* 27, no. 1 (October 2017): 279–301.

———. "Metaphysical Dependence: Grounding and Reduction." In *Modality: Metaphysics, Logic, and Epistemology*, edited by Bob Hale and Aviv Hoffman, 109–36. Oxford: Oxford University Press, 2010.

———. "Real Definition." *Analytic Philosophy* 56, no. 3 (September 2015): 189–209.

Scanlon, T. M. *Moral Dimensions: Permissibility, Meaning, Blame.* Cambridge, MA: Harvard University Press, 2008.

Shoemaker, David. "Response-Dependent Responsibility; or, A Funny Thing Happened on the Way to Blame." *Philosophical Review* 126, no. 4 (October 2017): 481–527.

———. "Response-Dependent Theories of Responsibility." In *The Oxford Handbook of Moral Responsibility*, edited by Dana Nelkin and Derk Pereboom, 304–24. Oxford: Oxford University Press, 2022.

Journal of Ethics and Social Philosophy
vol. 28, no. 3 · october 2024

https://doi.org/10.26556/jesp.v28i3.3286

RATIONALITY AND RESPONDING TO NORMATIVE REASONS

Mohamad Hadi Safaei

The reasons-responsiveness theory of rationality holds that rationality is a matter of responding correctly to one's normative reasons.[1] According to this theory, you are rationally criticizable (that is, you are not fully rational) when one of your attitudes is not a correct response to your normative reasons.[2] Moreover, it is metaphysically impossible that you respond correctly to your normative reasons while still being rationally criticizable. There is nothing more to rationality than responding to normative reasons.[3]

Associating rationality with normative reasons, reasons-responsiveness theory promises to vindicate an interesting claim about the normative significance of rationality. Since the balance of normative reasons determines both what you ought to do and what you are rationally required to do, then what

[1] Among the proponents of a reasons-responsiveness account of rationality are Williams, "Internal and External Reasons"; Parfit, "Reasons and Motivation"; Schroeder, "Means-End Coherence, Stringency and Subjective Reasons"; Sylvan, "What Apparent Reasons Appear to Be"; Kiesewetter, *The Normativity of Rationality*; and Lord, "The Coherent and the Rational" and *The Importance of Being Rational*.

[2] A normative reason is commonly understood to be a consideration that counts in favor of a response. See Scanlon, *What We Owe to Each Other*. It is also widely assumed that a consideration that is a normative reason for an agent to respond in a certain way is a fact or true proposition. See Dancy, *Practical Reality*; Parfit, *On What Matters*; Raz, *From Normativity to Responsibility*; and Broome, "Reasons."

[3] In contrast, structuralism about rationality holds that rationality is fundamentally a matter of satisfying structural requirements of rationality. See Broome, *Rationality through Reasoning*. These requirements include the consistency requirements on beliefs and intentions, the instrumental requirement to intend to do what one believes that is a necessary means for her intended ends, and the enkratic requirement, which requires you to act in accordance with your all-things-considered normative judgment about what you ought to do. Some authors have argued that there are two distinct phenomena under the name "rationality." See Worsnip, "The Conflict of Evidence and Coherence" and *Fitting Things Together*; Fogal, "Rational Requirements and the Primacy of Pressure"; and Fogal and Worsnip, "Which Reasons?" Structural rationality is a matter of satisfying structural requirements of coherence, and substantive rationality is a matter of responding to normative reasons. In this paper, I take issue with a unificationist reasons-responsiveness theory that holds that rationality simpliciter is only a matter of responding to reasons.

you are rationally required to do would always be the same thing as what you ought to do. By a similar reasoning, we can show that rational permissibility and normative permissibility are one and the same thing. This thesis is called the *Strong Normativity of Rationality*.[4]

Although it is an elegant theory, in recent decades, many philosophers have argued against the reasons-responsiveness account by providing counterexamples that purport to show that this theory cannot explain the irrationality of incoherent attitudes.[5] Typically, the focus has been cases in which one is rationally criticizable for being akratically or instrumentally incoherent. Below is an example of akratic irrationality.[6]

> *Akratic Irrationality*: The stuff in the glass in front of Bernard looks and smells like gin and tonic and has been served to Bernard in response to his request for a gin and tonic. Bernard is thirsty and badly wants to drink a gin and tonic. Deliberating on all these, he rationally comes to believe that he has decisive reason to drink what is in the glass. Given the intuitive rationality of Bernard's normative judgment, reasons-responsiveness theory would predict that Bernard's belief is a proper response to his normative reasons. Unfortunately, unbeknownst to Bernard, the glass contains petrol, and arguably, this fact provides him with a strong reason against drinking the contents of the glass. After all, Bernard wants to drink a gin and tonic, and he has every reason to avoid drinking petrol.

[4] On the other hand, the Weak Normativity of Rationality states that whenever you are rationally required to ϕ, you have some reason to ϕ, but this reason is not always overriding, and there are possible situations where what you are rationally required to do and what you ought to do might come apart. Similarly, according to the Weak Normativity of Rationality, what you are rationally permitted to do might diverge from what you are normatively permitted to do. According to the Special Normativity of Rationality, whenever you are rationally required to ϕ, then this fact about rationality is—or gives you—a normative reason to ϕ. The latter idea is what Broome calls the thesis of the Normativity of Rationality. See Broome, *Rationality through Reasoning*, 192. See also Worsnip, "Making Space for the Normativity of Coherence." Proponents of reasons-responsiveness theory typically reject the Broomean idea of the special normativity of rationality because according to the reasons-responsiveness view, rationality is not an independent source of normative requirements or normative reasons, but its requirements are ultimately grounded on normative reasons of different kinds like prudence, morality, evidence, etc.

[5] Broome, *Rationality through Reasoning* and *Normativity, Rationality and Reasoning*; Worsnip, "The Conflict of Evidence and Coherence" and "Reasons, Rationality, Reasoning"; Fogal, "Rational Requirements and the Primacy of Pressure"; Fogal and Worsnip, "Which Reasons?"; Brunero, *Instrumental Rationality*; and Lee, "The Independence of Coherence."

[6] For similar examples, see Fogal and Worsnip, "Which Reasons?"; and Broome, *Rationality through Reasoning*, 104–5.

> So, at least *prima facie*, it would be plausible to say that Bernard has decisive reason to refrain from drinking from his glass.

If Bernard were to respond correctly to all his normative reasons, he would end up at irrationality. By responding to his decisive normative reason, he would refrain from doing what he himself rationally believes he ought to do. But presumably, it is a necessary condition for being fully rational that one at least intends to do what one rationally believes one has decisive reason to do.[7] The example allegedly shows that not every case of irrationality is explainable in terms of one's failure in responding correctly to the balance of one's normative reasons.[8]

In response, the standard maneuver for reasons-responsiveness theorists is to appeal to a highly plausible distinction, which goes back to Aristotle and lies at the center of Kant's account of the moral worth of actions, between *acting in accordance with* normative reasons and *responding to* normative reasons.[9] The rough idea is that an agent's action is a response to her normative reasons just in case she performs the action *in virtue of* the fact that she has those normative reasons. That is, there must be an explanatory connection between the agent's action and the normative fact that her reasons support performing that action. In the above example, Bernard is rationally criticizable for acting against his rational normative judgment if his action is not a genuine response to his normative reasons. After all, Bernard does not know that the glass contains petrol. If he is not aware of the latter fact, there cannot be an explanatory connection between his action (refraining from drinking the glass) and the normative significance of the fact that the glass contains petrol. Thus, Bernard fails to respond to his reasons, and that explains why he is irrational.

[7] Notice that this example does not straightforwardly indicate that rationality has structural requirements. Of course, one possible explanation is to appeal to the enkratic requirement of rationality that requires you to intend to do what you believe you ought to do. But this is not the only explanation, and one might try to provide an alternative explanation of the same phenomena without appealing to the "requirements" of rationality. See Fogal, "On the Scope, Jurisdiction, and Application of Rationality and the Law."

[8] The idea is that, according to the reasons-responsiveness view, whenever you are irrational for having an incoherent combination of mental attitudes, at least one of those attitudes is not a correct response to your reasons. In other words, in cases of irrational incoherence, there is nothing especially wrong about the combination. Thus, in explaining cases of incoherence, the reasons-responsiveness view either argues that one of the attitudes is normatively deficient or tries to explain the apparent irrationality away. This latter strategy is the standard maneuver for the preface and lottery cases. See Lord, *The Importance of Being Rational*, 51–55; and Kiesewetter, *The Normativity of Rationality*, 254, 257. In what follows, I will discuss and argue the insufficiency of both strategies.

[9] Lord, *The Importance of Being Rational*, 217–22.

Generally, the reasons that can contribute to determining what an agent is rationally required to do (and what she ought to do) consist of the ones that the agent is able to respond to in this demanding sense of responding to reasons, which requires the existence of an explanatory relation between her performance and the normative fact that her reasons support that performance. In a technical term, only the reasons that she *possesses* can determine what is rational for her to do. Again, since Bernard does not possess the fact that the glass contains petrol, this fact cannot make it rational for him to refrain from drinking from the glass.

My aim in this paper is to show that even this perspectivist maneuver fails to explain the intuitive irrationality of practical akrasia.[10] To explain the irrationality of practical akrasia (that is, the irrational mismatch between believing that you ought to ϕ and lacking an intention to ϕ), proponents of the reasons-responsiveness view argue that whenever such a mismatch is irrational, it is either because you possess decisive reason to give up your belief that you ought to ϕ or because you possess decisive reason to intend to ϕ. My central argument against the reasons-responsiveness account is to show that there can be situations in which your possessed reasons permit you to believe that you ought to ϕ while simultaneously you possess sufficient reasons not to intend ϕ. In such situations, the reasons-responsiveness view allows for an irrational mismatch between believing that you ought to ϕ and lacking an intention to ϕ. The main premise of my argument is that the possession of a normative reason for action does not guarantee that there is available evidence for being subject to that reason. Consequently, one can possess a reason to act without being in a position to know that one possesses that reason.[11]

The plan of the paper is as follows. First, in section 1, I argue that the best explanation of the distinction between acting in accordance with a normative reason and responding to that reason involves appealing to one's competence or knowledge about how to respond to that reason. But as I explain in section 2, someone might possess a practical competence to respond to her decisive practical reasons to perform an action without having the parallel theoretical

[10] Here I am concerned with practical akrasia. There is a similar debate in the epistemology literature about the irrationality of epistemic akrasia and whether evidentialism, as a version of reasons-responsiveness view, can explain that epistemic irrationality. See Worsnip, "The Conflict of Evidence and Coherence"; Lasonen-Aarnio, "Enkrasia or Evidentialism?"; Titelbaum, "Rationality's Fixed Point"; Littlejohn, "Stop Making Sense?"; Horowitz, "Epistemic Akrasia"; and Greco, "A Puzzle about Epistemic Akrasia."

[11] I am grateful to an anonymous reviewer for this journal for helping me to better articulate and present my case against the reasons-responsiveness view. Here, I borrowed phrasing from the reviewer's comments.

competence to rationally conclude her deliberation by believing that she has decisive reason to perform that action. If possessing a normative reason is a matter of having the ability to respond to that reason, and responding to a reason is grounded in facts about manifesting one's competence about how to respond to that reason, then the mismatch between one's practical and theoretical competences can give rise to a normative mismatch between one's possessed reasons for action and one's possessed reasons about what to believe about one's reasons for action. In the final section, I conclude by considering whether it could be rationally permissible to act against one's own normative judgment about what one ought to do if one's reasons for action diverges from one's reasons for one's judgment about what one ought to do.

1. Responding to Normative Reasons

As I mentioned, my argument against the reasons-responsiveness theory of rationality relies on a competence-based account of the nature of responding to normative reasons. This section provides a very brief and concise defense of that competence-based account.

A simple account of responding to normative reasons has it that one's ϕ-ing is a response to a normative reason R if and only if (i) R is, as a matter of fact, a normative reason to ϕ, and (ii) R appropriately motivates one to ϕ.[12] In other words, one counts as responding to a normative reason just in case one's motivating reason for acting coincides with the reason normatively justifying the action.[13] Recently, some philosophers have argued that this conception of responding to normative reasons is problematic.[14] For one thing, let us consider Kant's famous shopkeeper, who, out of mere concern for building his own business, makes sure that he always charges fair prices. As it happens, his

[12] Markovits, "Acting for the Right Reasons," 205. Spelling out the modifier "appropriately" here requires finding a way to filter cases of deviant causal chains in which a consideration that is a reason for you to ϕ somehow causes you to φ, but it does not motivate you to ϕ in the relevant way. As Davidson puts it, "not just any causal connection between rationalizing attitudes and a wanted effect suffices to guarantee that producing the wanted effect was intentional. The causal chain must follow the right sort of route." See Davidson, "Freedom to Act," 78. For suggestions on how to handle this problem, see Turri, "Believing for a Reason"; and McCain, "The Interventionist Account of Causation and the Basing Relation."

[13] For a highly illuminating discussion of the common distinction between normative reasons and motivating reasons, see Alvarez, "Reasons for Action."

[14] Sliwa, "Moral Worth and Moral Knowledge"; Lord, *The Importance of Being Rational*, ch. 5; Isserow, "Moral Worth and Doing the Right Thing by Accident"; Johnson King, "Accidentally Doing the Right Thing"; Cunningham, "Is Believing for a Normative Reason a Composite Condition?" and "Moral Worth and Knowing How to Respond to Reasons."

actions are always motivated by the very considerations that make them right. Since he wants to earn a reputation as a morally good retailer, he is always particularly careful to consider and act for morally significant features. However, it is obvious that in acting in accordance with moral requirements, the shopkeeper is not genuinely responding to his normative reasons even though his right-doings are motivated by the same considerations that make his actions right (for example, facts about the fairness of a charge).

In response, proponents of the simple account might seem to have at least two options at their disposal. First, they could argue that although the fact about fairness is indeed *part* of the shopkeeper's motivating reason for action, his motivating reason also includes other considerations, the most important of which is that acting fairly is good for business. In that case, there would not be a complete coincidence between his motivating reasons and the normative reasons that make his action right. But what makes for responding to normative reasons is a one-to-one correspondence between one's normative and motivating reasons. For another option, proponents might be inclined to appeal to the instrumental/noninstrumental distinction between one's motivating reasons. The idea is that facts about fairness motivate the shopkeeper's actions only to the extent that fairness promotes financial gain. That is, the shopkeeper's noninstrumental motivating reasons merely include facts about what is pivotal for his business. But then one can argue that responding to normative reasons requires coincidence between one's normative reasons and one's noninstrumental motivating reasons.

However, even such maneuvers cannot save the day for the simple account. To illustrate, let us take a look into an example proposed by Paulina Sliwa.[15] Jean's friend has an important meeting, but she missed the bus that she normally takes to work; arriving late would be a major embarrassment. Out of a noninstrumental desire to spare her friend a major embarrassment, Jean gives her a ride. Other things being equal, Jean's action is the one she has decisive reason to do. The question is whether Jean's action is a response to her decisive normative reason to help her friend. Sliwa argues that even though sparing one's friend a major embarrassment always constitutes a normative reason to help one's friend, there might be circumstances in which one has other weightier reasons against helping one's friend. It is a good thing about Jean that she has some motivation to spare her friend an embarrassment, but having that motivation does not guarantee that Jean's action is a response to the fact that in that particular situation, sparing her friend an embarrassment is a *decisive* normative reason to help her. In other words, it might be a mere accident that

[15] Sliwa, "Moral Worth and Moral Knowledge," 6.

what motivates Jean is the same as what makes her action morally right. Intuitively, accidentally doing the thing that one has decisive normative reason to do is not an instance of responding to decisive normative reasons. How are we to capture this nonaccidentality condition on responding to normative reasons?

Very roughly, one might suggest, following Lord, that *A*'s ϕ-ing is a response to a reason-giving fact, *F*, just in case the normative fact that *F* is a normative reason to ϕ explains why *A* ϕ-ed.[16] Your action cannot be a response to a normative reason unless there is an explanatory connection between your action and the normative features of the fact that motivate you to act accordingly. For you to respond to a reason, it is not sufficient that the reason-giving fact somehow motivates you to act in accordance with that reason. *A*'s ϕ-ing is a proper response to a reason only if the reason-giving fact causes *A* towards ϕ-ing *in virtue of* its normative property that it is a reason for *A* to ϕ. Similarly, one's action is a response to the fact that one has *decisive* reason to act accordingly only if the normative fact that one has that *decisive* reason explains why one performs the action. And that is why Jean's action is not a response to the fact that she has *decisive* reason to help her friend, because Jean may not be sensitive to the decisiveness of the reason for which she acts.

Now, if responding to a normative reason requires that one's action is caused by the very fact that one has that normative reason, then it is natural to suppose that for that causal relation to obtain, one should be aware both of the reason-constituting fact and of the normative fact that one has that normative reason and thus be motivated by one's knowledge that one has that normative reason.[17]

However, although responding to normative reasons requires that one somehow recognizes the reason-giving force of the relevant considerations, this recognition need not and should not be spelled out in terms of knowing (or believing) that one has that reason. Among other things, it is not the case that everyone has the ability to have a belief about normative reasons. Presumably, some adults and most children can respond to their reasons and get credit for doing the right thing while they lack the concept of a normative reason, and thus, they cannot have a belief about those reasons. Requiring the presence of a normative belief about reasons overintellectualizes responding to reasons.[18] Furthermore, those who have the concept of a normative reason are not always required (even implicitly) to believe that they have the relevant reason to be

[16] Lord, *The Importance of Being Rational*, 135–40.

