From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Philosophy

2018

Practical Wisdom & Meta-Normative Reflection: Essays on Moral and Legal Normativit

Atiq, Emad

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01s1784p44v

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

PRACTICAL WISDOM & META-NORMATIVE REFLECTION

Essays on Moral and Legal Normativity

Emad Hanzala Atiq

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE BY

THE DEPARTMENT OF PHILOSOPHY

Advisers: Michael Andrew Smith & Gideon Rosen

June 2018

© Copyright by Emad Hanzala Atiq, 2018. All rights reserved.

Abstract

This dissertation's aim is to secure meta-normative foundations for ethics and legal theory. It consists of seven self-contained but interconnected papers. The first four form the first part of the dissertation, entitled "Meta-Normative Foundations for Ethics." These essays raise and address challenges facing a class of related meta-normative theories. The defensive maneuvers are all in the service of vindicating an original meta-normative theory for understanding moral thought and talk: Quasi-Realist Contextualist Expressivism, described in some detail in the preface. The second part, entitled "Meta-Normative Foundations for Law," consists of three essays that apply the results of meta-normative investigation to practical legal problems. The aim is show that general meta-normative theorizing has a substantial practical payoff: it is crucial for understanding legal rules in American common and constitutional law.

In chapter 1, "The Hard Problem of Supervenience," I raise and address a challenge facing expressivists: that of explaining why normative properties supervene on repeatable non-normative properties.

In chapter 2, "How to be Impartial as a Subjectivist," I explain why expressivism and related views do not have unacceptable ethical implications concerning moral disagreement.

In chapter 3, “On Ground as a Guide to Realism,” I raise problems for the dominant view on what distinguishes quasi-realism about normative properties from robust realism, according to which “real” properties play a certain explanatory role.

In chapter 4, “The Normative Distinction between Quasi-Realism & Realism,” I argue that the difference between realism and quasi-realism can be characterized in normative terms.

In chapter 5, “Juridical Obligations at the Edge of Legality,” I argue that the nature of legal normative claims entails that judges have no legal obligation to follow the law in all cases.

In chapter 6, “Legal vs. Factual Normative Question,” I argue that the difference between convention-dependent and convention-independent normative facts entails a difference in how judges should treat normative questions that arise under common law.

In chapter 7, “The Limits of Law in the Evaluation of Mitigating Evidence,” developed together with Erin Miller, we argue that the difference between legal and moral decision-making is a matter of constitutional significance.

Table of Contents

Preface vii
Acknowledgments xvii
Part 1. Meta-Normative Foundations for Ethics 1
Chapter 1. The Hard Problem of Supervenience 1
Introduction 2
I. "The Least Controversial Thesis in Metaethics" 6
II. "The Expressivist Side-Step" 16
III. The Hard Problem for Plan-based Expressivism 21
IV. A General Challenge 30
V. Ways of Biting the Bullet 33
Chapter 2. How to be Impartial as a Subjectivist 40
Introduction 41
I. The Argument from Disagreement 43
II. The Nature of Impartiality 48
III. Why Alternative Responses are Unsatisfying 63
IV. Objections and Replies 69
Chapter 3. On Ground as a Guide to Realism 80
I. A Finean Principle of Ground 80
II. The Grounding Principle Should be Rejected 82
III. The 'No Ground-Essence Mediation' Reply 88
IV. Does the Principle Remain Theoretically Useful? 92
Chapter 4. The Normative Distinction between Moral Realism & Quasi-Realism 99
Introduction 99
I. The Distinguishing Claim: REALIST ANGST 100
II. Quasi-Realist Accounts of Realist Angst 102
III. Realist Accounts of REALIST ANGST 106
IV. Objections & Replies 108
Conclusion 111
Part 2. Meta-Normative Foundations for Law 114
Chapter 5. Juridical Obligations at the Edge of Legality 114
Introduction 115
I. Reasonable & Persistent Disagreement About the Concept of Law 119
II. Competing Theories of the Concept Diverge Extensionally in Hard Cases 125
III. Judges Have Legal & Moral Permission to Ignore Legality in Hard Cases 132
IV. Objections and Replies 152
V. A Reply to Dworkin 155
Conclusion 161
Chapter 6. Legal vs. Factual Normative Questions under the Common Law 162
Introduction 164