[17] Sliwa, "Moral Worth and Moral Knowledge"; and Johnson King, "Accidentally Doing the Right Thing."

[18] Lord, *The Importance of Being Rational*, 1035; and Sylvan, "What Apparent Reasons Appear to Be."

able to respond to that reason. You might properly respond to a reason without believing that you have such a reason. For example, consider a situation in which, by controlling your doxastic reactions, a scientist makes sure that you could never have a normative belief about what you ought to do. It seems plausible that by merely acting on your doxastic states, the scientist cannot prevent you from performing the action that you have reason to perform. In many familiar circumstances, we easily and automatically perform the actions that are required of us without bothering ourselves with thinking about what we ought to do. Moreover, as Cunningham and Howard have argued, there are many cases in which one genuinely performs an act of a certain kind while believing that one's action is not of that kind.[19] For example, you can cook the best pie in the world all the while believing that it is one of the worst. In a similar way, you might correctly respond to all your normative reasons while unfairly criticizing yourself for failing to discharge your obligation. Thus, we need to find a middle ground between the simple account of responding to reasons that fails to provide sufficient conditions and the intellectualist account that falls short of coming up with the necessary conditions for responding to normative reasons.

Fortunately, we can find that ground in the concept of a normative *competence* to treat and respond to reasons. For example, Lord, among others, suggests that we respond to a normative reason to ϕ just in case our ϕ-ing is a manifestation of our *knowing how* to use that reason to ϕ.[20] Importantly, as Lord emphasizes, the know-how condition for responding to a normative reason does not require that the agent believes that she has that reason. This know-how is a competence that disposes you to get things right and can guarantee the explanatory connection between your action and the fact that your action is supported by the reasons.[21] If your action is a manifestation of such a competence to treat and respond to reasons, then it cannot be a mere accident that you perform the action that is actually supported by the reasons.[22]

As I argued at the outset, the reasons-responsiveness account maintains that rationality is about responding to reasons. Focusing on *responding* to reasons implicates the idea that the only reasons that can contribute to what one is rationally required (or permitted) to do—that is, the reasons that one

[19] Cunningham, "Moral Worth and Knowing How to Respond to Reasons"; and Howard, "One Desire Too Many."

[20] Lord, *The Importance of Being Rational*, 116–23. See also Cunningham, "Moral Worth and Knowing How to Respond to Reasons."

[21] Lord, *The Importance of Being Rational*, 117.

[22] Cf. Mantel, *Determined by Reasons*; Isserow, "Moral Worth and Doing the Right Thing by Accident"; and Howard, "One Desire Too Many."

possesses—are the ones that one is able to respond to. Now, if responding to a reason is a matter of manifesting one's competence to treat and respond to that reason, then possessing a reason is partially grounded upon having such competencies. However, it is noticeable that merely having a general ability or competence to respond to a normative reason is not sufficient for possessing that reason. Moreover, one needs to have the opportunity to manifest that ability or competence when the time comes. For instance, the fact that someone has a heart attack is a normative reason for a highly qualified surgeon to perform heart surgery to help that person. But if the surgeon does not have access to the necessary equipment to perform such a complicated surgery, then she does not possess that reason. Since the surgeon does not possess this reason, we cannot criticize her for failing to respond to it, even if she has the general ability to do so. Furthermore, the fact that one possesses the general ability to correctly respond to a kind of reason (e.g., reasons of beneficence) does not guarantee that with respect to all possible situations, one can determine whether the reasons of beneficence are overriding or defeated by other contrary reasons. One does not possess a normative reason in a particular situation if correctly working out the weight of that reason against the background of other present reasons in that particular situation goes beyond one's general ability and competence to respond to that kind of normative reason. Thus, to possess a normative reason in a particular situation, one also needs to have a specific ability to correctly calculate the normative significance of that reason in that particular situation.[23] We can sum up these points in a tripartite account of possessing normative reasons: for you to possess a reason R to ϕ is for you (i) to be aware of the fact that R obtains, (ii) to have the general competence to treat and respond to R as the reason it is, and (iii) to be in a position to appropriately manifest your competence to treat and respond to R as the reason it is.[24]

Now, one might think that if we accept a competence-based account of responding to reasons and of possession, then it would immediately follow that there are possible situations in which an agent competently responds to her decisive normative reason to perform an action while at the same time

[23] For illuminating discussions about the distinction between the general and specific ability to respond to a reason and having the opportunity to do so, see Way and Whiting, "Reasons and Guidance"; Lord, *The Importance of Being Rational*, 235–37; and Schwan, "What Ability Can Do."

[24] Following Lord, one may argue that since you cannot be in a position to manifest your general competence to treat and respond to R unless you have that general competence and also know that R obtains, then this tripartite definition boils down to a simple claim to the effect that to possess a reason R just is to be in a position to appropriately manifest one's competence to treat and respond to R as the reason it is (*The Importance of Being Rational*, ch. 3).

she rationally believes that she lacks sufficient reason to act accordingly. But things are not as simple as they appear. The mere fact that she can competently respond to her normative reasons without believing that she has those reasons does not imply that the agent can also *rationally* form a mistaken judgment about the force and direction of her practical reasons. To argue for that conclusion about the structure of one's normative reasons, we need to defend other premises, and that is the task of the next section.

2. The Rational Criticizability of Normative Judgments

In this section, I present an argument against the reasons-responsiveness account to the effect that there can be situations in which your possessed reasons permit you to believe that you ought to ϕ while simultaneously you possess sufficient reasons not to intend to ϕ. In such situations, the reasons-responsiveness view allows for an irrational mismatch between believing that you ought to ϕ and lacking an intention to ϕ. The upshot is that the reasons-responsiveness view cannot account for an important dimension of rationality—that is, it cannot explain why some interesting instances of practical akrasia are irrational. The argument relies on a premise that the possession of a practical normative reason does not guarantee that there is available evidence for being subject to that reason. The rough idea is that one might be in a position to successfully manifest one's practical competence to respond to a decisive practical reason but fail to be in a position to successfully conclude one's deliberation about whether one possess sufficient reasons for performing that action. The existence of the conditions for possessing a practical reason does not guarantee the presence of the necessary conditions for possessing decisive epistemic reasons for believing that one has such a practical reason. To have an intuitive sense of the argument, let us consider the following example.

> *Apt Emotions*: Jane is a morally good person. The special thing about Jane is that, unbeknownst to her, whenever she becomes aware of the morally relevant facts of her own situation, her emotions competently guide her to do what those facts normatively support, such that she has a perfect track record of successfully doing what is morally required of her. Unfortunately, she decides to choose philosophy as her career. In her first year, she attends a course in morality, and her professor convinces her with a bunch of sophisticated philosophical arguments that one never is permitted to break one's promises—that is, facts about one's promises are always normatively overriding. One day, Jane finds herself in a situation where there is a fact of the matter, say, someone being

> in urgent need, that provides Jane with a strong reason to break her own promise. Jane is aware of all the relevant facts about her situation, and, as always, her competent emotions strongly motivate her toward doing the right thing (in this case, helping the other person). Nonetheless, when she begins to reflectively deliberate on the relevant features, weighing them against each other, she concludes that she has decisive reason to keep her promise and to refrain from helping the person. Being convinced by those sophisticated philosophical arguments, she thinks that consideration of the person in need cannot defeat the normative force of her own promise. Fortunately, when the time comes, she shows weakness of will and fails to act as she herself believes she ought to. Her emotions retain dominance and guide her toward breaking her promise by helping the person. Jane criticizes herself for demonstrating an irrational weakness.[25]

What should we say about the normative status of Jane's action (i.e., her helping the other person)? Before attending the ethics course, if Jane were in a similar situation, her competent emotions would have guided her to do what is morally right—that is, helping the other person—despite the fact that this required breaking her promise. In that situation, the correct verdict is that Jane's action is a correct response to her decisive reason to help the person since her motivation to help them is a manifestation of her normative competence to act according to her decisive reasons. In other words, in that situation, it is a fact that Jane correctly responds to her decisive reason to help the person, and this fact is grounded in the fact that her action is a manifestation of her normative competences. But one feature of the grounding relation is that the ground necessitates the grounded. Whenever the ground exists, the grounded also obtains. Now, in the above example, Jane's action is again, a manifestation of her normative competence. And if the underlying fact about the manifestation of competence is present, then what that fact grounds should also be present.

[25] As some readers may know, the case has a background in the literature of morality. It is a version of Mark Twain's character Huckleberry Finn that was first introduced by Bennett, "The Conscience of Huckleberry Finn." A similar example also can be found in Weatherson, "Do Judgments Screen Evidence?" Most notably, Nomy Arpaly has discussed a bunch of related examples in order to argue that there are cases of rational akrasia. Arpaly, "On Acting Rationally against One's Best Judgment." However, as I will argue, Arpaly holds that in such cases, the agent's normative belief about what she ought to do is rationally criticizable because it is not a correct response to the agent's evidence. See Arpaly, "On Acting Rationally against One's Best Judgment," 498–500, 503, 505. Regarding Apt Emotions, we have every reason to conclude that Jane's judgment about what she ought to do is rational and fully supported by the evidence she possesses.

Thus, we must conclude that Jane's action is a correct response to the normative reasons that decisively support breaking her promise to help the person in need. (More about this below.)

What about the status of Jane's normative judgment? First of all, I think most of us feel inclined to say that Jane cannot be criticized for her doxastic response regarding what her normative reasons support. Despite knowing that the fact that someone needs help constitutes a good normative reason to help, Jane fails to correctly conclude that this consideration is normatively sufficient for breaking the promise. However, her failure to correctly calculate the practical significance of that consideration is not necessarily a failure of rationality. After all, we are assuming that Jane's tutor, an infamous philosopher, has put forward persuasive philosophical arguments in favor of a view according to which one is never morally permitted to break one's promises. It is due to possessing this misleading piece of evidence that Jane fails to conclude that she ought to do the beneficent thing. Jane inculpably lacks the theoretical competencies to find the flaw in the spurious sophisticated arguments.[26] Thus, we cannot legitimately expect her to infer from what she knows that in her particular situation, the reason of beneficence outweighs the normative significance of promise-keeping.[27] The fact that Jane's doxastic reaction to her normative situation is *not criticizable*, I submit, suggests that the normative judgment she holds is *rational* in the sense of being a correct doxastic response to all her *possessed* epistemic

[26] When I write that Jane's doxastic failure is due to her lack of theoretical competencies to find the flaws in those misleading philosophical arguments, I do not mean that Jane lacks the *general* ability or competence to successfully judge that she ought to do the beneficent thing. It is just that she is not in an ordinary position to successfully *exercise* her general ability to work out the balance of reasons through conscious deliberation. As I argue, possessing a cluster of considerations, *R*, as a sufficient reason to believe that one ought to ϕ requires that one be in a position to manifest her general ability and competence to treat and respond to *R* as a sufficient reason for judging that one ought to ϕ. Thus, in Apt Emotions, Jane fails to possess sufficient reasons for judging that she ought to do the beneficent thing because she is not *in a position to exercise* her general theoretical ability to determine whether the reason of beneficence is overriding. Thanks to an anonymous reviewer for pressing me to address this delicate issue.

[27] As Kieran Setiya argues, there seems to be a tight connection between rationality and our legitimate expectations about an agent's actions and attitudes. As a general principle, someone is rationally criticizable in φing only if she could be legitimately expected not to ϕ. Setiya, "Against Internalism," 275–77. Here, Setiya refers to Michael Smith, who claims that "[one] thing we can legitimately expect of rational agents as such is that they do what they are rationally required to" (*The Moral Problem*, 85).

reasons. If there is any possibility of rational false belief regarding normative matters, Jane's situation seems to be an evident instance of that type.[28]

In what follows, I will try to clarify how the fact that Jane's judgment is intuitively not criticizable shows that her judgment is rational.[29] The general argument goes as follows. First, Jane's doxastic reaction is not criticizable in the sense that she is *excused* for her failure to find out that she ought to do the beneficent thing. Second, Jane is excused for her doxastic failure because she is not rationally *required* to believe that she ought to do the beneficent thing. Third, if Jane is not rationally required to judge that she ought to do the beneficent thing, then she is rationally *permitted* to hold a different judgment, that is, to believe that she is permitted or even ought to keep her promise. Based on that rational permission, Jane's normative judgment that she ought to keep the promise constitutes a *rational* doxastic response.

When we treat an agent's normative judgment as not being criticizable, one of the following might be the case: (i) the agent's doxastic response is not rationally evaluable in the first place (that is, it is subject to an exemption); (ii) the agent's doxastic response is rationally evaluable, and it constitutes a praiseworthy achievement of knowingly believing the truth about the normative issue; or (iii) the agent's response is rationally evaluable, and it constitutes an objective yet blameless normative failure (that is, the agent's failure to find the truth about the normative issue is subject to an excuse).[30]

Clearly, Jane's doxastic reaction to her normative situation is not subject to an exemption. There is, for instance, no psychological barrier for her to revise her belief about what she has most reason to do; she possesses all the general abilities for revising her judgement. Needless to say, Jane's normative judgment is an unfortunate objective failure to correctly determine what action she has most reason to perform. So she cannot be praised for an achievement as to

[28] Some philosophers have argued that one cannot rationally make mistakes about some particular normative matters of fact. For example, Titelbaum argues that it is always irrational to have false beliefs about the requirements of rationality ("Rationality's Fixed Point"). Similarly, according to Littlejohn's account of rationality, there is a special class of propositions about the requirements of rationality that we cannot make rational mistakes about (Littlejohn, "Stop Making Sense?"). Unfortunately, I cannot examine and respond to these arguments here. For recent interesting discussions, see Field, "It's OK to Make Mistakes"; Worsnip, "The Conflict of Evidence and Coherence"; and Killoren, "Why Care about Moral Fixed Points?"

[29] I am indebted to an anonymous reviewer for helping me to clarify this issue. In doing so, I have borrowed phrasing from the reviewer's comments.

[30] For an illuminating discussion about the notion of exemption, see Wallace, *Responsibility and Moral Sentiments*, ch. 6. For recent discussions about the nature and normative role of excuses, see Baron, "Excuses, Excuses"; and Sliwa, "The Power of Excuses."

know the truth about the balance of her normative practical reasons. Jane's normative judgment seems to fall under the third category. She cannot be blamed for her failure to find the truth about the normative significance of the relevant reason-giving facts. Why? For one thing, Jane's inculpable lack of the relevant philosophical competencies provides an excuse for her doxastic failure; it appears to be irrational for Jane to simply ignore those philosophical arguments without having anything specific to say about why they are misleading. For instance, if Jane knew about the reliability of her emotions, she would possess sufficient grounds to suspect that those philosophical arguments are misleading. However, as the example suggests, she does not know about her interesting emotional competencies.

Now, the idea is that the same considerations that provide an excuse for Jane's doxastic failure can also make it the case that she did not possess sufficient reasons to believe the truth about what she ought to do in the first place. Taking account of all her possessed epistemic reasons, Jane's false judgment that she ought to keep the promise is to be considered as a rational doxastic response. The following considerations provide motivations to take this further step.

First, it is noticeable that one cannot be excused for ϕ-ing unless one is *rationally permitted* to ϕ. If one is not rationally permitted to ϕ, then one is rationally required to not-ϕ. But the existence of a rational requirement to not-ϕ excludes the possibility of one's being excused for ϕ-ing. If you are rationally required to perform an action, then you cannot rationally make an excuse for failing to perform that action. The whole point of introducing the notion of a *rational requirement*, distinct from an all-things-considered objective "ought" of reasons, is to determine whether an agent's excuse is or is not acceptable. Thus, there is no such thing as blameless, excused, or inculpable irrationality. The fact that *S* is excused for her ϕ-ing suggests that *S* is not rationally required to refrain from ϕ-ing. And the fact that *S* is not rationally required to refrain from ϕ-ing means that *S* has a rational permission to ϕ—that is, there is a way for *S* to rationally ϕ.[31]

[31] The connection between excusability and rational permission is a central feature of recent debates over rationality. As a proponent of the reasons-responsiveness theory of rationality, Lord suggests that the most fundamental feature of the kind of rationality that is at stake in recent debates in metaethics and epistemology is its connection to a certain kind of *blame* or *criticism* that we express with words like "senseless," "stupid," "idiotic," and "crazy" (*The Importance of Being Rational*, 4). See also Scanlon, *What We Owe to Each Other*, 25–30; Parfit, *On What Matters*, 33; and Kiesewetter, *The Normativity of Rationality*, 39. It is exactly this essential feature of the property of rationality that explains why the rationality of one's attitudes and actions is to be determined by the reasons that one possesses, that is, the factors that fall within one's epistemic and practical perspective. For a related

To illustrate, consider Jenny, who always comes home at nine o'clock in the evening, and the first thing she does is to flip the light switch in her hallway. She did so this evening. Jenny's flipping the switch caused a circuit to close. By virtue of an extraordinary series of coincidences, which were unpredictable in advance, the circuit's being closed caused a release of electricity (a small lightning flash) in her neighbor's house. Unluckily, her neighbor was in its path and was therefore badly burned.[32] From the point of view of all the objective facts, Jenny's flipping the light switch is impermissible—that is, she *objectively ought* to refrain from flipping the switch. But intuitively, Jenny is excused for acting against this objective impermissibility. And Jenny's excusability can be explained in terms of another deontic notion to the effect that Jenny was *rationally permitted* to flip the light switch. If she was rationally required to refrain from flipping the switch, then there would be no ground to excuse her for doing something that causes her neighbor being badly burned.

The same, I think, is true about Jane's doxastic response to her normative situation. From the objective point of view, Jane excusably fails to believe the truth about what she ought to do. The explanation of why Jane is excused for her objective doxastic failure lies in the fact that there was no *rational requirement* on Jane to believe that she has most reason to break her promise. If Jane is not rationally required to hold the judgment that she has most reason to break her promise, then she is rationally permitted to believe otherwise. And if Jane is rationally permitted to judge that she has most reason to keep her promise, then we must conclude that her normative judgment is rational.

Moreover, the fact that excusability entails rational permission can be explained on the basis of the more fundamental fact that the elements that excuse an agent's failure to act in accordance with normative reasons are the same ones that undermine her possession of the relevant reasons. The considerations that excuse can fulfil their normative function by defeating one's possession of the relevant normative reasons. For instance, the fact that such and such series of extraordinary coincidences are unpredictable in advance provides an excuse for Jenny's objective normative failure because it shows that Jenny does not possess the relevant objective normative reasons. And if Jenny does not possess the relevant objective normative reasons, then she cannot be rationally required to comply with the demands of those reasons. Likewise, the excusability of Jane's doxastic failure can be explained in terms of the fact that she does not possess sufficient epistemic reasons to believe that she ought to

illuminating discussion to the effect that full excusability entails a rational permission, see Bruno, "Being Fully Excused for Wrongdoing."

[32] The case is borrowed from Thomson, *The Realm of Rights*, 229. For an influential discussion of this case, see Scanlon, *Moral Dimensions*, 47–52.

do the beneficent thing. The fact that Jane lacks the specific theoretical competencies to find the flaws in the spurious philosophical arguments and the fact that Jane does not know that her emotional reaction is a highly reliable indicator of what she ought to do together explain why her false normative judgment is excused: they make it the case that Jane does not possess the fact that someone needs help as a *decisive* reason to believe that she ought to do the beneficent thing.[33] But if Jane does not possess decisive reason to believe that she ought to do the beneficent thing, then she possesses sufficient reason to conclude her deliberation by judging that she ought to keep the promise. Thus, according to the reasons-responsiveness account, her judgment that she ought to keep the promise is rational. This completes my argument that Jane's normative judgment is rational.[34]

In reply, friends of the reasons-responsiveness view might insist that if Jane's action is a genuine response to the fact that she has decisive reasons to break the promise and do the beneficent thing, then she must also possess the same decisive reasons for believing that she ought to break the promise and do the beneficent thing. In that case, her judgment that she ought to keep her promise would turn out to be irrational, at least in a strong and objective sense of the term "rationality."[35] This is what Markovits argues for regarding the classic case of Huckleberry Finn.[36] Out of laudable sympathy, Huck helps his friend Jim, a fugitive, to escape from slavery. However, in the grip of the racist ideology of his culture, he criticizes himself for stealing from Miss Watson, whom he takes to be Jim's "owner." Like Jane, Huck acts against his judgment about what he ought to do, even though his action is, according to Markovits, a correct response

[33] Needless to say, Jane possesses the fact that someone needs help as *some* reason to believe that she ought to help that person. But she does not possess this fact as a *decisive* epistemic reason to believe that she ought to do the beneficent thing in that particular situation.

[34] It is worth noting that arguing for the rationality of Jane's normative judgment is important for the plausibility of my case against the reasons-responsiveness view, for it makes it even clearer why Jane's practical failure to act in accordance with her own normative judgment is irrational. The enkratic principle of rationality requires agents to conform with what they believe they ought to do. Now, if one rationally holds a judgment about what they ought to do, then one seems to have no way of being fully rational unless one acts in accordance with one's own judgment.

[35] For an influential and classic presentation of the distinction between weak and strong rationality, see Goldman, "Strong and Weak Justification." We need to remember that proponents of the reasons-responsiveness account do not endorse dualism about rationality. See Lord, *The Importance of Being Rational*, ch. 7. They support unificationism: the idea that rationality is a function of responding to reasons. For dualism about rationality, see Fogal, "Rational Requirements and the Primacy of Pressure"; Worsnip, *Fitting Things Together*; and Fogal and Worsnip, "Which Reasons?"

[36] Markovits, "Acting for the Right Reasons," 216–17.

to his normative reasons to help Jim. The idea is that Huck's sympathy puts him in direct contact with the normative significance of the fact that Jim is a human being. But Huck intuitively shows some degree of irrationality in acting against his own normative judgment about what he ought to do. Presupposing a reasons-responsiveness theory of rationality, Markovits suggests that if Huck's action is a correct response to his normative reasons, then the apparent irrationality must have something to do with his normative belief. Markovits suggests that the rational criticizability of Huck's normative judgment can be explained in terms of the fact that the "process" by which Huck forms his belief is "deeply flawed."

To examine Markovits's point, it is helpful to discuss a related example suggested by Amia Srinivasan, which involves a clearly objectionable normative belief:

> *Domestic Violence*: Radha lives in rural India, and her husband, Krishnan, regularly beats her. After the beatings, Krishnan often expresses regret for having had to beat her but explains that it was Radha's fault for being insufficiently obedient or caring. Radha finds these beatings humiliating and guilt-inducing; she believes she has only herself to blame and that she deserves to be beaten for her bad behavior. After all, her parents, elders, and friends agree that if she is beaten, it must be her fault, and no one she knows has ever offered a contrary opinion. Moreover, Radha has thoroughly reflected on the issue and concluded that given the natural social roles of men and women, women deserve to be beaten by their husbands when they misbehave.[37]

Srinivasan maintains that Radha's normative belief that she deserves to be beaten is *intuitively* unjustified, and the explanation lies in the fact that her belief is "the product of a convincing, and systematic, patriarchal illusion" about the role of men and women in society.[38] The idea is that if the correct intuitive verdict about Radha's judgment is that she is not only mistaken but also unjustified, then the same verdict and explanation should be true about Huck's and Jane's normative conclusions. There is something importantly in common between the normative judgments of all these figures.

The first point to note is that even if there is an intuitive sense in which Radha's, Huck's, and Jane's beliefs are unjustified, they are all excused for their normative failures. We cannot legitimately expect Radha and Huck to resist the force of the bad ideologies of their cultures. Likewise for Jane. As I argued, Jane

[37] Srinivasan, "Radical Externalism."

[38] Srinivasan, "Radical Externalism," 399.

is not philosophically competent enough to find the flaw in those philosophical arguments that suggest that we are never permitted to break our promises, but it is not fair to criticize Jane for lacking the sophisticated philosophical competence. Having such competence is not part of being a rational agent. The whole point is that the reasons-responsiveness view cannot capture and explain this important dimension of rational excusability.

Second, it is dubious in the first place whether the reasons-responsiveness theory can adequately explain the purported intuitive data that Radha's normative belief is unjustified. One might think, as Srinivasan does, that the example actually supports the traditional externalist theories of justification that tend to explain the rationality of one's doxastic reaction in terms of facts about the reliability of one's doxastic states. Admittedly, the testimonial reasons on which Radha based her normative belief are highly unreliable indicators, but Radha possesses no reason to question the reliability of her misleading reasons. Similarly, Markovits is right that Huck's moral reasoning is deeply flawed, but Huck does not possess any reason to conclude that there is something objectionable about the conclusion of his moral reasoning.[39] Similarly, even though Jane's emotions put her in direct contact with the genuine normative force of helping the person in need, she does not know how to use this fact as a defeater for all those philosophical arguments that allegedly support the idea that she should always keep her promises. In a nutshell: even if we admit that Jane's normative judgment about her reasons remains unreliable and thus, in an important sense, unjustified, the reasons responsivist who wants to acknowledge the role of one's competence and know-how in responding to reasons and possessing them cannot explain the justificatory status of Jane's belief in terms of the normative reasons she possesses.

The latter point about reasoning might seem to suggest another line of resistance for proponents of the reasons-responsiveness view. In their recent paper, Way and Whiting argue that whenever S justifiably believes that she ought to ϕ, then as a matter of fact, S ought to ϕ.[40] In the language of reasons, if S believes for sufficient reasons that she has decisive reasons to ϕ, then she has decisive reasons to ϕ. This is what they call *ought infallibilism*. If ought infallibilism is true, then either Jane's belief that she ought to keep her promise is not sufficiently supported by reasons, or she has decisive reason to keep the promise

[39] The point is that the flaw in Huck's reasoning lies not even in the general rules that Huck follows but rather in the unsound premises that he takes, albeit excusably, for granted. Huck's moral reasoning begins with an unfortunate socially given belief to the effect that helping Jim escape amounts to stealing from Miss Watson. And Huck knows that he is not morally permitted to help satisfy a friend's desire if so doing requires thievery.

[40] Way and Whiting, "If You Justifiably Believe that You Ought to ϕ, You Ought to ϕ."

and therefore lacks decisive reasons to do the beneficent thing. Either way, reasons-responsiveness theory does not provide Jane with rational permission to act against her own normative judgment.

Way and Whiting's argument for ought infallibilism goes as follows.

1. Since it is correct reasoning to move from the belief that you ought to ϕ to deciding to ϕ, then if you justifiably believe that you ought to ϕ, you would be justified in deciding to ϕ.
2. If you justifiably believe that you ought to ϕ, then you can have no other justified attitude from which you could correctly reason to decide not to ϕ. This is because if you can reason from one of your justified attitudes toward deciding not to ϕ, this shows that your justification for believing that you ought to ϕ is defeated.
3. If you have no other justified attitude from which you could correctly reason to decide not to ϕ, you lack justification for deciding not to ϕ.
4. If you are justified in deciding to ϕ, and you lack justification for deciding not to ϕ, then you ought to decide to ϕ. Assuming that reasons for attitudes are restricted to object-given reasons, since you ought to decide to ϕ, you ought to ϕ.
5. Therefore, if you justifiably believe that you ought to ϕ, then you ought to ϕ.

The core idea is that since Jane can correctly reason from her justified belief that she ought to keep her promise to deciding to keep her promise, she must have justification to decide to keep her promise and refrain from helping the person in need. Moreover, Jane cannot correctly reason from her belief that someone needs help to deciding to break her promise and help the person because if it is a correct reasoning route available to Jane, then her normative belief that she ought to keep her promise would turn out to be unjustified. So, either Jane's normative belief is not justified, or she lacks decisive reasons to break her promise and do the beneficent thing.

I have two responses. The first is against the second premise of the above argument. Way and Whiting suggest that if one has a justified attitude from which one can correctly reason toward deciding not to ϕ, then one lacks justification for believing that one ought to ϕ; that is, one's justification for that attitude can defeat whatever justification one has for believing that one ought to ϕ. In Apt Emotions, Jane's emotions lead her to break the promise in response to the reason that someone needs help. This transition seems to be correct reasoning from Jane's justified belief that someone needs help toward breaking the promise; after all, it is an instance of responding to a decisive reason, and at least in a broad sense of the term "reasoning," responding to decisive reasons constitutes correct reasoning. Moreover, this reasoning route is available to Jane because of her emotional character. But having such an incredible emotional character does not guarantee that there is a parallel correct reasoning route available to Jane to reason from her justified belief that someone needs help toward revising her normative belief that she ought to keep her promise. As I understand it, Apt Emotions highlights the possibility of an asymmetrical structure between the reasoning routes that are available to an agent. Why should there be such an asymmetry here?[41]

The rough idea is this: Jane's emotional character puts her in touch with the normative fact that in her particular situation, the reason-giving significance of helping the person is relatively weightier than the normative significance of keeping her promise. Other things being equal, this normative recognition, provided by her emotional competence, would constitute sufficient grounds for Jane to believe that her reasons decisively support doing the beneficent thing. However, as to the details of the example, other things aren't equal. For one thing, Jane has strong albeit misleading evidence, provided by some spurious philosophical arguments, that she is never permitted to break her promises. She might also have been introduced to some recent neuroscientific findings that purportedly show that the moral testimony of one's emotional dispositions is generally unreliable.[42] As long as Jane is not in a position to successfully exercise her theoretical competencies to find the flaw in those sophisticated arguments, she cannot *rationally* rely on her emotional reaction to judge that she ought to help the person. From her deliberative viewpoint, whatever reason she has to break her promise is defeated by the relevant philosophical arguments. At the same time, the fact that Jane lacks such relevant theoretical competencies does not preclude her from being in a position to correctly respond to the fact that someone needs help as a decisive reason to help the person. Jane's action is a manifestation of an emotional competence that correctly detects the overriding normative force of doing the beneficent thing. Despite all those misleading arguments, Jane recognizes that the reason-giving significance of helping the person is weightier than the normative force of keeping her promise. And her action is firmly guided and nonaccidentally motivated by her direct access to this normative fact.

To further clarify the latter point, consider Katie, who senses a strong fear whenever she sees a dangerous spider. Suppose that there is a lawlike, reliable

[41] Thanks to an anonymous reviewer for pressing me to address this question. As they have argued, this challenge targets the core of my case against reasons-responsiveness theory and needs to be dealt with, even if, in the end, we agree that Way and Whiting's argument is unsound for other reasons.

[42] For instance, she may have been introduced to Greene, "The Secret Joke of Kant's Soul" but not to Berker, "The Normative Insignificance of Neuroscience."

connection between the perceptual appearance of a spider and the danger it poses. Unbeknownst to her, Katie is highly sensitive to the nuances of perceptual appearance of different types of spiders, and her fear is always a competent response to the detection of that lawful connection. Unfortunately, Katie receives misleading information from a scientist that spiders of a particular type are harmless. One day she sees a member of that type and believes that it is harmless. Thankfully, however, her fear does not listen to what she believes and causes her to run. Now, it seems to me that despite her competent emotional reaction to the appearance of the dangerous spider, Katie cannot correctly reason from her recognition of the dangerousness of that spider to the conclusion that the testimony of the scientist is misleading. After all, she does not know that she possesses such an extraordinary capacity. At the same time, it seems highly plausible that Katie's fear is a correct and competent response to her normative situation. It is the result of her accurate identification of the danger of that spider. I suggest that something similar happens to Jane and her morally competent emotions. Jane's emotional character puts her in a position to correctly discern the moral significance of the facts in her situation, and she successfully responds to her recognition by doing the beneficent thing. However, she cannot treat her emotional reaction as a conclusive reason against the misleading philosophical arguments and believe that she has decisive reason to break her promise. Like Katie, she is inculpably ignorant about the reliability of her emotional capacities.[43]

One way to make Jane's normative situation intelligible is to say that it is a case of acting for a decisive moral reason without knowing or even being in a position to know that one acts for such a reason.[44] Arpaly nicely brings out the same point when discussing the case of Huckleberry Finn:

> Talking to Jim and interacting with him, Huckleberry constantly *perceives* data (never deliberated upon) that amount to the impression that Jim is a full person, just like Huckleberry himself. While he never

[43] One might object that if Jane cannot rationally revise her normative judgment on the basis of her successful recognition of the overriding force of the fact that someone needs help, then, and because of this, her act of doing the beneficent thing is not a genuine response to the same normative fact. In reply, we must notice that the reasons-responsiveness theory of rationality aims to provide a *reductive* explanation of facts about rationality in terms of facts about competent response to normative reasons. Thus, proponents of this view cannot coherently challenge my verdict that Jane's action is a *competent response* to her normative situation adverting to the fact that it is not *rational* for her to revise her normative judgment on the basis of her recognition of the overriding normative force of benevolence.

[44] For interesting related discussion in epistemology, see Lasonen-Aarnio, "Unreasonable Knowledge"; Kelly, "Peer Disagreement and Higher-Order Evidence," 157–58; and Worsnip, "The Conflict of Evidence and Coherence."

> deliberates on his perceptions, they prompt him increasingly to act toward Jim as a friend.… The idea that we can sometimes act for moral reasons without knowing that we act for moral reasons is not strange when posed against the background of epistemology and psychology, where many have maintained that we can know without knowing that we know, believe without believing that we believe, or *act for a reason without knowing that we act for a reason*.[45]

For another take, one might be inclined to understand Jane's reaction in terms of the notion of performative expertise, as Cholbi suggests:

> Performative experts have the ability to "get it right" within a particular domain, without being able to articulate or justify how the expert gets things right. The skilled marksman or musician displays performative expertise when she hits the target or executes a beautiful musical performance without being able to elaborate the steps by which she achieved these ends or even the criteria for excellence in meeting those ends.[46]

In a nutshell, although Jane is not a *deliberative expert* with respect to moral questions in the sense that we cannot seek advice from her about the normative matters, she functions, in virtue of her emotional competencies, like a moral compass or a performative expert who knows how to correctly react to her normative situation.[47] One way or another, Jane seems to be trapped in an asymmetrical normative situation.