I. How Courts Deal with Normativity Under the Law/Fact Distinction .. 175
II. A Framework for the Classification of Normative Questions under the Law/Fact Distinction 198
III. Judicial Sensitivity to Conventional and Pre-Conventional Norms ... 212
IV. A Framework for Sixth Amendment Jurisprudence: Does the "Life-to-death" Judicial Override Infringe on the Jury's Fact-finding Responsibility? 246
Conclusion 255
Chapter 7. The Limits of Law in the Evaluation of Mitigating Evidence 258
Introduction 260
I. A Troubling Case of Restrictive Consideration: The Causal Nexus Requirement for SED 267
II. Unrestricted Consideration of Deprivation Evidence: Theory and Practice 280
III. The Constitutional Argument Against Restrictive Consideration of Deprivation Evidence 295
Conclusion 316

Preface

It is controversial whether and to what extent metaphysical enquiry into the nature of normative thought and talk bears on first-order normative enquiry—that is, enquiry into what we ought to do or think. This dissertation represents a qualified defense of the view that normative enquiry needs and benefits from meta-normative reflection.

The dissertation is comprised of seven chapters, grouped into two parts. The chapters are written as independent papers and can be read in any order, with the exception of Chapter Four. The first part of the dissertation, entitled "Meta-Normative Foundations for Ethics," spans Chapters One through Four. These chapters focus on the metaphysical foundations for normativity, in general, and moral thought and talk, in particular. Each essay raises and addresses a challenge confronting a class of related meta-normative theories. The essays are primarily defensive, and the various defensive maneuvers amount to an indirect defense of an original meta-normative theory—QUASI-REALIST CONTEXTUALIST EXPRESSIVISM ("QCE")—further described below.

The second part of the dissertation, entitled "Meta-Normative Foundations for Law," spans chapters Four through Seven, with the content of the final chapter having been developed together with Erin Miller. These chapters apply the results of meta-normative investigation to practical legal problems. With QCE providing the background metaphysical framework for understanding normative claims, these final three essays interpret prominent legal rules in American common law and constitutional law in light of meta-normative facts: among other things, the nature of legal normative claims; and

the difference between convention-dependent and convention-independent normative facts. These final chapters have sections devoted to legal theory and analysis, especially the final two.

Since the vindication of QCE is the animating principle behind the first part of the dissertation and presupposed by the second, and yet no positive arguments for this controversial position are offered elsewhere in the dissertation, I offer, here, some of my reasons for being attracted to the view. These reasons that favor QCE have not been recognized in the literature and will be the subject of future work. While the present discussion is not intended to persuade anyone, I hope to say enough to motivate interest in my project: an investigation into the view's viability and normative consequences.

Here is a first-pass characterization of the view's defining commitments, beginning with expressivism:

EXPRESSIVISM	Normative terms, like "good" and "ought" are used to express the speaker's motivational, emotional, and other non-cognitive attitudes. E.g. to assert that one ought to help the poor *just is* to express a desire or plan to help the poor.

Full-blown expressivists insist that normative terms always play this non-cognitive attitude expressive role. Contextualist expressivists deny this. According to contextualists, the semantic contribution of normative terms shifts depending on the sentential context. The view with its commitment to relational meaning is analogous to a Fregean view of semantic content. Unlike Hybrid-Expressivists who believe that normative terms in *every* context have an expressive *and* descriptive role, Contextualist-Expressivists think the two roles come apart. In certain contexts, assertions involving

normative terms are just ordinary descriptive statements. E.g., when I assert Tom believes lying is wrong, I am reporting that Tom accepts the proposition that lying is wrong.

CONTEXTUALISM — Normative assertions *in the paradigm case* amount to the expression of a non-cognitive attitude. However, normative language can be embedded in sentential contexts that render the overall assertion purely world-descriptive. What is being described can be understood within the general expressivist framework—it does not involve essential reference to irreducibly normative entities, except in a thin sense (see quasi-realism below).

The final defining commitment of QCE is quasi-realism, a deflationary view of facthood, truth, properties, and propositions.

QUASI-REALISM — Assertions of the form '*it is a fact that* one ought to help the poor' and '*it is true that* one ought to help the poor' are equivalent to (have the same content as) the assertion that one ought to help the poor. Adding 'it is a fact that' or 'it is true that' to a sentence contributes nothing to its sense. Property- and proposition-talk receive similar deflationary treatment.

These very rough characterizations of the defining commitments of QCE stand in need of much refinement. But they are close enough to the truth for us to be able to fruitfully discuss the motivations behind the overall package.