The second point about Way and Whiting's argument concerns their first premise. Undoubtedly, it is correct reasoning if you move from the belief that you ought to ϕ to deciding to ϕ. Accordingly, there is a sense in which you are always permitted to reason in that way. But the question is: How are we to understand and interpret such permission? Way and Whiting suggest that the correctness of enkratic reasoning entails that if you believe that you ought to ϕ based on sufficient reasons, then you have sufficient reason for deciding to ϕ. They hold that you are normatively permitted to reason from your (justified) normative belief that you ought to ϕ to deciding to ϕ in the sense of normative permission that is related to normative reasons. However, this is not the only possible interpretation. One might, following John Broome, argue that we should account for the correctness of reasoning rules in terms of rational

[45] Arpaly, "Moral Worth," 229–30 (emphasis added).

[46] Cholbi, "Moral Expertise and the Credentials Problem," 235. See also Weinstein, "The Possibility of Ethical Expertise."

[47] For related discussion about the notion of a performative expert, see Shepherd, "Practical Structure and Moral Skill."

permissions.[48] Enkratic reasoning is a correct rule because you are rationally permitted to move from your (rational) belief that you ought to ϕ to deciding to ϕ. In other words, the rationality of your normative judgment that you ought to ϕ can make it the case that you are rationally permitted to decide to ϕ. From such rational permission, one cannot draw the kind of permission that is related to normative reasons unless one presupposes the reasons-responsiveness theory of rationality. But that is exactly what is at issue here.

Way and Whiting nonetheless might insist that there is a close connection between correct reasoning and reasons. And I do admit that if my belief that I ought to ϕ is correct, then it would certainly be correct for me to decide to ϕ. According to a plausible account, a belief in *p* is correct if and only if *p* is true. Thus, my belief that I ought to ϕ is correct if and only if it is true that I ought to ϕ. And if it is true that I ought to ϕ, then it is obviously correct for me to decide to ϕ. But we must be careful not to conflate the fact that a belief is correct with the fact that one possesses sufficient reasons for having that belief. Correctness is a function of all the objective normative reasons out there, whether possessed or unpossessed.

Finally, one might follow Karen Jones in arguing that we must distinguish between two kinds of relations in which one stands vis-à-vis reasons.[49] According to the first kind, which we can call *responding* to reasons, agents guide their actions according to a conception of the reasons as *normative reasons*. This kind of action-guidance requires the capacity for reflection, and the agent must be able to judge the balance of the reasons she has. This is something that we can attribute to the *agent* as the thing that she does. But there can be another relation with reasons, namely, *tracking* reasons, which does not require the agent to guide her action through reflection about her reasons. In this sense, an agent can reliably track the reasons in a nonreflective way even though she cannot deliberatively guide herself towards performing that action with the help of a judgment about the reasons she has.

In Apt Emotions, as long as Jane's reflective abilities are concerned, she rationally judges that she has decisive reason to keep her promise, and it is not rationally possible for her to believe that the fact that someone needs help makes it the case that she has decisive reason to break her promise. And thus, she is not in a position to *respond* to the reason she has for helping that agent. The correct description of Jane's performance is to say that she merely tracks the relevant decisive reason in virtue of the sub-agential, reliable capacity she has for tracking reasons. She acts as a reason tracker, not as a reason responder.

[48] Broome, *Rationality through Reasoning*.

[49] Jones, "Emotion, Weakness of Will, and the Normative Conception of Agency," 189.

Now, the suggestion is that if Jane cannot act as a reason responder with respect to the fact that someone needs help, this fact cannot contribute to what is *rational* for her to do. The reasons that can make a difference to the rational status of one's actions and attitudes are the reasons one can respond to—but not the reasons that one can only track.

Two points are in order. First, remember that in the previous section, I argued that whether an agent's (doxastic or practical) reaction is a correct response to her normative reasons is a matter of the existence of an explanatory relation between her performance and the relevant normative truths about reasons, and that explanatory connection can obtain even in the case of tracking reasons. When one tracks a normative reason to ϕ, the fact that one has such a reason can and does explain one's ϕ-ing.

Second, and more importantly, I explained at the outset that one of the most interesting features of the reasons-responsiveness account of rationality is that this theory can vindicate the normativity of rationality in the sense that the same things—that is, (possessed) normative reasons—determine both what one ought to do and what one is rationally required to do. I think it is highly plausible that the set of reasons that contributes to what one ought to do (or what one is justified to do) must include the reasons that one can only track. If we restrain the potent reasons to the reasons that one can only respond to in this demanding sense of responding to reasons, there cannot be any normative truth about what children or some adults ought to do. Presumably, children and some adults lack the concept of a reason, or they lack relevant reflective abilities required for responding to reasons. But if we assume that the potent reasons include all and only the reasons that one can reflectively respond to, then it would be inappropriate to say that those who lack such reflective abilities ought to do something or that they have justification to perform an action. But this is utterly unacceptable. The potent, ought-making reasons must include the reasons that one can merely track. Thus, if it is true that the same things must determine what one ought to do and what one is rationally required to, as the reasons-responsiveness theory assumes, then we must allow the reasons that one can only track among the rationalizing reasons.

3. Concluding Remarks: The Irrationality of Akrasia

In the previous section, I have argued for and explained why it is metaphysically possible that the demand of one's reasons for action diverges from what is rational for one to believe about one's reasons. There are situations in which, from the perspective of normative reasons, an agent is required to act against her rationally held judgment about what she ought to do. And this suggests that the reasons-responsiveness theory of rationality is untenable because in those situations, the agent's action against her own normative judgment about what she ought to do is plainly irrational. That is, there is something to rationality that is beyond the reach of the reasons-responsiveness theory.

As a last resort, the proponents of the reasons-responsiveness view might be tempted to question the latter idea, arguing that it is not always irrational to act against one's own judgment about what one ought to do. And I do agree that there are cases in which one's acting against one's own normative judgment cannot be rationally criticized. But the structure of those cases is markedly different from cases in which there is a mismatch between one's reasons for action and one's reasons for having an attitude towards what the reasons demand. For example, suppose that an evil demon ensures that whenever you come to rationally believe that you ought to do something, you fail to intend to bring it about. In that case, your failure to adjust your intentions with what you believe about what you ought to do is not an instance of irrationality, even though from an objective point of view, your mental states involve incoherency. You cannot be rationally criticized for the failure, since the failure is excusable. Or consider a more mundane situation in which you have come to believe that you are normatively required to save someone's life from a grave danger, but your fear paralyzes you such that you cannot take the necessary course of action. Again, it seems highly plausible that you are not rationally criticizable for your failure because we cannot legitimately expect people to overcome such a complete though local inability.[50] But we cannot assume that whenever an agent responds to her decisive reasons to perform an action while she rationally believes that she ought to take another option is a case where she loses her ability to materialize her normative judgment to the extent that her failure turns out to be rationally excused.

In response, one might argue that the above cases do not exhaust all of the cases of rational akrasia. After all, everyone has to agree that correctly responding to reasons is at least part of what rationality consists in. Thus, even if there are cases where the demands of one's normative reasons require one to be in an otherwise irrationally incoherent state, since the rational force of responding to reasons is always greater than the rational significance of being coherent, then overall, one is always rationally required to follow the guide of one's possessed reasons.[51] Moreover, it is not clear what kind of normative achievement one would demonstrate if one were to act in accordance with one's false normative judgment. Would it be any better for you to refrain from performing the action

[50] But cf. Broome, *Rationality through Reasoning*, 173.

[51] Arpaly, "Moral Worth," 36.

that you have decisive reason to perform just because you rationally believe that you ought to take another option?

I have two points in response. First, we must notice that sometimes agents are caught in unfortunate dilemmatic situations in which there is no chance of perfect normative success. Old examples include situations in which our evidence decisively suggests that we ought to adopt an incorrect belief such that acting on that belief would bring about a disaster. We cannot explain the importance of rationality unless we adopt a long-run perspective. Even though our disposition to follow the lead of our rational normative judgments does not always guide us towards performing the best option, it is still a highly valuable disposition, and we cannot arbitrarily prevent its manifestation if we aren't in a position to determine whether we are in a case where it leads us astray.

Second, the above argument presupposes that facts about rationality are or entail facts about decisive reasons. The idea is that if in your acting against your own judgment about what you ought to do you are actually responding to the balance of your normative reasons, then there cannot be anything criticizable about your performance. But from the perspective of normative reasons, that there is nothing criticizable about your performance does not imply that you cannot be criticizable from the perspective of rationality unless we already assume that rationality is a matter of responding to normative reasons.

There is something distinctively wrong about acting against one's rational judgment about what one ought to do. And we cannot explain that distinctive kind of failure in terms of facts about normative reasons. Now, whether we should identify this particular type of failure as a failure of rationality or a failure of another sort is, I suspect, a verbal dispute.[52] My case against the reasons-responsiveness theory of rationality provides a new argument for skepticism about the strong normativity of rationality.[53] The normativity of rational requirements cannot be explained in terms of the demands of decisive normative reasons.[54]

Institute for Research in Fundamental Sciences, Tehran
hadisafaei@ipm.ir

[52] See Lasonen-Aarnio, "Enkrasia or Evidentialism?" and "Coherence as Competence."

[53] Kolodny, "Why Be Rational?"

[54] For comments and discussion, I am grateful to Mohsen Eslami, Faraz Ghalbi, Hossein Khalaj, Marziyeh Lotfi, Mahmoud Morvarid, Amir Saemi, Mohsen Zamani, and two anonymous referees for the *Journal of Ethics and Social Philosophy*.

REFERENCES

Alvarez, Maria. "Reasons for Action: Justification, Motivation, Explanation." In *Stanford Encyclopedia of Philosophy* (Winter 2017). https://plato.stanford.edu/archives/win2017/entries/reasons-just-vs-expl/.

Arpaly, Nomy. "Moral Worth." *Journal of Philosophy* 99, no. 5 (May 2002): 223–45.

———. "On Acting Rationally against One's Best Judgment." *Ethics* 110, no. 3 (April 2000): 488–513.

Baron, Marcia. "Excuses, Excuses." *Criminal Law and Philosophy* 1, no. 1 (January 2007): 21–39.

Bennett, Jonathan. "The Conscience of Huckleberry Finn." *Philosophy* 49, no. 188 (April 1974): 123–34.

Berker, Selim. "The Normative Insignificance of Neuroscience." *Philosophy and Public Affairs* 37, no. 4 (October 2009): 293–329.

Broome, John. *Normativity, Rationality and Reasoning: Selected Essays*. Oxford: Oxford University Press, 2021.

———. *Rationality through Reasoning*. Chichester: Wiley-Blackwell, 2013.

———. "Reasons." In *Reason and Value: Themes from the Moral Philosophy of Joseph Raz,* edited by R. Jay Wallace, 204–28. Oxford: Oxford University Press, 2004.

Brunero, John. *Instrumental Rationality: The Normativity of Means-End Coherence*. New York: Oxford University Press, 2020.

Bruno, Daniele. "Being Fully Excused for Wrongdoing." *Pacific Philosophical Quarterly* 104, no. 2 (August 2022): 324–47.

Cholbi, Michael. "Moral Expertise and the Credentials Problem." *Ethical Theory and Moral Practice* 10, no. 4 (August 2007): 323–34.

Cunningham, J. J. "Is Believing for a Normative Reason a Composite Condition?" *Synthese* 196, no. 9 (November 2017): 3889–910.

———. "Moral Worth and Knowing How to Respond to Reasons." *Philosophy and Phenomenological Research* 105, no. 2 (August 2021): 385–405.

Dancy, Jonathan. *Practical Reality*. Oxford: Oxford University Press, 2000.

Davidson, Donald. "Freedom to Act." In *Essays on Actions and Events,* 63–81. Oxford: Oxford University Press, 2001.

Field, Claire. "It's OK to Make Mistakes: Against the Fixed Point Thesis." *Episteme* 16, no. 2 (August 2017): 175–85.

Fogal, Daniel. "On the Scope, Jurisdiction, and Application of Rationality and the Law." *Problema* 12 (December 2018): 21–57.

———. "Rational Requirements and the Primacy of Pressure." *Mind* 129, no. 516 (October 2020): 1033–70.

Fogal, Daniel, and Alex Worsnip. "Which Reasons? Which Rationality?" *Ergo* 8, no. 11 (2021).

Goldman, Alvin. "Strong and Weak Justification." *Philosophical Perspectives* 2 (1988): 51–69.

Greco, Daniel. "A Puzzle about Epistemic Akrasia." *Philosophical Studies* 167, no. 2 (January 2013): 201–19.

Greene, Joshua. "The Secret Joke of Kant's Soul." In *Moral Psychology*, vol. 3, edited by Walter Sinnott-Armstrong, 35–79. Cambridge, MA: MIT Press, 2007.

Horowitz, Sophie. "Epistemic Akrasia." *Noûs* 48, no. 4 (May 2013): 718–44.

Howard, Nathan. "One Desire Too Many." *Philosophy and Phenomenological Research* 102, no. 2 (September 2019): 302–17.

Isserow, Jessica. "Moral Worth and Doing the Right Thing by Accident." *Australasian Journal of Philosophy* 97, no. 2 (April 2018): 251–64.

Johnson King, Zoe. "Accidentally Doing the Right Thing." *Philosophy and Phenomenological Research* 100, no. 1 (August 2018): 186–206.

Jones, Karen. "Emotion, Weakness of Will, and the Normative Conception of Agency." In *Royal Institute of Philosophy Supplement*, edited by Anthony Hatzimoysis, 181–200. Cambridge: Cambridge University Press, 2003.

Kelly, Thomas. "Peer Disagreement and Higher-Order Evidence." In *Disagreement*, edited by Richard Feldman and Ted A. Warfield, 111–74. New York: Oxford University Press, 2010.

Kiesewetter, Benjamin. *The Normativity of Rationality*. Oxford: Oxford University Press, 2017.

Killoren, David. "Why Care about Moral Fixed Points?" *Analytic Philosophy* 57, no. 2 (June 2016): 165–73.

Kolodny, Niko. "Why Be Rational?" *Mind* 114, no. 455 (July 2005): 509–63.

Lasonen-Aarnio, Maria. "Coherence as Competence." *Episteme* 18, no. 3 (September 2021): 353–76.

———. "Enkrasia or Evidentialism? Learning to Love Mismatch." *Philosophical Studies* 177, no. 3 (November 2018): 597–632.

———. "Unreasonable Knowledge." *Philosophical Perspectives* 24, no. 1 (2010): 1–21.

Lee, Wooram. "The Independence of Coherence." *Synthese* 199, nos. 3–4 (February 2021): 6563–84.

Littlejohn, Clayton. "Stop Making Sense? On a Puzzle about Rationality." *Philosophy and Phenomenological Research* 96 (December 2015): 605–27.

Lord, Errol. "The Coherent and the Rational." *Analytic Philosophy* 55, no. 2 (January 2014): 151–75.

———. *The Importance of Being Rational*. Oxford: Oxford University Press, 2018.

Mantel, Sussane. *Determined by Reasons: A Competence Account of Acting for a Normative Reason*. New York: Routledge, 2018.

Markovits, Julia. "Acting for the Right Reasons." *Philosophical Review* 119, no. 2 (April 2010): 201–42.

McCain, Kevin. "The Interventionist Account of Causation and the Basing Relation." *Philosophical Studies* 159, no. 3 (February 2011): 357–82.

Parfit, Derek. *On What Matters*, vol. 1. Oxford: Oxford University Press, 2011.

———. "Reasons and Motivation." *Aristotelian Society Supplementary Volume* 71, no. 1 (July 1997): 99–130.

Raz, Joseph. *From Normativity to Responsibility*. Oxford: Oxford University Press, 2011.

Scanlon, T. M. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge, MA: Belknap Press, 2008.

———. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press, 1998.

Schroeder, Mark. "Means-End Coherence, Stringency and Subjective Reasons." *Philosophical Studies* 143, no. 2 (February 2008): 223–48.

Schwan, Ben. "What Ability Can Do." *Philosophical Studies* 175, no. 3 (March 2018): 703–23.

Setiya, Kieran. "Against Internalism." *Noûs* 38, no. 2 (May 2004): 266–98.

Shepherd, Joshua. "Practical Structure and Moral Skill." *Philosophical Quarterly* 72, no. 3 (July 2022): 713–32.

Sliwa, Paulina. "Moral Worth and Moral Knowledge." *Philosophy and Phenomenological Research* 93, no. 2 (June 2015): 393–418.

———. "The Power of Excuses." *Philosophy and Public Affairs* 47, no. 1 (July 2019): 37–71.

Smith, Michael. *The Moral Problem*. Cambridge, MA: Blackwell, 1994.

Srinivasan, Amia. "Radical Externalism." *Philosophical Review* 129, no. 3 (July 2020): 395–431.

Sylvan, Kurt. "What Apparent Reasons Appear to Be." *Philosophical Studies* 172, no. 3 (March 2015): 587–606.

Thomson, Judith Jarvis. *The Realm of Rights*. Cambridge, MA: Harvard University Press, 1990.

Titelbaum, Michael. "Rationality's Fixed Point." In *Oxford Studies in Epistemology*, vol. 5, edited by Tamar Szabó Gendler and John Hawthorne, 253–94. Oxford: Oxford University Press, 2015.

Turri, John. "Believing for a Reason." *Erkenntnis* 74, no. 3 (May 2011): 383–97.

Wallace, R. Jay. *Responsibility and Moral Sentiments*. Cambridge, MA: Harvard University Press, 1994.