Why EXPRESSIVISM? The rules governing normative thought and talk are easily mastered. It takes very little for individuals, even at a very young age, to become

competent users of terms like "good" and "should." By competent user I mean a user who has mastered the rules governing meaningful usage of the terms. At the same time, it seems most individuals are wildly mistaken about which things are good and about what they ought to do. It is very tempting to think that such individuals are *wrong* about the good—that is, they are apt targets for criticism. This is not the 'Moral Twin Earth' intuition that persons who apply "good" to very different things than we do genuinely disagree with us about a common subject matter. Traditionally, expressivism has been motivated by Moral Twin Earth type cases and the idea that a meta-normative theory needs to secure the possibility of disagreement in such cases. I think the focus on disagreement has obscured the more compelling reason for embracing EXPRESSIVISM, which is that it renders possible a straightforward vindication of criticism directed at individuals, on Moral Twin Earth or otherwise, who have divergent views about the good. Expressivists can explain both the easy mastery of normative language and the criticizability of users of normative language. The expressivist's account of the rules governing meaningful application of the goodness predicate reveals why the rules are easily mastered. All it takes to meaningfully use normative vocabulary is to have motivations and to see the point in communicating one's motivations to others. On the expressivist picture, there is similarly no mystery concerning why people are mistaken about what ought to be done and why they are often appropriately criticized. To say that *Tom is wrong* about the good is just to direct a kind of negative non-cognitive attitude towards Tom and his motivations. The appropriateness of this non-cognitive attitude is entirely a function of its moral permissibility, which in many cases is easily secured. Descriptivist views – both of the naturalist and non-naturalist variety – that aspire to

provide extensionally adequate accounts of normative terms struggle with explaining why the rules governing meaningful use of such terms are easily mastered and yet normative truth remains elusive and renders agents open to critique. Descriptivists think the judgment that Tom is wrong about the good involves a description of the world. In order to vindicate the critical judgment that Tom is wrong, descriptivists need to show that the relevant description is accurate, which is far from straightforward. For the expressivist, since the critical stance is entirely non-cognitive, it does not raise the same alethic questions of warrant.

Another reason for favoring EXPRESSIVISM is that it, unlike standard descriptivist views, secures the possibility not just of justified true beliefs in ethics but that of ethical *certainty* based on ordinary ethical reasoning. We are often warranted based on ordinary ethical reasoning in being maximally confident in our moral judgments. We can be maximally certain, e.g., that it is not OK to kick puppies for fun. Ordinary ethical confidence presents standard descriptivist views in metaethics with a challenge distinct from that of providing a general epistemology for ethics. The challenge is to explain why ordinary ethical confidence remains justified once one accepts a descriptivist metaethics. Mainstream descriptivist views entail that moralizers take on (and overlook) a substantial risk of error when they make ethical judgments. This is because the *truth-maker* for the judgment "It is not OK to kick puppies for fun" turns out to be an esoteric proposition about which agents should be considerably uncertain in ordinary circumstances. For example, according to a popular version of analytic naturalism, what makes it the case that it is not OK to kick puppies is that *a counterfactual version of agents with consistent desires and no false non-normative beliefs would desire that one not kick puppies* (T_1). On a popular

version of synthetic naturalism, the truth-maker for the judgment is the fact that *the act of puppy-kicking has whichever natural property it is that best satisfies a functional role specified by moral thought and talk in general* (T_2). According to standard non-naturalism, the truth-maker is the fact that *the act of puppy-kicking has a sui-generis property that is non-natural—neither invoked by the natural sciences nor characterizable in terms of natural properties* (T_3). T_1, T_2, and, T_3, even if true propositions, are ones we cannot, in ordinary contexts, have maximal credence in. The propositions are at best non-obvious. Furthermore, consideration of their truth is entirely bypassed in ordinary ethical reasoning. Based on plausible principles governing credences, it is easy to show that a moralizer with maximal credence in the truth of some such descriptivist metaethics should be considerably uncertain about her ordinary ethical judgments. There are descriptivist theories that avoid this problem. But the most viable ones undermine ordinary ethical confidence.