Way, Jonathan, and Daniel Whiting. "If You Justifiably Believe that You Ought to ϕ, You Ought to ϕ." *Philosophical Studies* 173, no. 7 (July 2016): 1873–95.

———. "Reasons and Guidance (or, Surprise Parties and Ice Cream)." *Analytic Philosophy* 57, no. 3 (July 2016): 214–35.

Weatherson, Brian. "Do Judgments Screen Evidence?" Unpublished manuscript, 2010. https://brian.weatherson.org/quarto/posts/jse/jse.html.

Weinstein, B. D. "The Possibility of Ethical Expertise." *Theoretical Medicine and Bioethics* 15, no. 1 (March 1994): 61–75.

Williams, Bernard. "Internal and External Reasons." In *Moral Luck: Philosophical Papers 1973–1980*, 101–13. Cambridge: Cambridge University Press, 1981.

Worsnip, Alex. "The Conflict of Evidence and Coherence." *Philosophy and Phenomenological Research* 96, no. 1 (September 2015): 3–44.

———. *Fitting Things Together: Coherence and the Demands of Structural Rationality*. New York: Oxford University Press, 2021.

———. "Making Space for the Normativity of Coherence." *Noûs* 56, no. 2 (February 2021): 393–415.

———. "Reasons, Rationality, Reasoning: How Much Pulling-Apart?" *Problema* 12 (December 2018): 59–93.

Journal of Ethics and Social Philosophy
vol. 28, no. 3 · october 2024

https://doi.org/10.26556/jesp.v28i3.3777

EDUCATIONAL JUSTICE AND THE VALUE OF EXCELLENCE

Tammy Harel Ben Shahar

Promoting educational justice and nurturing educational excellence are two values many hold dear. Education systems declare their commitment to realizing both, yet in many cases, there is inescapable tension between them. While educational justice typically entails prioritizing the needs of low achievers in decisions concerning institutional design, educational practices, and resource allocation, developing educational excellence presumably requires preferring the needs of those already educationally advantaged.

For example, prioritizing educational excellence might require investing scarce educational resources in developing gifted programs instead of providing these resources to children with low and average abilities, even if they have yet to obtain rudimentary educational skills. Other policy decisions involve irresolvable tension between the two goals, even when there is no shortage of resources. Thus, the most beneficial student assignment policy for low achievers typically involves mixed-ability classes, whereas separating high-ability students to designated programs may be preferable for developing excellence. Assignment policy cannot reconcile these opposing requirements and requires prioritizing one over the other. And finally, pedagogical and curriculum choices can benefit students with a specific set of abilities and be less suitable for others. Therefore policy decisions as well as everyday classroom practices often require making tough decisions and prioritizing the development of the abilities of some students over the abilities of others.

The tension described above between developing excellence on the one hand and developing more mundane abilities on the other hand is one of the most basic tensions in the educational justice literature. It is also the starting point and the motivational driving force of this paper, which aims to contribute to the discussion of this distributive dilemma through the exploratory examination of the concept of excellence.

One approach to resolving the tension is to compare the case of ability to the distribution of wealth. Theories of distributive justice contend with similar issues and must evaluate policies that will affect the distribution of wealth between individuals who are unequally well-off. The solutions offered by theories of distributive justice might be applicable to the dilemma at hand. Notwithstanding significant variation between their prescriptions, many theories of distributive justice accord initial priority to those who are less well-off. Despite that priority, in the appropriate circumstances, many theories of justice allow the better-off to benefit despite their favorable starting point. For example, our concern for the disadvantaged should not prevent allocating resources to ensure political participation for all, including the wealthy. Also, when investment in those already better-off results in large returns, and especially when these returns benefit everyone, including the worst-off (by creating jobs, technological developments, etc.), accumulation of additional wealth by the well-off is often considered justified.

The same considerations may apply in the educational domain, creating a duty to invest in all children, including those who are already better-off, especially when investing in high-achieving students can potentially benefit many people, including the least well-off.[1]

Yet despite these similarities, an interesting disanalogy emerges between the educational domain and distributive justice more generally. Theories of distributive justice routinely require inhibiting the accumulation of wealth for the sake of redistributing it to the less well-off. If the urgent needs of the poor require it, we would not be especially troubled if none of those who are better-off become extremely rich.[2] Taking from the rich is an inseperable component of distributive justice.

On the other hand, in the educational domain, despite the moral importance of promoting the educationally disadvantaged, hindering the development of educational excellence at the very top of the distribution of abilities is not treated with the same indifference. As Harry Brighouse states, "If we worry too much about ensuring that the least advantaged get a fair shot at labour market advantage, we jeopardize the production and discovery of excellence."[3] As opposed to the loss of wealth, the loss of excellent talent seems to many to be intuitively undesirable, even if that loss facilitiates the development of the educationally disadvantaged. This suggests that developing excellence is special in some way that renders the standard considerations of distributive justice inapplicable.

Yet while many might share the intuitive aversion toward policies that "jeopardize the production of excellence," philosophers have not developed

[1] Brighouse and Swift, "The Place of Educational Equality"; Meyer, "Talent Advancement"; and Harel Ben Shahar, "Distributive Justice in Education and Conflicting Interests."

[2] Beyond a certain threshold, some even argue that surplus wealth has "zero moral weight," since people are already fully flourishing. See Robeyns, "Having Too Much."

[3] Brighouse, "Educational Equality and School Reform," 39.

a principled moral argument that supports this intuition and examines its implications. They have not discussed *why* developing excellence is important—specifically, whether it is important "for its own sake" or for instrumental reasons—nor do they provide a systematic account of how we ought to balance the concern for developing excellence with duties we have toward the least advantaged (in resource allocation, student assignment, or education practices). Within the educational justice debate, the importance of developing excellence "is more asserted than argued for."[4] For example, Brighouse attests that he values high ability and even values it "for its own sake," although he admits he "can't give much of a justification for valuing it."[5] Elizabeth Anderson also maintains that "the development of human talents is a great intrinsic good, a good to the person who has it, and a good to others," but does not explain why we should value developing human talent for its own sake nor how this consideration should be factored in complicated real-life decisions.[6]

Instead, there is some philosophical discussion concerning educational practices that focus on nurturing excellence, such as gifted education and private or selective schools.[7] These contributions (some of them supportive of practices that prioritize excellence and others that criticize them) discuss the definition of educational excellence and spell out the tension between developing it and promoting other educational goals. But these too do not scrutinize the value of developing high ability, which they take for granted.

The lack of principled examination of the value of excellence is unfortunate, as it might lead to misguided decisions in distributive dilemmas of the sort presented above; more specifically, the aversion to loss of excellence might result in giving the development of excellence more than its due moral weight, thereby undermining policy aimed at promoting students with low abilities.

To contribute to an informed discussion of the tension between developing excellent versus low or average abilities, this paper takes a closer look at excellence. It is an exploratory project that aims to discover what makes developing excellent ability valuable, whether it is valuable for its own sake, and whether the value of developing excellence is special compared to developing ability at any other level. If developing excellence is indeed unique, further questions

[4] Brighouse, "Educational Equality and School Reform," 40.

[5] Brighouse, "Educational Equality and School Reform," 39–40.

[6] Anderson, "Fair Opportunity in Education," 615.

[7] See, for example, Sapon-Shevin, "Playing Favorites"; Meyer, "Educational Justice and Talent Advancement"; Merry, "Educational Justice and the Gifted"; Swift, *How Not to Be a Hypocrite*; Mazie, "Equality, Race and Gifted Education"; Mason, "Fair Equality of Opportunity and Selective Secondary Schools"; and Harel Ben Shahar, "Ability and Ability Grouping."

involve whether it should outweigh the value of developing ability at other levels and how it affects our duties to the educationally disadvantaged.

The conclusion, in a nutshell, is that developing excellent ability is valuable in numerous instrumental and noninstrumental ways to those who possess it as well as to others, but all kinds of value created by developing excellence are created also by developing ability at other, lower levels. Since the same *kind* of value is created by developing all abilities (albeit in different amounts), decisions concerning education policy (such as resource allocation, pedagogy, student assignment, and so on) should be made by weighing the gains and costs of alternative educational options and would thereby result more often than not in favoring the development of abilities at the lower end of the "ability continuum." Thus, I suggest that in many cases, our intuitive aversion to restricting programs aimed at developing excellence is misguided.

Sections 2–5 of the paper explore several ways in which excellence is valuable. Section 2 examines the immense instrumental (financial and vocational) value of developing excellent abilities for the individuals possessing them and for others who enjoy their exercise. Both of these types of value, I argue, are not special to developing excellent abilities and are generated by developing ability at any level, including the low and average levels of ability. And while we may think that developing excellence generates more value than developing abilities in the lower range, I offer several explanations why in many cases—if not most cases—developing the abilities of the least advantaged is in fact more beneficial than developing excellence.

Sections 3–6 discuss the value of developing excellence that is not vested in its vocational or financial consequences. Gaining a deep understanding of the world, appreciating art and literature, and developing high-order thinking skills are valuable for individuals not only because of what they can "do" with these abilities but also as an end in themselves (section 3). Further, we might think that human excellence is impersonally valuable, meaning that it is a good thing even if it is not good for anyone in particular—a possibility developed in section 4. Finally, in section 5, I examine how excellence is valuable because it elicits inspiration, which is a noninstrumentally valuable human experience.

Analysis of these three noninstrumental types of value leads, perhaps surprisingly, to similar conclusions as the examination of vocational and financial value in section 2. In other words, I argue that developing ability at any level creates the same kind of benefit, although the amount of value may vary. Enjoying rational capacities, appreciating reading, and comprehending the world we live in are valuable at all levels of ability, not only at the highest range. For example, working hard to improve one's guitar-playing skills even if the result is amateurish is valuable for the same reasons as honing virtuosic guitar-playing abilities. Impersonal value and inspiration can also be generated, I argue, by developing lower abilities. Accomplishments that are mediocre in absolute terms can inspire awe if obtaining them involves overcoming extraordinary difficulties.

If this is the case, there is nothing special in the value that is created by developing excellence, and the value of developing excellence is on par with developing abilities in general. As a result, developing excellence cannot stand as an independent consideration in debates concerning policy design or resource allocation, let alone automatically override the concern for developing low abilities. Instead, decisions regarding policy and resource allocation must assume that each choice entails developing some students' abilities and should consider *how much* value is developed in each case and at what cost. "Demistifying" excellence by showing that the value it creates is comparable to the value created by developing all human ability serves to reassure us that although we may be intuitively averse to compromising the development of excellent human potential, it is often the inescapable and justified outcome of what we are morally required to do.

The final section of the paper offers some guidance for balancing the value of developing excellence with the value of developing ability at other levels. Since developing abilities at all levels is valuable for the same kinds of reasons, decision-making must be sensitive to facts and weigh all the relevant moral considerations. Philosophers can contribute to this, as many already do, by offering careful, empirically informed analysis of specific practices and general principles. A sophisticated understanding of the value of excellence (and other abilities) is indispensable in such endevors. I argue further that although developing ability has various kinds of value, when confronted with concrete cases of balancing, our primary concern should be with the vocational and financial (i.e., instrumental) gains of developing ability. The noninstrumental value of developing ability, which is the focus of this paper, is typically less morally urgent as well as less tangible than some of the instrumental gains of developing ability and therefore should be relegated to secondary status.

Before proceeding to these conclusions, however, I set the stage by defining the concepts *ability* and *excellence*.

1. Ability and Excellence

While there may be various ways to understand excellence, my focus on education and the development of excellence therin means that it is useful first to define ability. However, defining ability in the context of education is "complex and fraught with difficulty."[8] The "chronic ambiguity" that the concept suffers from is related to the fact that "ability" has several different meanings and is used in many different contexts.[9] There are countless different human abilities and talents; some need serious work and training to develop, while others come naturally to most people. Even within the educational domain, numerous types of abilities are relevant—including specific skills such as solving mathematical problems and more general capabilities such as critical thinking.

"Ability" is also often used interchangeably with other terms, including intelligence, IQ, talent, aptitude, skills, and more.[10] All of these have slightly different meanings and are used by scientists and educators to describe different things. To make things even more complicated, not all abilities that philosophers refer to in their work can be measured by empirical tools, creating discrepancies between theoretical policy recommendations and what is possible in practice.

Absent a single "correct" definition, the appropriate understanding of ability depends on the topic and context of the discussion and needs to be explicitly stated in each case. For the purposes of this paper, I focus on the type of abilities that are specifically relevant to schools, namely, those developed by schools and by educators. This, I think, is the way in which most of the philosophical work on educational justice uses the term.

This very tentative definition, however, needs further explanation. Specifically, there are two questions that need to be answered to make the definition more precise. First, when discussing educational justice and especially duties concerning the development of abilities, we must understand what abilities schools can (and should?) develop and perhaps also obtain knowledge concerning the abilities they actually do develop as a matter of fact. (Schools inevitably vary significantly in their success in developing abilities.) The second question concerns the difference between abilities that schools develop and the abilities that schools measure, since it is often the case that schools measure (and reward) only one subset of the important abilities they develop. Thus, while schools can nurture various "soft" skills, such as communication skills, time management, problem-solving, and leadership skills, these are rarely measured in any systematic way. An account of the duties of educational justice concerning ability should take all of these into consideration.

As to the first issue—namely, which kinds of abilities schools can and do develop—the practice of education (and schooling more specifically) is based

[8] Terzi, "On Educational Excellence," 96. See also Harel Ben Shahar, "Ability and Ability Grouping"; Marley-Payne, "Rethinking Nature and Nurture in Education"; and Robb, "Talent Dispositionalism."

[9] Harel Ben Shahar, "Ability and Ability Grouping," 401.

[10] Robb, "Talent Dispositionalism."

on a working assumption according to which schools are able to develop students' abilities (at least some abilities and to some extent). The assumption seems self-evidently true, since schools are clearly successful in developing abilities to perform certain actions.[11] For example, schools teach children to read and write and to solve basic mathematical problems. Most would also agree that schools (when they are adequate) can develop additional abilities that are not as particular as these skills. For example, schools develop "domain independent" skills such as critical thinking, the ability to construct and evaluate logical arguments, and more.

The possibility that educational interventions performed in schools can improve "general ability" or intelligence is more contested. The consensus in the scientific community is that abilities in general (and intelligence more specifically) are only partly hereditary, and abilities are a result of a "dynamic interplay between genes and experience."[12] Education (and schools more specifically) can therefore potentially affect general ability. Indeed, some studies show that simply attending school has positive effects on tests that evaluate domain-independent cognitive skills.[13] Educational interventions are especially promising for young children and children whose environments are not sufficiently nurturing.[14] The possibility of successful intervention in such cases may have significant import in terms of the scope of efforts we are morally required to invest in children whose background circumstances may have impaired their ability. Despite this optimistic possibility, except in extreme cases (such as children who have been abused), the effect of educational interventions on general abilities is probably limited, and general ability is a relatively stable property of individuals.[15]

[11] Thompson, "A Limited Defense of Talent as a Criterion for Access to Educational Opportunities"; and Harel Ben Shahar, "Ability and Ability Grouping."

[12] Sweatt, "The Emerging Field of Neuroepigenetics," 624; Carroll, *Human Cognitive Abilities*; Marley-Payne, "Rethinking Nature and Nurture in Education"; and Harel Ben Shahar, "Redefining Ability, Saving Educational Meritocracy."

[13] Ceci and Williams, "Schooling, Intelligence, and Income;" McCrea, Mueller, and Parrila, "Quantitative Analyses of Schooling Effects on Executive Function in Young Children"; Burrage et al., "Age and Schooling Related Effects on Executive Functions in Young Children"; and Bergman Nutley, "Gains in Fluid Intelligence after Training Non-Verbal Reasoning in 4-Year-Old Children."

[14] Finn et al., "Cognitive Skills, Student Achievement Tests, and Schools."

[15] Some critics argue that educational interventions merely improve test-taking skills and cannot affect general ability. See Steinberg, "My House Is a Very Very Very Fine House"; Finn et al., "Cognitive Skills, Student Achievement Tests, and Schools"; and Neisser et al., "Intelligence."

The second challenge is the discrepancy between the abilities that schools develop and the abilities that schools measure. Tests administered in schools, especially standardized tests, often focus on knowledge and a narrow subset of skills, failing to evaluate other cognitive abilties and skills.[16] Psychological and emotional skills (such as self-regulation, coping with frustration, and resilience), social skills, and social and cultural capital are also developed in schools, and they are important in the production of excellence; yet they are not measured in tests and evaluations. Tests are also notoriously prone to biases, and therefore their reliability in measuring even narrow abilities is questionable.[17] An account of educational justice that focuses only on the abilities that are currently measured in schools may be overly narrow and overlook many abilities developed in schools (as well as inequalities in the development of these abilities).

For the sake of this article, I choose a definition that encompasses more than the abilities developed and measured by schools, following Lorella Terzi's recent conceptual analysis of educational excellence.[18] Her definition of ability is pluralistic in two ways. First, it includes capabilities that are detected in tests but also what Terzi characterizes as qualitative achievements involving deep understanding, critical skills, creativity, etc.; and second, it includes abilities in various areas that are developed in schools (traditional academic subjects) but also art, physical abilities, and more.[19] This definition does not take schools as they are—not all schools develop and measure all of these skills—but it also does not significantly depart from contemporary schools as we know them, nor does it adopt a completely idealized version of schooling.