Ordinary ethical confidence presents less of a challenge to the expressivist. According to the expressivist, the judgment "It is not OK to kick puppies for fun" does not involve a description of the world. It involves the expression of a desire-like attitude—e.g., a desire not to kick puppies. The judgment that this expression of desire is warranted is itself an ethical judgment: again, the expression of a desire. Accordingly, the metaethics does not entail any novel truth-related questions of warrant that might be overlooked in ordinary ethical reasoning. Expressivism, by design, is entirely compatible with maximal confidence in one's first-order judgments about puppy-kicking based on whatever it is that ordinary ethical reasoning takes to be epistemically sufficient. Expressivist quietism concerning when and why ethical judgments are warranted is usually portrayed as a vice when it comes to making sense of ethical uncertainty. It has

gone unnoticed that this feature of expressivism turns out to be a virtue in relation to ethical certainty. The argument from ordinary ethical confidence suggests a new way of motivating expressivism that has several advantages over traditional arguments based on Moral Twin Earth thought experiments, motivation-internalism, and the open question argument.

Why CONTEXTUALISM? The context-sensitivity of terms in ordinary language is well-motivated. Expressivists about normative language should recognize that there are opaque or oblique sentential contexts where normative terms do not make their usual contribution to the meaning of a sentence. E.g., when I assert that Tom thinks lying is wrong I am not expressing my own motivations or trying to express Tom's motivations. Rather, I am simply reporting which normative proposition Tom accepts. Trying to explain opaque usage of normative vocabulary based on the motivation expressive account robs expressivism of its primary virtue—its simplicity and intuitiveness. It is better to treat these opaque contexts as generating ordinary descriptions. CONTEXTUALISM plays an important role in the second part of the dissertation. It suggests a helpful reconstrual of the debate between legal positivists and legal anti-positivists about the nature of claims of the form "there is a legal obligation to pay taxes." Positivists famously claim that such assertions amount to specifications of what we do around here—which normative rules are conventionally followed, while anti-positivists insist that such claims partly involve ordinary moral assertions the content of which is somehow fixed by our conventions. CONTEXTUALISM suggests a descriptivist understanding of positivist claims about law and an expressivist understanding of anti-positivism. According to positivists, talk about what we legally ought to do is descriptive

of the normative assertions that are conventionally embraced or accepted by the relevant legal officials. According to anti-positivists, talk of what we legally ought to do necessarily involves the expression of some pro-attitude towards the prescribed action, where the reasons for an agent's being 'for' the action might have something to do with our conventions. This is a controversial characterization of the debate, but there is much to be said in its favor.

Finally, why QUASI-REALISM? To clarify, the quasi-realism I embrace is *presumptive* rather than dogmatic, which is to say it takes ordinary terms like "existence," "facthood," and "truth" as amenable to deflationary treatment but remains open-minded about the possibility of the introduction of technical 'hard' or non-deflationary metaphysical concepts. It is, accordingly, a very weak thesis. A quasi-realist in my sense needn't be resistant to traditional metaphysics. She simply places the burden on metaphysicians who insist on using terms like 'fact' in a non-deflationary sense to explain what they mean and insists on the right to apply the ordinary concept of a fact in her preferred way (because ordinary usage is most compatible with deflationism). Relatedly, the overall meta-normative theory—QCE—is compatible with the introduction of technical vocabulary that resembles ordinary normative vocabulary while being entirely descriptive. Naturalists and non-naturalists alike are entitled to technical concepts of goodness* and ought* so long as the role of such concepts is well-defined; and we should be interested in such concepts insofar as they track a real and important subject matter.

Part 1 of the dissertation, entitled: "Meta-normative Foundations for Ethics," begins with Chapter 1: "The Hard Problem of Supervenience." In this chapter, I argue

that the fact that the normative supervenes on the non-normative has been misunderstood, resulting in sham or incomplete explanations of the relevant phenomenon. Normative supervenience, properly understood, poses a significant explanatory challenge even for expressivists. I suggest, ultimately, that the relevant supervenience thesis is best understood by expressivists as a normative truth rather than a conceptual one. Chapter 2: "How to be Impartial as a Subjectivist," defends expressivism and related views from an objection of David Enoch's: namely that such views make it impossible to be impartial while standing one's ground in moral disagreements. I disarm Enoch's objection with an original account of what impartiality consists in. While Enoch's argument is principally intended as a *reductio* of subjectivism, it easily generalizes to threaten even the expressivist. My response on behalf of the subjectivist is one the expressivist can also help herself to. Chapter 3: "On Ground as a Guide to Realism," explores what distinguishes full-blooded realists from quasi-realists about ethical features. I raise problems for Kit Fine's criterion for realism, which distinguishes "real" features by assigning them a distinctive explanatory role. I argue that Fine's approach, which has proven influential in metaethics, presupposes a false principle of explanation and should be rejected. Finally, Chapter 4: "The Normative Distinction between Quasi-Realism & Robust Realism," explores a different means of distinguishing quasi-realism from robust realism: based on whether a view acknowledges certain normative assertions—a conditional claim with a meta-normative antecedent and normative consequent—as meaningful.