Moving on from the concept of "ability" to "excellence," I part ways with Terzi's definition. Terzi defines educational excellence as high but not extraordinary achievement, whereas I am interested in abilities in the highest range, of the kind that schools assume justify special programs and treatment such as gifted education. This choice is driven by the underlying dilemma that motivates the paper, namely, whether (and when) the importance of developing excellence outweighs our concern for the educationally disadvantaged. The strongest case for prioritizing excellence can be made, I think, by considering the development of outstanding rather than merely high abilities, and I define

[16] Gardner, "Multiple Intelligences"; and Bloomberg, "Multiple Intelligences, Judgement, and Realization of Value." But see also White, "Illusory Intelligences?"

[17] Erwin and Worrell, "Assessment Practices and the Underrepresentation of Minority Students in Gifted and Talented Education"; Ford, "Desegregating Gifted Education"; Garda, "The New IDEA"; and Steinberg, "My House Is a Very Very Very Fine House."

[18] Terzi, "On Educational Excellence."

[19] Terzi, "On Educational Excellence," 98, 101.

excellence accordingly. It is these extraordinary abilities that are needed to create sterling accomplishments that are deemed especially socially valuable.[20] If excellence should be given priority over fostering ability in general, the best justification might be found at the very top of the scale.[21] Clearly, though, the conclusions of this exploration will have normative import for high ability more widely construed.

Another comment concerns the possibility that students may demonstrate excellent abilities in one domain and mediocre or even low abilities in other domains.[22] Although we are accustomed to thinking of students as "high ability," "gifted," or "low ability" without making distinctions between different academic abilities, this is often oversimplistic. For our discussion, this means that policy-making needs to be able to make nuanced decisions, sometimes prioritizing a specific student in one domain (say, foreign languages) and deeming that same student a low priority in another (art, for example).

2. Developing Excellent Ability as a Means to Other Ends

I should say at this point that the categorization of the different types of value, especially the distinction between instrumental and noninstrumental value, is itself the subject of much dispute and theorization.[23] For example, developing high ability is valuable because it enables a person to enjoy literature. This value could be classified as noninstrumental because it does not lead to any financial or vocational rewards. It could, however, be classified as instrumental, depending on one's notion of intrinsic value. Since hedonists consider pleasure as intrinsically valuable, developing high ability even if only for personal

[20] Cooper, *Illusions of Equality*; and Kramer, *Liberalism with Excellence*.

[21] Another difference between my conception of excellence and Terzi's is that I am interested in excellence of individuals, whereas Terzi focuses on excellence as a property of education systems. See Terzi, "On Educational Excellence," 93n3. My interest in the excellence of individuals stems from the aim of the paper, namely, to address the tension between nurturing individuals with excellent abilities and promoting the education of those less advantaged. Doing so requires examining the value of developing the excellent abilities of high-achieving students.

[22] Terzi, "On Educational Excellence"; Allen, *Education and Equality*; and Hurka, *Perfectionism*, 167. Although students with high cognitive abilities tend to perform well across different areas of academic abilities. See Deary et al., "Intelligence and Educational Achievement." Our interest lies in excellence with respect to a plurality of abilities (including artistic and athletic), which makes it more likely that different students may excel in different things. Ultimately then, students with excellent abilities are not a homogeneous and distinct social group.

[23] See, for example, Korsgaard, "Two Distinctions in Goodness"; and Anderson, "Value in Ethics and Economics."

enjoyment is instrumentally valuable. For the sake of our discussion, however, the choice of terms is unimportant. My task is to think clearly about the different ways in which developing abilities is good (and for whom) and to examine whether this value may justify prioritizing developing excellent abilities compared to fostering abilities at other levels. To forward this aim, we need not commit to specific classifications of types of value as long as we provide a precise characterization of each of the ways in which excellence is valuable.

Human abilities, widely understood, are a means to pursue all life plans and human endeavors. As such, schooling that develops abilities has value for the individuals educated because they lead to vocational and financial benefits. Excellent abilities generate especially high value for individuals since they open up a wide range of valuable options—opportunities for higher education, high-paying and meaningful jobs, and more.

While nurturing excellent abilities is indeed extremely valuable for the individuals who have them, the same kind of value is also created through developing the ability of students who are less able. Nurturing those who currently demonstrate low abilities and improving their abilities (especially abilities that are valued in the employment market) can create access to a wider range of jobs and ensure that individuals are more financially independent and able to lead more autonomous lives.

What about the value that society derives from developing excellent abilities? Society benefits from people developing high abilities because their exercise leads to advancement in science, culture, and human thought. The outstanding human achievements that result from the exercise of high abilities (in health care, transportation, and communications, for example) improve the well-being of all members of society, including those least able. Impeding the development of excellent abilities by divesting in programs that nurture them therefore might result in the loss of these valuable things. This value, one might argue, does not have a counterpart at the lower levels of ability because developing excellence involves pushing humanity forward in ways that would be impossible for those with lesser capabilities.

While I do not dispute the unique role of people with high abilities in advancing humankind or that their contribution is a good reason (probably the best reason) to invest in nurturing excellence, I insist that developing basic and intermediate abilities is also extremely instrumentally beneficial for society.[24] And despite some differences that will be described, this value, as well as the domains in which it is expressed, is of the same nature as the value of developing excellent abilities.

[24] Harel Ben Shahar, "Distributive Justice in Education and Conflicting Interests."

Advancing the abilities of those at the lower part of the ability spectrum benefits society since it significantly reduces expenditure on welfare, crime and law enforcement, as well as health care for a range of health conditions that are associated with poverty and lack of education, such as diabetes and substance abuse. The resources society currently invests in remedying these social (and medical) ills could instead be directed to other endeavors that improve the wellbeing of all members of society, such as ensuring access to quality and advanced health care, funding scientific research, supporting culture, and more. Cultivating abilities at the low and middle ranges also has a direct positive effect on other members of society, including those with high ability. Crime and other social problems affect not only the least well-off but also other members of society who might be victims of these crimes or of public health problems characteristic of poor and uneducated population.

So increasing longevity, improving public health, and vitalizing scientific research are the kinds of social benefits that can be gained by developing the ability of those at the lower end of the ability spectrum. These are of course the same benefits society gains by nurturing excellent abilities, as discussed above. The difference between advancing abilities at various levels then is not the *kind* of value created nor the *domains* in which this value may be manifested (health, science, culture, and more). Rather, the difference lies in the *way* in which abilities translate into social benefits, and these can vary vastly between those with low abilities and those with high abilities, on account of the different circumstances and characteristics of those groups. Developing high ability typically contributes to society by nurturing the people who will lead innovation, whereas developing ability at lower levels can contribute to society by preventing social problems, accommodating growth, and enabling society to invest in promoting well-being and development.

There may also be differences in the *quantity* of value created by developing ability at different levels, but as I will now explain, developing high ability is not always the socially beneficial choice.

Having concluded that developing ability at any level creates the same kind of instrumental value, we are left with questions of proper distribution. How to balance the relative instrumental value of developing ability at different levels depends on the specific circumstances of each case. Sometimes, it may be especially important to invest in developing excellence, such as when there is a shortage in scientists or when society is facing a public health crisis. In other cases, prioritizing low ability may be more socially beneficial, for example, in a society with high illiteracy rates.

Furthermore, perhaps surprisingly, despite the value of developing exceptional abilities described above, developing abilities at the lower end of the spectrum is often more morally important than developing high ability. First, for the individual developing their ability, it is likely that marginal utility diminishes with regard to abilities in education, meaning that basic skills such as reading, writing, and basic arithmetic bring larger gains to people who develop them than do more advanced abilities. Acquiring these basic skills is especially beneficial because they are preconditions to more human activities and projects than extremely advanced skills. For example, while mastering high-level calculus is instrumental for certain occupations and projects, there are almost no human projects in modern society that do not necessitate reading and performing basic mathematical actions. As a result, the instrumental benefits we gain from basic skills are greater than those we derive from high-level skills, and similar resources are likely to bring higher returns when invested in developing those with lower ability.

Admittedly, there may be cases in which a small improvement for individuals with especially low abilities requires huge investment of resources or cases in which gains at the top levels of ability generate especially high gains, such as abilities needed for deciphering the human genome. Also, benefits may follow nonlinear patterns so that the gains do not neatly correlate with different levels of ability. The examination of gains and costs would therefore have to be performed at a high level of specificity.

To make things even more complicated, developing excellence is an insatiable goal. While the minimal abilities needed for successfully joining the workforce or for accessing higher education can be determined quite specifically (depending on specific job requirements, admission policies, etc.), developing excellence is more elusive. Even the highest ability can be further improved, so excellence defies attempts to define its end, and there is no such thing as "sufficient" investment in it. As a result, the demands of excellence on the limited resources available to education may be endless, whereas above a certain threshold of ability, the gains from further improvement may not rise proportionately.

In terms of social benefits, uncertainty exists with regard to the *exercise* of developed abilities. People may develop abilities but fail to exercise them (for various reasons including personal and motivational), and the social benefit from their development may ultimately come to nothing. When comparing instrumental gains and costs, we have to keep in mind that developing ability does not guarantee that the expected value will be realized through its exercise. While uncertainty qualifies discussion of potential costs and gains of developing any kind of ability, I think it is especially hard to predict the outcomes of developing excellent abilities. Mundane abilities can (and must) be exercised in a wide range of activities and occupations. Converesly, only one in so many people who develop excellent ability actually achieves the kind of feats that make high ability so instrumentally valuable for society—such as finding a cure for cancer or writing a literary masterpiece. Others will develop excellent ability but fail to create extraordinary social value. They may utilize their abilities to their private benefit alone. They may also fail to make these contributions because producing works of genius takes more than high ability (requiring creativity, time, effort, and luck). Unfortunately, we cannot tell in advance who will produce these excellent achievements. Maximizing abilities would perhaps be the best strategy to cope with this problem, assuming that only some of those who develop their ability will in fact "deliver" on their promise. However, given limited resources and the mutually exclusive needs of different students, this is impossible, so the uncertainty must be calculated into the value that society gains from developing these exceptional abilities.

But a utilitarian calculation of costs and gains is only part of the input required for calculating vocational and financial value. Weighing the value of ability as a means to other ends is also subject to moral constraints. Thus the equal moral status of individuals would prevent following utilitarian considerations if those imply depriving an individual of a fundamental human right. For example, we might think that we should not give absolute priority to the disadvantaged if it meant denying free education to advantaged students. Moreover, principles of justice will affect how we weigh the different benefits gained. Those committed to a sufficientarian principle of educational justice, for example, assign more moral weight to developing abilities below the adequacy threshold, even when those create the same benefits as abilities above the threshold. Since most theories of justice prioritize the worst-off in some way, improvements on the lower side of the ability scale would usually end up being more morally important, all things considered, than those at the top end of the distribution.

3. The Value of High Ability as an End in Itself

When people talk of education (and other things too) as having intrinsic value, what they often mean is that developing ability has value as an end in itself.[25] Possessing high ability seems to make one's life better simply in virtue of having it, even if it has no beneficial effect in terms of access to employment and even when life might be happier or easier if one did *not* possess excellent abilities.

[25] Korsgaard discusses the term "intrinsic value" and argues two separate distinctions should be drawn: between instrumental (as a means) and final (as an end), and between intrinsic (value within the object) and extrinsic (value related to something else). See Korsgarrd, "Two Distinctions in Goodness." See also Anderson, *Value in Ethics and Economics*, 3.

For example, in *How Not to Be a Hypocrite*, Adam Swift states, "It matters to me that my children grow up to be able to appreciate—I mean really appreciate—Shakespeare. It matters because, other things equal, I think people who can appreciate Shakespeare live more fulfilling lives than those who can't."[26] In other words, even if appreciating Shakespeare is not a means to anything else and if one's life is just as enjoyable without it, life is made better by being able to appreciate Shakespeare.[27]

George Sher describes the intrinsic value of gaining a "wide and deep knowledge of the world, and of one's place in it" and how lives are made better by having "scientific, historical, and social insight."[28] Developing rationality and human capabilities in general can also be thought of as valuable in this way, above and beyond the instrumental benefits they may generate.[29]

Admittedly, the value of developing ability is derivative of the value of possessing this ability, since it is usually impossible to have a certain ability without going through the process of procuring it. However, the process of *developing* ability through intellectually stimulating and challenging learning is itself valuable as an end.[30] Overcoming intellectual challenges, solving puzzles, and discovering new things bring about intellectual pleasure and a sense of worth and fulfillment, which is why even disregarding the possibility that abilities may help realize other ends, they are of value. This also explains why unexercised abilities can be valuable to those who develop them. They can become a part of people's identity, thereby enriching their lives, and they can contribute to one's self esteem and sense of achievement.[31] Developing high ability therefore is valuable for the individual possessing it, not merely as a means to some other end but also as an end in itself.

The end value of developing ability, I argue, is created at all levels of ability, from the very highest to the lowest. Developing excellent ability does not differ *in kind* from value created by developing other levels of ability. It is valuable to develop the high ability needed to "really" appreciate Shakespeare, but it is also valuable in the same way to develop ability sufficient to appreciate other,

[26] Swift, *How Not to Be a Hypocrite*, 26.

[27] I think that the most plausible interpretation of Swift is that appreciating Shakespeare makes a life better, not more pleasurable. Other actions might generate comparable enjoyment, but that enjoyment would be less valuable than appreciating Shakespeare.

[28] Sher, *Beyond Neutrality*, 121. See also Hurka, *Perfectionism*; Kramer, *Liberalism with Excellence*; and Sypnowich, *Equality Renewed*.

[29] Hurka, *Perfectionism*; and Sher, *Beyond Neutrality*.

[30] Merry, "Educational Justice and the Gifted."

[31] Harel Ben Shahar, "Distributive Justice in Education and Conflicting Interests"; and Robb, "Talent Dispositionalism."

less demanding forms of literature. Gaining any understanding of the world we live in, rather than the deepest understanding of it, to give another example, is valuable as an end and makes people's lives better, other things being equal, than lives in which they have no such understanding. This, I think, is true even in cases of singular abilities such as those of medalist athletes or musical prodigies. The development of excellence at those heights becomes one of the most defining parts of the athlete's or musician's identity and is tightly linked to their self-worth and sense of accomplishment. This same kind of value (albeit perhaps weaker) is generated in cases of amateur marathonists, for example, who gain a sense of accomplishment and empowerment from meeting self-set goals such as improving their time or extending the distance of their run. Running becomes a part of who they are and how they define and present themselves to others.

In other words, as several perfectionist philosophers stress, placing value on developing human capabilities (as an end in itself) does not necessarily entail elitism.[32] Thomas Hurka, for example, states that the perfectionist good of rationality can be performed either at a theoretical level or at a practical level.[33] Therefore, it is not only the philosopher that can live a good life according to the perfectionist standard but also the shopkeeper who is required to make innumerable decisions based on rational deliberation. Realizing one's rational capacity, according to Hurka's account of perfectionism, does not necessarily entail maximizing cognitive ability but rather developing the ability to lead one's life on the basis of rational decision-making. In fact, the value of leading rational lives as an end in itself can actually have an egalitarian pull because it grounds a claim for enabling as many people as possible to develop their capacity for rationality rather than investing in those who are already able to practice rationality but who can nonetheless develop their rational abilities further.[34]

As described regarding ability as a means to other ends, circumstances affect *how much* noninstrumental value is generated from developing ability. People may have different ends, different levels of awareness of their abilities, and different attachment to them. Sometimes when people have exceptionally high ability in a certain domain, it becomes especially important to them, so improving it is extremely valuable. On the other hand, there may be cases when small improvements at the very bottom range of ability make a big difference by introducing people to new areas of interest that significantly enrich their lives and become a part of their identity. This raises complicated questions of

[32] Arneson, "Perfectionism and Politics."

[33] Hurka, *Perfectionism.*

[34] Arneson, "Perfectionism and Politics"; Nussbaum, *Frontiers of Justice*; Sypnowich, "A New Approach to Equality"; and Campbell, Nyholm, and Walter, "Disability and the Goods of Life."

quantification, which I will say more about in section 4. The important point for now is that the *kind* of end value created by developing excellence does not differ from that created by developing ability in general. If so, any intuition we might have that suggests that it is especially wasteful to not develop excellent abilities (and therefore that education policy should be designed to ensure their development) is misguided.

Can excellence bring about end value for others as well as for the individual possessing it? David Cooper answers in the affirmative, arguing that when "some scale the heights," unique value is created for those who observe it, regardless of any other benefit it may create. We should not, he argues, be concerned in the same way for a "general, marginal improvement in the amateur playing of string quartets, or at the times clocked by run-of-the-mill club runners; but [in] seeing the highest standards of musicianship maintained and advanced, with seeing great athletes break new barriers."[35]

Notice however, that what creates enjoyment and appreciation for others is the *exercise* of excellent abilities and not their development or existence *per se*. Possessing excellent ability is a precondition for creating great works of art, literature and science, which are valuable for individuals who derive pleasure and appreciation from them. But developing or possessing excellent ability, as opposed to exercising it, does not seem to have final value for anyone except the person possessing it. Think of an extremely gifted painter who irrationally believes that he is obligated to never create a single work of art. It seems unlikely that an ability unpracticed, or practiced only in private, is still valuable for others.[36]

[35] Cooper, *Illusions of Equality*, 55. Not any capability developed creates value. As Lorella Terzi points out, the notion of excellence relates to our theory of good, so perfecting abilities that are unvaluable (such as the ability to count grass blades) or have social disvalue (such as the ability to plan and execute perfect crimes) does not bear intrinsic value. Terzi, "On Educational Excellence," 96.