Part 2 of the dissertation is entitled: "Meta-normative Foundations for Law," and it begins with Chapter 5: "Juridical Obligations at the Edge of Legality," which argues

based on ecumenical assumptions about the nature of legal facts that judges have little to no legal obligation to follow a rule based on its legality when its legality, though genuine, is sufficiently uncertain. Chapter 6: “Legal vs. Factual Normative Questions,” argues that the difference between convention-dependent and convention-independent normative facts entails a difference in how normative questions arising in contract and tort law should be treated under a well-known common law distinction. The final chapter, Chapter 7: “The Limits of Law in the Evaluation of Mitigating Evidence,” is co-authored with Erin Miller. It offers an account of what it means for judges to follow a rule out of a sense of legal (rather than moral) obligation emphasizing law's distinctive conventionality. We use this account of legally motivated decision-making to argue that whenever the law explicitly demands moral reflection from judges, as in the evaluation of mitigating evidence during criminal sentencing, legally motivated decision-making violates the law.

Acknowledgments

My interest in meta-normative theory was sparked by the threat of error theory in ethics. At first, this threat took the form of the problems of personal identity and free will, to which I was introduced as an undergraduate by Gideon Rosen in 2007. Gideon's presentation of the problems with his characteristic magnetism left me wondering about the ethical significance of metaphysics; and it altered my life trajectory. I hope he will forgive me for all the demands I have made on his time since. Had it not been for Gideon's intellectual mentorship, support, and generosity with his time, this dissertation would have been far less than what it is.

I thank and credit Michael Smith for drawing my attention to a more general threat to ethics which would become the central theme of my graduate work: the possibility that every normative claim involves a mistaken thought. It was Michael's seminars in metaethics, his patient critique and challenge, that prompted me to sort out my own metaethical views. The non-cognitivism I am drawn to is motivated by an ambition fostered in me by Michael of securing the possibility not just of justified true beliefs in ethics and related domains, but that of ethical certainty based on ordinary ethical reasoning.

In addition to Gideon & Michael, I was supervised in different capacities by Sarah McGrath and Philip Pettit. I am extraordinarily fortunate to have been taught by them. The papers that comprise this dissertation benefited especially from conversations with and written feedback from Sarah and Philip. Their engagement with my work was consistently helpful and has meant a lot to me. Several other persons played a significant

role at critical stages of my academic career to date, through their sponsorship and support. I thank them in the order in which they appeared in my life: Tom Kelly, Simon Blackburn, Sebastian Köhler, Joshua Knobe, Ian Ayres, Daniel Markovits, Gideon Yaffe, Patrick Weil, Jo Kelly, and Johann Frick.

To the extent that this dissertation contains original ideas, they were inspired by and refined through conversation with my peers. I owe Nat Tabris, in particular, a major debt of gratitude for being just as ready to discuss with me the problem of universals as American constitutional law. My papers bear the mark of Alisabeth Ayars, who prepared me for the winter of workshops and job-talks with her persistent demands for clarity and precision. I am indebted also to Adam Lerner, Sam Preston, and Eric Hubble for their contributions to my work.

Erin Miller, with whom I co-authored Chapter 7, was my most significant interlocutor through graduate school. It would be very hard to overestimate her influence on my work. The published papers to come out of this dissertation would not have made it past the initial review stage had it not been for Erin's multiple "revise and resubmit" orders prior to submission. I am grateful most of all to her for her insights, her criticisms, her kindness, and her love. I regret all the ways in which I was a burden to her.

Finally, I thank my family for their love and support over the years. I am very fortunate to have had not one, but two sets of parents instill in me respect for scholarly toil.

To my grandparents.

Part 1. Meta-Normative Foundations for Ethics

Chapter 1. The Hard Problem of Supervenience

Abstract. I argue that the conceptual truth that the normative supervenes on the non-normative has been misunderstood, resulting in incomplete explanations for why it holds. Normative supervenience, correctly understood, poses a significant explanatory challenge even for expressivists, who have long claimed the ability to explain supervenience a primary virtue of expressivism. The conceptual truth that demands explanation is the supervenience of the normative on the *repeatably* non-normative. After rendering the repeatability constraint precise, I show that expressivism is inconsistent with supervenience so defined. This result is derived using the expressivist's preferred semantics for normative terms, together with constitutive facts about the non-cognitive attitudes essentially involved in normative thought and talk. I suggest, ultimately, that the problem is not unique to expressivism: our usage of normative terms seems disciplined by repeatable supervenience, but it turns out to be very hard to say why.

.

Introduction

Almost everyone writing in metaethics agrees that the normative supervenes on the non-normative.[1] Two things identical in their non-normative features cannot differ in their normative features. Moreover, most philosophers who accept normative supervenience agree that normative supervenience holds as a matter of *conceptual* necessity. Anyone competent with our concept of the normative should see that normative properties cannot vary without some non-normative property varying.

Any metaethical theory that accepts this starting point must explain why normative supervenience holds as a matter of conceptual necessity. There is some question whether metaethical non-naturalists can meet this burden; but most metaethical theories have a credible account.[2] Expressivists and reductive naturalists, in particular, offer elegant explanations for the conceptual truth of normative supervenience.[3] Given this, we can call the challenge of explaining why normative properties supervene on non-normative properties the Easy Problem of Supervenience.

There is a harder problem—the Hard Problem of Supervenience—that demands as much attention. It is the problem of explaining the conceptual truth that normative properties supervene on *repeatable* non-normative properties. Roughly, non-normative properties are repeatable when it is possible for numerically distinct individuals to share

[1] There are a few contrarians. See e.g. Rosen (MS); Fine (2002).
[2] Non-naturalist realists are supposed to have a hard time explaining supervenience. See discussion in McPherson (2012) and *infra* Section II.
[3] See discussion *infra* Section II.

them. The conceptual truth to be explained is that if two things are alike in all repeatable non-normative respects, they are alike in every normative respect. I argue that the intuitions that support ordinary supervenience militate just as strongly in favor of repeatable supervenience. But the two have been conflated in recent discussion, with explanations for ordinary (or unrestricted) supervenience being taken as complete explanations for everything that needed explaining. The problem of explaining repeatable supervenience is hard even for expressivists and reductive naturalists, for whom supervenience was meant to pose no challenge at all.[4] Partial as I am to expressivism, I will be chiefly concerned in what follows with the problem as it arises for the expressivist.

The problem is this: because repeatable supervenience is a conceptual truth, expressivist semantics for normative terms should entail that a normative judgment inconsistent with repeatable supervenience is incoherent. The judgment's coherence should be ruled out by the correct account of *what it is* to deploy a normative concept. But the semantics don't entail this. According to the expressivist, to make a normative judgment is to express a motivational state: a controlling desire, preference, or plan, or something of the sort.[5] The incoherence of supervenience-flouting normative judgments thus needs to be explained in terms of constraints on the relevant motivational states. For example, if, *per impossibile*, the judgment that two non-normatively identical individuals are normatively different were coherent, this judgment would amount to the expression of some motivation to treat them differently. So long as this kind of motivation is impossible, this fact rules out the intelligibility of the supervenience-flouting normative

[4] See Blackburn (1971); Smith (2004).
[5] See e.g. Blackburn (1984); Gibbard (2003).

judgment. But, I argue, the motivational states we are familiar with are mostly *not* constrained by repeatable supervenience.[6] Moreover, on the most plausible versions of expressivism, the motivational states that underwrite normative thought and talk will be supervenience-flouting. So, far from explaining the conceptual truth of repeatable supervenience, expressivism seems inconsistent with the thesis that repeatable supervenience holds as a matter of conceptual necessity.[7]

Since I suspect the Hard Problem to be a general one in metaethics, I do not take this result to be devastating for the expressivist. A claim I will not defend here but do elsewhere is that there is a general puzzle confronting meta-normative theories: our usage of normative terms seems disciplined by repeatable supervenience, and it turns out to be very hard to say why. Expressivists have made a mistake in thinking that they can provide a deep explanation for this feature of our linguistic practice, and they are not the only ones who have made this mistake.

Section One clarifies the supervenience phenomenon. After stating the repeatability condition on the supervenience base precisely, I offer three arguments in support of the condition. The first is a direct argument for repeatable supervenience that appeals to what is inconceivable. The second and third appeal to the uses to which supervenience has been put in metaethics.

---

[6] I consider and reject a view that concedes that our motivational states are supervenience-flouting but insists that, given the functions of normative language, only those motivational states that respect supervenience count as normative judgments. I show that such accounts that appeal to the functions of normative language illicitly assume the fact that needed explaining. See discussion *infra* Sections III and IV.