[36] Korsgaard offers a similar example concerning a beautiful painting that is locked up permanently in a closet, arguing that the good is conditional on someone seeing it. See Korsgaard, "Two Distinctions in Goodness," 196. Note that the distinction between having an ability and exercising it, which is quite powerful regarding musical, athletic, or artistic abilities, is harder to sustain when we think of cognitive abilities such as contemplation, critical reasoning, or understanding. These abilities are exercised all the time through spontaneous reactions to stimuli in our surrounding world, and they generate end value for individuals possessing them. Enjoyment of other people's excellent cognitive abilities usually does not occur spontaneously but rather in response to accomplishments such as books or inventions that more obviously require diligence and hard work.

4. The Intrinsic Value of Excellent Ability

Though there may be various possible ways to understand the term "intrinsic value," I refer to it here in the following sense: when an object has intrinsic value, it means that its goodness lies in its properties and does not depend on it being good for anyone.[37] This type of value is harder to grasp but has been alluded to, for example, in order to ground the value of nature independently of its value for humans.[38] In his canonical work on value, Moore suggested that things have intrinsic value when they are valuable "in isolation"—namely, if they are good even if nothing else exists at all.[39]

Excellent human ability can be good in such an impersonal sense: not good for anybody in particular but in the abstract. Excellent athletic abilities, musical abilities, or mathematical genius can be valuable *simpliciter* in the same way we think that beauty or nature is good: valuable not because of the pleasure or increase in well-being that it brings to a specific agent but because some things, excellent things especially, are good in themselves.

Not all philosophers endorse the concept of impersonal value, and those who do disagree upon the specific goods that have such intrinsic value. But assuming we ascribe intrinsic value to human abilities, this does not yet entail that only high abilities have intrinsic value. When we value nature—a tree, for example—it would be odd to ascribe value only to the tallest tree, the greenest one, or the one that yields the most fruit. True, it is reasonable to value the Great Barrier Reef more highly than just any random part of the ocean, but the difference between the two is vested in *how much* we value them rather than in the kind of value they have. The entire ocean is arguably still valuable in and of itself, so it would be worthy of protection and sustaining, even if there were no humans around to appreciate it. Similarly, the most persuasive version of the view that attributes intrinsic value to human capabilities involves ascribing such value to abilities of any level. Kant's approach toward intrinsic value demonstrates this. "The good will," understood as the practice of fully rational choices, is the only thing intrinsically valuable according to Kant.[40] Rational choices, however, are made by people with a range of abilities rather than only by people with the highest rationality.[41]

[37] Korsgaard, "Two Distinctions in Goodness"; Green, "Two Distinctions in Environmental Goodness"; and Langton, "Objective and Unconditioned Value."

[38] Green, "Two Distinctions in Environmental Goodness."

[39] Moore, "The Conception of Intrinsic Value."

[40] For a discussion of this, see Korsgaard, "Two Distinctions in Goodness."

[41] Hurka, *Perfectionism.*

5. THE VALUE OF INSPIRATION

There is an aspect of the noninstrumental value of developing excellence that seems unique to developing excellence (rather than any ability)—namely, the value of inspiration. Yet as I soon demonstrate, even inspiration is not unique to excellent ability.

Extraordinary human abilities inspire people. They set an example for the unlimitedness of human spirit, they ignite the imagination, they move and motivate us, and they can even create a sense of community and solidarity between the people sharing the experience. Feeling awe when observing outstanding abilities is a valuable human experience that can enrich our lives, even if it has no further specific beneficial consequences. Noticing the special value of excellent feats, Matthew Kramer goes as far as to argue that "the excellence of the society through its furtherance of sterling accomplishments will heighten the level of self-respect which each of its members is warranted in experiencing."[42] People belonging to a society that creates such excellence feel warranted self-respect, and this, according to Kramer, justifies governmental support of actions needed to foster outstanding achievements.

Although I consider inspiration to be valuable as an end in itself, it can also have instrumental value because it motivates people to excel and to persevere in the face of difficulty (but this would be considered together with other instrumental benefits of developing excellence). Note that inspiration is a reaction to excellent achievements (or the exercise of abilities) but also directly to the *development* of extraordinary abilities, since the effort and talent involved in developing outstanding abilities is itself subject to admiration.

At first brush, it might seem that inspiration is elicited only when people scale the heights, and as such, it provides at least one sense in which excellent abilities are uniquely valuable. Upon closer examination, however, the value of inspiration is also, I argue, not reserved solely for excellent abilities. It is warranted not only when abilities are high in absolute terms but also when abilities are *comparatively* high. In a neighborhood basketball scene, for example, a local hero can elicit inspiration even if her abilities are only exceptional compared to her amatuer friends. Even in an imaginary dystopian scenario in which human excellence dwindles significantly (due to denying resources to the brightest, for example), the good of admiration could still exist. It would simply be directed toward relatively outstanding abilities instead of toward excellent abilities according to an absolute scale.

[42] Kramer, *Liberalism with Excellence*, 36.

Further, we are also inspired when confronted with people who succeed in developing abilities against all odds, even if the ensuing abilities are not excellent in absolute terms. For example, the achievements of Paralympic athletes may fall short of the highest possible human abilities in absolute measures of speed or height, but the abilities developed and demonstrated warrant awe equal to or indeed greater than the excellent abilities developed by athletes without disabilities.[43]

We might, however, be able to distinguish between two different types of inspiration: one is the response to effort, grit, and perseverance, whether or not the outcome is objectively excellent; the other is the awe we feel when we behold outstanding accomplishments. This second kind of emotional reaction is unrelated to the effort invested in it, much like the emotional response we might experience when we see a beautiful landscape, sunset, or butterfly. Indeed, I concede that when we see magnificent works of art or listen to a divine masterpiece, we may experience a strong emotional reaction simply in virtue of the beauty of what we are witnessing. But I think that even in these cases, awe is related to the ability needed to make such a perfect creation. If the subject of our admiration were easily accomplished, I suspect it would not elicit the same emotional reaction.

As a result, I contend that even the value of inspiration can be gained in response to abilities across the board. If all of the above is persuasive, then the value of developing excellent ability is not of a unique kind, and whatever value it has is created also (to varying degrees) by developing ability at all levels. It follows then that education policy, including resource allocation, pedagogy, student assignment, and other issues, should be determined by weighing the value and disvalue created by alternative possible educational policies.

6. Concluding Remarks

Developing abilities is extremely valuable for the individuals possessing them, for others, and even in impersonal and intrinsic ways that do not depend on the abilities being good for anyone in particular. The discussion above was an exploratory one, aiming to understand the different value that is created by developing human ability and specifically to determine whether developing excellence involves the creation of special value that is not created in developing ability at other levels—low ability, average ability, and even high (but

[43] Notably, however, writers and activists in the disabilities movement have referred critically to the fact that people with disabilities are often regarded as "inspiring" for doing the most ordinary things such as working, getting married, raising children. See, for example, Grue, "The Problem with Inspiration Porn."

not quite excellent) ability. This examination revealed that although we might intuitively think that developing excellence is valuable in a way that developing other types of ability is not, this is actually not the case: the differences between developing ability at different levels are vested not in the *kind* of value derived but in *how much* value is created in each case.

Where does this leave us in terms of the distributive dilemma that motivated the paper? Namely, assuming that excellence is not valuable in unique ways, how should we address educational dilemmas that involve tension between developing excellence and developing ability at lower levels? Should we invest scarce resources in programs for gifted children or in funding educational aides for students with low abilities? Should teachers choose materials that will challenge high achievers if students with average or low abilities would gain more from other curricular choices? And should we allow ability grouping even though it is not the most desirable assignment policy for children with low abilities, if it will push forward the very best students?

Balancing the expected gains and costs of educational options according to their effect on all levels of ability is a complicated task—empricially and normatively—that cannot be performed properly here. I will, however, venture to provide some guiding comments that can be gleaned from the discussion above.

Evaluating the relative weight of ability's value involves two separate questions. The first concerns value of the same kind, for example, figuring out whether one policy is more instrumentally valuable than another. The second must factor in different kinds of value for a comprehensive evaluation of educational options. The first issue is theoretically less difficult but involves taking into consideration a lot of information that is not always available. For example, it would matter how financially rewarding it is to develop high abilities in a certain society; what ends specific people have and what abilities are needed to pursue them; whether there are alternative pathways (apart from schools) to developing certain abilities; how many people in society have the potential to develop specific abilities and how many people with those abilities are needed for society to prosper; how many people in society have substandard abilities and what the social costs of that reality are; whether specific individuals have one or more excellent abilities; how costly it is to develop (excellent and low) abilities; and more.

Since these considerations and many more should be factored in each case, designing simple and conclusive guidelines for decision-making is impossible, even with regard to the first challenge—evaluating value of the same kind. The most promising way forward is through empirically informed philosophical discussion of specific educational practices. Philosophers of education engage in this important work routinely, and the observations made in this paper accentuate the importance of continuing this line of research. The analysis here contributes to such projects by clarifying the various aspects of value that excellence and ability in general have and prompting those who take part in these debates to give excellence its due weight.

The second challenge, namely how to provide an integrated assessment of different and incommensurable kinds of value, is theoretically more complicated. But it is also, I argue, more pressing in theory than in practice. For the sake of real-life decision-making, we should usually give predominant weight to the instrumental value of developing ability. Other types of value, including end value, intrinsic value, and the value of inspiration, are insignificant except in special cases, as we will see shortly.

The different types of value attached to developing ability are typically created simultaneously. Individuals seldom gain one without the other. The instrumental financial and vocational benefits of developing (both high and low) ability have a tangible effect on people's well-being. They enable people to become independent and live autonomous lives; they improve people's chances of pursuing higher education and having interesting and meaningful occupations instead of working in menial, boring, and demeaning jobs or perhaps even being involved in crime. By developing abilities, society gains productive citizens and reduces costs associated with poverty and crime. Likewise, the instrumental benefits society gains from developing the abilities of high achievers are also concrete: inventions lead to improvements in life expectancy, health, and economic growth, potentially improving well-being for many individuals.

As opposed to the palpable benefits described above, the value of developing ability as an end in itself is quite amorphous. We value our abilities as an end in themselves, meaning that our lives are better with them. But the noninstrumental gains are secondary compared to the instrumental benefits that permeate every single aspect of our lives. In the balance between obtaining the concrete gains that developing ability has to offer on the one hand and the value of "really appreciating Shakespeare" on the other, one might reasonably prioritize the former. The same, I argue, can be argued for intrinsic value and the value of inspiration. While we may accept that developing ability is intrinsically valuable and may inspire others, neither seems as morally urgent or weighty as some of the more practical instrumental aspects of developing ability.

Luckily, when abilities are developed, both kinds of value are created. So while the moral importance of developing ability is vested primarily in the instrumental benefits it generates, noninstrumental value is created at the same time and spread (even if unequally) across the whole spectrum.

The upshot is that ethical consideration of educational policy-making should focus predominantly on the instrumental value of developing ability. Considering intrinsic value may be appropriate, however, in special cases when it provides an important and unique consideration. For example, fostering a disabled person's artistic abilities could be very valuable for that individual as an end in itself even if it does not create any vocational or financial gains (or even if it does not make that person happy) given how it can fill their lives with meaning. Still, instrumental value typically provides policymakers with the most morally significant information and should therefore be at the center of decision-making processes.

Developing students' abilities, ideally to their maximal potential, is one of the goals of education. Surely, many educational practices are able to attend to the needs of students with high and low ability alike, and efforts should be directed to develop and implement pedagogies that make this possible. Additionally, sufficient educational resources should ideally be directed to multiple ends, meeting the needs of children with diverse needs and abilities. Still, in many cases, distributing resources and designing educational policy entail prioritizing either the development of basic competencies or the development of excellent ones. Clarifying the value of excellence helps us to strike a balance between these competing aims and to accord excellence its appropriate moral weight.[44]

University of Haifa
tharel@univ.haifa.ac.il

REFERENCES

Allen, Danielle. *Education and Equality*. Chicago: University of Chicago Press, 2016.

Anderson, Elizabeth. "Fair Opportunity in Education: A Democratic Equality Perspective." *Ethics* 117, no. 4 (July 2007): 595–622.

———. *Value in Ethics and Economics.* Cambridge, MA: Harvard University Press, 1993.

Arneson, Richard J. "Perfectionism and Politics." *Ethics* 111, no. 1 (October 2000):

[44] The author wishes to thank Sigal Ben-Porath, Steve De Wijze, David Enoch, Adi Goldiner, Kirsten Meyer, Catherine Robb, Yair Sagi, Idit Shafran Gittleman, Barbara Vetter, Dustin Webber, and participants of the workshop on Developing Talent held by the Human Abilities Centre for Advanced Studies in the Humanities in Berlin for their helpful comments. The research was supported by the Israel Science Foundation grant no. 848/19.

37–63.

Bergman Nutley, Sissela, Stina Söderqvist, Sara Byrde, Lisa B. Thorell, Keith Humphreys, and Torkel Klingberg. "Gains in Fluid Intelligence after Training Non-verbal Reasoning in 4-Year-Old Children: A Controlled, Randomized Study." *Developmental Science* 14, no. 3 (2011): 591–601.

Bloomberg, Doug. "Multiple Intelligences, Judgement, and Realization of Value." *Ethics and Education* 4, no. 2 (2009): 163–75.

Brighouse, Harry. "Educational Equality and School Reform." In *Educational Equality*, edited by Graham Haydon, 15–70. London: Continuum International Publishing Group, 2011.

Brighouse, Harry, and Adam Swift. "The Place of Educational Equality in Educational Justice." In *Education, Justice, and the Human Good: Fairness and Equality in the Education System*, edited by Kirsten Meyer, 14–33. New York: Routledge, 2014..

Burrage, Marie S., Claire Cameron Ponitz, Elizabeth A. McCready, Priti Shah, Brian C. Sims, Abigail M. Jewkes, and Frederick J. Morrison. "Age and Schooling Related Effects on Executive Functions in Young Children: A Natural Experiment." *Child Neuropsychology* 14, no. 6 (2008): 510–24.

Campbell, Stephen M., Sven Nyholm, and Jennifer K. Walter. "Disabiltiy and the Goods of Life." *Journal of Medicine and Philosophy* 46, no. 6 (December 2021): 704–28.

Carroll, John B. *Human Cognitive Abilities: A Survey of Factor Analytic Studies.* Cambridge: Cambridge University Press, 1993.

Ceci, Stephen J., and Wendy M. Williams. "Schooling, Intelligence, and Income." *American Psychologist* 52, no. 10 (1997): 1051–58.

Cooper, David E. *Illusions of Equality.* New York: Routledge, 1982.

Deary, Ian J., Steve Strand, Pauline Smith, and Cres Fernandes. "Intelligence and Educational Achievement." *Intelligence* 35, no. 1 (2007): 13–21.

Erwin, Jesse O., and Frank C. Worrell. "Assessment Practices and the Underrepresentation of Minority Students in Gifted and Talented Education." *Jounral of Psychoeducational Assessment* 30, no. 1 (February 2012): 74–87.

Finn, Amy S., Matthew A. Kraft, Martin R. West, Julia A. Leonard, Crystal E. Bish, Rebecca E. Martin, Margaret A. Sheridan, Christopher F. O. Gabrieli, and John D. E. Gabrieli. "Cognitive Skills, Student Achievement Tests, and Schools." *Psychological Science* 25, no. 3 (March 2014): 736–44.

Ford, Donna Y. "Desegregating Gifted Education: Seeking Equity for Culturally Diverse Students." In *Rethinking Gifted Education*, edited by James H. Borland, 143–58. New York: Teachers College Press, 2003.

Garda, Robert A. "The New IDEA: Shifting Educational Paradigms to Achieve Racial Equality in Special Education." *Alabama Law Review* 56, no. 4 (2005):

1071–134.

Gardner, Howard. *Multiple Intelligences: New Horizons in Theory and Practice*. New York: Basic Books, 2006.

Green, Karen. "Two Distinctions in Environmental Goodness." *Environmental Values* 5, no. 1 (February 1996): 31–46.

Grue, Jan. "The Problem with Inspiration Porn: A Tentative Definition and a Provisional Critique." *Disability and Society* 31, no. 6 (2016): 838–49.

Harel Ben Shahar, Tammy. "Ability and Ability Grouping." In *Handbook of Philosophy of Education*, edited by Randell Curren, 401–12. New York: Routledge, 2022.

———. "Distributive Justice in Education and Conflicting Interests: Not (Remotely) as Bad as You Think." *Journal of Philosophy of Education* 49, no. 4 (November 2015): 491–509.

———. "Redefining Ability, Saving Educational Meritocracy" *Journal of Ethics* 27, no. 3 (September 2023): 263–83.

Hurka, Thomas. *Perfectionism.* Oxford: Oxford University Press, 1993.

Korsgaard, Christine M. "Two Distinctions in Goodness." *Philosophical Review* 92, no. 2 (April 1983): 169–95.

Kramer, Mathew. *Liberalism with Excellence.* Oxford: Oxford University Press, 2017.