[7] Despite some recent criticisms of the expressivist's account of supervenience (Zangwill 1997, 110–11; Sturgeon 2009, 84–87; Dreier 2015), the precise vulnerability I discuss and its general upshot have been missed altogether. These critics fail to press on points where the expressivist's explanation of supervenience seems most vulnerable. Moreover, they appear to be in the minority in expressing any skepticism whatsoever about the expressivist's ability to explain supervenience.

Section Two examines the expressivist's resources for explaining repeatable supervenience. Alan Gibbard's plan-based expressivism provides a useful starting point for the analysis. On Gibbard's view, the non-cognitive attitude that underwrites normative judgments is that of a plan. Further, Gibbard thinks the non-normative properties on which the normative supervenes is the class of empirically discernable properties, and one cannot plan to treat empirically indiscernible situations differently.

Section Three demonstrates that the Gibbardian explanation of supervenience fails so long as the class of empirically discernable properties satisfies the repeatability constraint. The explanation only seems compelling on the assumption that there cannot be numerically distinct yet empirically indiscernible situations. But it falters once this assumption is set aside because it is possible to plan to treat numerically distinct yet empirically indiscernible situations differently.

Section Four generalizes the challenge for the expressivist in various ways. I argue that the motivational states essentially involved in normative thought and talk *must* be ones that flout supervenience, on any repeatability-respecting definition of the supervenience base. The result is that it is always meaningful or intelligible, given expressivist semantics, to judge two numerically distinct states of affairs to be normatively different despite their being identical in base non-normative respects.

Section Six takes stock and considers the expressivist's options in light of the Hard Problem. One option is to modify standard expressivist semantics for normative terms by positing a brute fact about such terms—unless one is in the business of expressing motivations that are disciplined by supervenience, one does not count as using such terms intelligibly. The problem with this approach, I argue, is that it takes as brute

rather than *explain* why our use of normative terms is disciplined by supervenience.[8] It goes against the overt explanatory ambitions of expressivists.[9] A different option, the more attractive one, may be to regard supervenience as an ordinary *normative* truth, rather than a conceptual one. On either way of going expressivists must abandon their claim to a dialectical advantage over their theoretical opponents of having to bite no bullets when it comes to supervenience.

I. "The Least Controversial Thesis in Metaethics"[10]

It would be difficult to find a proposition more widely accepted amongst meta-ethicists than that if two things share all of their non-normative features then they share all of their normative features. Normative properties—whether of acts, states of affairs, or whatever else—are said to *supervene* on the non-normative properties of their bearer. According to Michael Smith (2004, p8), the truth of supervenience "is accepted by nearly everyone writing on the nature of value" and its denial would be evidence of a very basic conceptual confusion.[11] Arguments for supervenience tend to be rare given the assumption that it is self-evident. One argument—the Argument from Cases—appeals to what seems true in individual cases (we shall see a few more shortly). Consider some act,

[8] I consider and reject attempts to explain this brute fact in terms of the functions of normative language. See discussion *infra* Sections III and IV.
[9] See discussion *infra* Sections II and IV.
[10] Rosen (MS).
[11] See e.g. Smith (2004: 208); Blackburn (1985); Enoch (2011). It is easy to multiply examples of theorists across the metaethical spectrum who endorse the thesis.

say an instance of lying, that is judged to be wrong. It seems inconceivable that there could be another act of lying that is identical in *all* non-normative respects—it is told for the same purpose and in the same type of situation, has the same effects, and so on—and yet fails to be wrong. [12] Fixing the non-normative features of an act of lying seems to settle its normative features. Such cases reflect a more general truth about the character of the normative, or so the argument from cases maintains.

There are different ways of stating precisely what this general truth amounts to, but we can fix on the following as a first approximation:

> SUPERVENIENCE: $\Box\ (\forall N \in \alpha)(\forall x)[Nx \rightarrow (\exists B \in \beta)(Bx\ \&\ \Box(\forall y)(By \rightarrow Ny))]$,
>
> where α is the class of normative properties, β the class of non-normative or base properties, and $\Box$ is a necessity operator.