Langton, Rae. "Objective and Unconditioned Value." *Philosophical Review* 116, no. 2 (April 2007): 157–85.

Marley-Payne, Jack. "Rethinking Nature and Nurture in Education." *Journal of Philosophy of Education* 55, no. 1 (February 2021): 143–66.

Mason, Andrew. "Fair Equality of Opportunity and Selective Secondary Schools." *Theory and Research in Education* 14, no. 3 (2016): 295–312.

Mazie, Steven. "Equality, Race and Gifted Education: An Egalitarian Critique of Admission to NYC's Specialized High Schools." *Theory and Research in Education* 7, no. 1 (2009): 5–25.

McCrea, Simon M., John H. Mueller, and Rauno K. Parrila. "Quantitative Analyses of Schooling Effects on Executive Function in Young Children." *Child Neuropsychology* 5, no. 4 (1999): 242–50.

Merry, Michael. "Educational Justice and the Gifted." *Theory and Research in Education* 6, no. 1 (2008): 47–70.

Meyer, Kirsten. "Educational Justice and Talent Advancement." In *Education, Justice, and the Human Good: Fairness and Equality in the Education System*, edited by Kirsten Meyer, 133–50. New York: Routledge, 2014.

Moore, G. E. "The Conception of Intrinsic Value." In *Philosophical Studies*, 253–75. London: Routledge and Kegan Paul, 1922.

Neisser, Ulric, Gwyneth Boodoo, Thomas J. Bouchard Jr., A. Wade Boykin, Nathan Brody, Stephen J. Ceci, John C. Loehlin, Robert Perloff, Robert J. Sternberg, and Susana Urbina. "Intelligence: Knowns and Unknowns." *American Psychologist* 52, no. 2 (1996): 77–101.

Nussbaum, Martha. *Frontiers of Justice: Disability, Nationality, Species Membership*. Cambridge, MA: Harvard University Press, 2006.

Robb, Catherine. "Talent Dispositionalism." *Synthese* 198, no. 9 (September 2021): 8085–102.

Robeyns, Ingrid. "Having Too Much." In *Wealth*, edited by Jack Knight and Melissa Schwartzberg, 1–44. New York: New York University Press, 2017.

Sapon-Shevin, Mara. *Playing Favorites: Gifted Education and the Disruption of Community.* Albany: State University of New York Press, 1994.

Sher, George. *Beyond Neutrality: Perfectionism and Politics.* Cambridge: Cambridge University Press, 1997.

Steinberg, Robert J. "'My House Is a Very Very Very Fine House' but It Is Not the Only House." In *The Scientific Study of General Intelligence: Tribute to Arthur R. Jensen*, edited by Helmuth Nyborg, 373–95. Oxford: Pergamon, 2003.

Sweatt, J. David. "The Emerging Field of Neuroepigenetics." *Neuron* 30, no. 3 (October 2013): 624–32.

Swift, Adam. *How Not to Be a Hypocrite: School Choice for the Morally Perplexed Parent*. London: Routledge, 2003.

Sypnowich, Christine. *Equality Renewed: Justice, Flourishing and the Egalitarian Ideal*. London: Routledge, 2017.

———. "A New Approach to Equality." In *Political Neutrality: A Reevaluation*, edited by Daniel Winstock and Roberto Merrill, 178–209. London: Palgrave, 2014.

Terzi, Laurella. "On Educational Excellence." *Philosophical Inquiry in Education* 27, no. 2 (2020): 92–105.

Thompson, Winston C. "A Limited Defense of Talent as a Criterion for Access to Educational Opportunities." *Educational Philosophy and Theory* 8 (2020): 833–45.

White, John. "Illusory Intelligences?" *Journal of Philosophy of Education* 42, nos. 3–4 (August/November 2008): 611–30.

Journal of Ethics and Social Philosophy
vol. 28, no. 3 · october 2024

https://doi.org/10.26556/jesp.v28i3.3691

RESCUE CASES, THE MAJORITY RULE, AND THE GREATEST NUMBER

Jonas Werner

In a recent paper, Tim Henning argues that the conclusion that we should save the greatest number in rescue cases can be established on procedural grounds without making use of the aggregation of interests. He first argues that we ought to respect the affected person's equal claims to have a say in the rescue decision and that this can be achieved only by the majority rule, which consists in giving each affected person an equal vote. Then he argues for the second claim that if everyone votes in their self-interest, then the greatest number will be saved. I present a class of cases in which the second claim fails. This establishes that even if self-interested voting is assumed, the majority rule does not always lead to the greatest number being saved.

1. Majority without Numbers

The claim that I will dispute in this note is that if we use the majority rule in rescue cases, then "in cases where each votes in their own self-interest, respect for their equal right to decide, or their autonomy, will lead us to save the greater number."[1] This section presents a rescue case in which this claim fails. The second section presents a variant of this case (which, other than the one discussed in this section, does not involve any probabilistic element) and discusses potential ways to weaken the claim that use of the majority rule leads to saving the greatest number.

Following Henning, I understand the majority rule as "a decision procedure that selects an option only if it receives at least as many votes from a relevant electorate of affected persons as any other option on the table" (758).[2] For simplicity, I will exclusively consider rescue cases in which two or more boats are about to sink, and the rescuer can save the passengers of at most one boat. An example might help to understand Henning's claim:

[1] Henning, "Numbers without Aggregation," 755 (hereafter cited parenthetically).

[2] See Novak, "Majority Rule," for a general discussion of the majority rule.

> *Base Case*: Let there be boats B_1 and B_2. Every boat is about to sink, and those on the boats will die if you don't rescue them. You can rescue the passengers of at most one boat. There are two persons on B_1, and there is one person on B_2.

Assume for the sake of argument that it is established that we should let majority rule determine which option we choose. The options are rescuing boat B_1 and rescuing boat B_2. For any passenger, to vote in their self-interest is to vote for the boat on which they are being rescued. Consequently, if everyone votes in their self-interest, then B_1 will be rescued. Rescuing B_1 is tantamount to saving the greatest number, namely, two persons instead of only one. The base case is in accordance with Henning's claim.

However, Henning's claim is false in the following case:

> *First Problem Case:* Let there be boats B_1, B_2, and B_3. Every boat is about to sink, and those on the boats will die if you don't rescue them. You can rescue the passengers of at most one boat. There are three persons on B_1, and there are two persons on each of B_2 and B_3. You can either go to B_1 and rescue it or steer your rescue boat into the fog, which results in a probabilistic process that gives each of B_2 and B_3 a 50 percent chance of being the boat you reach and rescue.

I assume that only options that can be ensured to obtain by you can be voted for. Of course, the passengers of B_2 hope that you will steer into the fog and that the probabilistic process leads to you rescuing B_2. However, there is nothing you (or anyone else) can do to ensure that the probabilistic process will yield a certain outcome. It seems absurd to allow people to cast votes for options that are such that no one can bring the option about. For this reason, I assume that in this case, the available options between which the passengers should vote are rescuing B_1 and steering the boat into the fog.

I will also make the assumption that in ordinary rescue cases, for an affected person to vote in their self-interest is for them to vote for the option that maximizes the chance that they are rescued. I take examples of extraordinary rescue cases to be cases where the passengers of some boats are meaningfully related to the passengers of other boats (e.g., their children or spouses), cases where some passengers have no interest in continuing to live, or cases where some options involve being rescued for extremely high costs (like the death of close relatives). Henning is explicit that in such cases, the majority rule might not lead to the greatest number being saved, so it is dialectically safe to set them to one side.[3]

[3] See Henning's case of Unanimous Choice and the discussion of his claim that "it is a *justified default assumption* that people want us to save the group to which they belong" (767).

With these assumptions in place, we can argue as follows. By the first assumption, you rescuing B_1 and you steering into the fog are the two options that can be voted for.[4] For the four passengers that are on one of B_2 and B_3, the option of you steering the boat into the fog maximizes their chance of survival (for it gives each of them a 50 percent chance of survival while the only other option available gives each of them no chance of survival). By the second assumption, if every one of the four passengers that are on one of B_2 and B_3 votes in their self-interest, every one of them votes for you steering the boat into the fog. The four passengers that are on one of B_2 and B_3 have a majority. The option the majority will vote for if everyone votes in their self-interest is such that there is an alternative option that leads to a greater number being saved. Therefore, in the First Problem Case, it is not the case that if everyone votes in their self-interest, then the greatest number will be saved.

One potential response, which I owe to an anonymous referee, is to discount the votes of those on B_2 and B_3. The persons on B_2 and B_3 might be seen as having a weaker claim than those on B_1, for your decision to steer into the fog would only give them a 50 percent chance of survival, while your decision to rescue B_1 would guarantee the survival of its passengers. While section 9 of Henning's paper only discusses discounting the votes of those who face lesser harms, some might hold that a lesser chance to avoid an equally severe harm should also lead to a discount. I am unsure whether discounting in the given case is plausible, for it seems that the plausibility of discounting is underwritten by the thought that those who have less to lose should have less of a say. This feature gets lost if we extend it to cases where everyone's life is at stake and having a lesser chance of being saved is what leads to a discount. In any case, the next section will present a further problem case in which the discounting response is not available.

2. Discussion

The First Problem Case might look like a counterexample that can be taken care of by insisting on the discounting response or by a slight weakening of the claim under discussion. One option for a weakening is to restrict the claim that if everyone votes in their self-interest, then the greatest number will be saved. This claim could be restricted to cases in which every affected person can vote for an option that guarantees that they will be rescued. This, one might

[4] I ignore the options that consist in holding lotteries between the two basic options. Clearly, holding a nontrivial lottery between the two basic options is not maximizing the chance of survival for those on boats B_2 and B_3. Given that they together already have a majority, we can assume that they do not vote for such a lottery if they are self-interested voters.

hope, takes care of the probabilistic element that creates problems in the First Problem Case.

It should be noted that this weakening would already undermine Henning's argumentation to a considerable extent, unless the weakening could be independently motivated (i.e., motivated without overtly or tacitly relying on the claim that the greatest number should be saved). Furthermore, it will not do, as the following case shows:

> *Second Problem Case:* Let there be boats B_1, B_2, and B_3. Every boat is about to sink, and those on the boats will die if you don't rescue them. You can rescue the passengers of at most one boat. There are three persons on B_1, and there are two persons on each of B_2 and B_3. The passengers can't communicate, but they are able to cast votes, and they know about the options available to you and that you will get the result and use the majority rule to decide what to do.

I claim that, given that their aim is to maximize their chances of surviving, it is instrumentally rational for the passengers on B_2 and B_3 to vote in favor of you holding a lottery that gives each of the passengers of B_2 and B_3 a 50 percent chance to be rescued. To see why I hold that this voting behavior is rational, take the perspective of the passengers of B_2. (The situation of those on B_3 is exactly analogous.) Casting a vote for B_2 being rescued (without any lottery) is irrational for them, for they cannot expect that there will be a majority for this. Given that they cannot communicate with the passengers of B_1, it is impossible to form an alliance with them. They have no reason to expect that the passengers of B_1 will vote for anything that differentiates between the passengers of B_2 and B_3. The setup treats B_2 and B_3 exactly alike. Therefore, no solution that creates any asymmetry between them is plausible to find a majority (without communication between the boats). The only options that treat B_2 and B_3 perfectly alike while giving each of their passengers a chance to survive is to hold a lottery that gives each of the passengers of B_2 and B_3 an equal chance to be rescued. Given that the passengers of B_2 and B_3 together have a majority, they have no reason to care about the chances of those who are not in a situation that is symmetrical to their own. So the passengers of B_2 should cast votes for a lottery that gives each of the passengers of B_2 and B_3 a 50 percent chance to be rescued, hoping that their counterparts on B_3 have similar thought processes.

Why did I stipulate that the passengers can't communicate with those on other boats? *Prima facie* one might think that the assumed voting behavior becomes more realistic if the passengers of B_2 and B_3 can share their thoughts. However, if they can also communicate with those on B_1, then the three passengers of B_1 will also try to make offers. The resulting situation is unstable insofar as the passengers of any two boats have an absolute majority (a majority of more than half of the electorate), and the passengers of no boat have an absolute majority on their own. Whatever the ensuing discussions would result in, in real-life cases, it is far from clear that the result would lead to those on B_1 being rescued.[5]

Revisiting the discounting response shows that it does not work in the second problem case: in this case, every person has the option to vote for a procedure that guarantees their survival. Discounting the passengers' votes only if they do not exercise this option amounts to letting the strengths of their votes depend on what they are voting for. This seems to be an implausible response that is hard to square with the idea that in majority decisions, the affected persons can autonomously decide on how to use their votes.

Of course, one could further restrict the claim that if everyone votes in their self-interest, then the greatest number will be saved. One might ban voting for lotteries. However, Henning himself explicitly claims that you ought to respect if affected persons vote for lotteries, even if you take doing so a "tragic mistake" (761). Another option would be to ban voting for lotteries *for strategic reasons* (i.e., to secure a majority).

Taking this further weakening into account, we arrive at the following claim. If a rescue case (1) involves no probabilistic element, (2) every affected person votes for the option they would vote for if they could dictatorially decide the outcome, and (3) everyone votes in their self-interest, then the majority rule guarantees that the greatest number will be saved.

I did not present any reasons to doubt this weaker claim. Restriction 1 takes care of probabilistic setups like the First Problem Case, and restriction 2 rules out strategic alliances like in the Second Problem Case. It might after all be the principle Henning has in mind.[6] When making his argument explicit, he uses the following premise:

> P4. If each affected person votes for the option in which she herself is saved, and if we let majority rule determine the option we realize,

[5] The passengers on B_1 might have a slightly more comfortable position, for they have a relative majority if no two boats are such that their passengers manage to form an alliance and agree to vote for the same option. Still, it remains the case that there is a possible alliance against them that has an absolute majority.

[6] Henning's dialectics against lottery voting (i.e., randomly drawing one of the cast votes) in sec. 8 of his paper speaks against the suspicion that he assumes 2. There he argues that we should allow voting for probability distributions. Then he shows that voting for probability distributions would in some cases of lottery voting give strategic voters the power to dictatorially influence the overall probability distribution (given the reduction of compound lotteries).

> then we will realize an option that saves at least as many people as any other option. (759)

This premise is applicable only to cases where each affected person has the option to vote for themself being saved (without any probabilistic element, so we might suppose). The further explication of Henning's argument consists in concluding from P4 and the intermediate conclusion that "in rescue cases we should let majority rule determine which option we realize" (759) the following final conclusion:

> C2. Thus, if each affected person favors the option in which she herself is saved and we follow the morally required procedure, then we will realize an option that saves as many people as any other option. (759)

The slight change in formulation from "votes for" to "favors" seems to indicate that Henning is either unaware of the possibility that affected persons rationally do not vote for the option they favor (to secure a majority) or that he wishes to tacitly preclude it.[7] However, it seems worthwhile to point out that the class of cases in which the majority rule leads to the greatest number being saved is severely restricted. The cases discussed in this note show that the proponent of the majority rule sometimes has to decide between ignoring majority votes and not saving the greatest number. The connection between numbers and majorities is hence not as close as one might have hoped.

One potential reaction is to discuss whether the restrictions needed to secure a tight relation between the majority rule and saving the greatest numbers can be independently motivated. A closer look at the ways in which the majority rule can be philosophically justified might help. Two prominent philosophical ideas in this respect are that the majority rule is justified because it encodes equal respect to the members of the electorate and that it is justified because it fosters the autonomy of the affected persons.[8] One might, for example, try to argue for a theory of autonomy that supports the majority rule only in the restricted class of cases in which it leads to the greatest number being saved. Whether this project can be successfully carried out is a question that will be left for another occasion.

A more steadfast reaction is to maintain that the majority rule tells us what we ought to do in rescue cases and to accept that we hence ought to save the

[7] Note that the argument is formally invalid if "favors" and "votes for" are not treated here as synonymous.

[8] The former idea can be found in Waldron, *The Dignity of Legislation*. The latter idea is alluded to in sec. 4 of Henning's paper and can also be found in Kelsen, "On the Essence and Value of Democracy."

greatest number only in a restricted class of cases. It should be noted that the Second Problem Case can be modified (by raising the number of boats with few passengers that are in pairwise symmetrical situations and raising the number of people on the remaining boat) to make the majority rule lead to drastic cases of failures to rescue the greatest number. A potential defense of the steadfast response that goes beyond a general defense of the majority rule might consist in arguing that the features that lead to a disconnect between the number rule and the majority rule (like, e.g., options for strategic voting) are morally relevant for independent reasons.[9]

Massachusetts Institute of Technology
jonas_w@mit.edu

REFERENCES

Henning, Tim. "Numbers without Aggregation." *Noûs* 58, no. 3 (September 2023): 755–77.

Kelsen, Hans. "On the Essence and Value of Democracy." In *Weimar: A Jurisprudence of Crisis*, edited by Arthur Jacobson and Bernhard Schlink, 84–109. Oakland: University of California Press, 2000.

Novak, Stéphanie. "Majority Rule." *Philosophy Compass* 9, no. 10 (October 2014): 681–88.

Waldron, Jeremy. *The Dignity of Legislation*. New York: Cambridge University Press, 1999.

[9] I wish to thank two anonymous reviewers for their comments, which led to significant improvements of this paper. Work on this paper was supported by the Swiss National Science Foundation, grant number P500PH_214177.