In English: *necessarily, for any normative property, if that property holds of some x, then there is some non-normative property that x has such that, necessarily, for any y that has the same non-normative property, y also has the normative property*.[13] The relevant non-normative property might be a complex property: it could be the property of being B_1 or B_2 or ….[14] There is relative agreement that the outermost necessity operator should be interpreted as conceptual necessity. It is supposed to follow simply as a matter

[12] See Smith (2004: 225-9) discussing ordinary moral practice; and Hare (1952: 145).
[13] There are weaker versions of the supervenience thesis, concerning the co-variance of normative and non-normative properties within possible worlds. But we will be occupied with the strong formulation in what follows. On different formulations of supervenience, see Kim (1984).
[14] The formulation assumes that non-normative properties are closed under infinitary Boolean operations. Since reductive naturalists deny this, it's not an entirely neutral formulation. One way of avoiding this problem is by talking about *sets* of normative properties. See Bader (2016). But that still does not get the naturalist on board who identifies a normative property with a simple non-normative property (like *being pleasing*), so long as one thinks a property cannot be both normative and non-normative. For a helpful discussion of how to accommodate such views, see McPherson (2012). Our argument does not turn on these complications.

of the concepts involved that the connection between normative and non-normative properties holds. The inner necessity operator should be interpreted as metaphysical necessity. The conceptually necessary truth is that it is metaphysically impossible for two things to be non-normatively identical but normatively distinct.[15] Though SUPERVENIENCE is sometimes interpreted as implying a fact about metaphysical dependence—roughly, that a thing's normative features obtain *in virtue of* its various non-normative features—we won't be concerned in what follows with anything stronger than the mere modal covariance of normative and non-normative features.[16]

My primary concern is with β—the class of non-normative properties on which the normative supervenes—and whether it needs to be restricted or further characterized. I am interested in the weakest nontrivial supervenience claim that stands a chance of being a conceptual truth. The narrower the class of base properties, the stronger the corresponding supervenience claim. But if the class is too narrow, the truth will not hold simply by virtue of the concepts involved (or the meanings of the terms).[17] For example, it may be true that normative facts supervene on the *physical* facts. But this is not a

[15] These assumptions about the types of necessity in play are generally agreed upon. See Dreier (2015); Enoch (2011); Blackburn's (1984); Smith (2004). More importantly, they are explicitly endorsed by expressivists.

[16] See Dancy (1981: 367, 380-2); (2004: 86) on the difference between "in virtue of" claims and supervenience.

[17] Too much precision in specifying the relevant non-normative properties would render the claim *synthetic*: possibly true, but not true in virtue of the concepts involved. It is not a *conceptual* matter precisely which non-normative properties the normative supervenes on. Consider the debate between deontologists and utilitarians. These theorists disagree about whether an act's *rightness*, a paradigmatic normative property, covaries with the non-normative property of being happiness maximizing. But their disagreement about does not appear to be a conceptual one. The truth or falsity of utilitarianism does not fall out of the very meaning of "rightness."

conceptual truth, since it is conceptually possible that a non-physical thing, God, exists, and makes a difference to the normative truths (Sturgeon 2009).[18]

The main aim of this section is to demonstrate that there must be some constraints on the supervenience base that follow from our very concept of a normative property. Leaving β unrestricted may serve the aim favored by theorists of stating the supervenience thesis as uncontroversially and modestly as possible (McPherson 2012; Rosen MS).[19] But it obscures the precise supervenience phenomenon that needs explaining.

Unrestricted supervenience is true but seems trivial. Consider the claim that the normative facts supervene on the non-normative facts with no restrictions on the latter. Among the non-normative properties are *haecceitistic* properties that can only be instantiated by particular individuals, like *the property of being Fred*. There may also be properties indexed to possible worlds, like the *property of being Fred in the actual world*. It is certainly true that two things cannot differ in normative respects without differing in some such non-normative respect. But that is because things that are alike in all of these

[18] The supervenience base for the normative was traditionally characterized in terms of *natural* properties, where a property counts as natural if it is invoked by the natural sciences (microphysical and causal/functional properties, for instance) or would be invoked by the sciences sufficiently idealized. See e.g. Jackson (1998); Smith (2004). See also exchange between Griffin (1996) and Smith (2004). It is sometimes suggested that all the familiar ways of defining the supervenience base render the thesis vulnerable to counterexample. See Sturgeon (2009). I take these skeptical arguments to show only that stating the claim precisely is hard; not that supervenience is false.

[19] One reason it has seemed tempting to leave the category of relevant non-normative properties unrestricted is to get the moral particularist on board. Moral particularists deny that there are exceptionless general principles linking the normative to the non-normative (Dancy 2004). It is true that my argument for constraining the supervenience base rejects an extreme kind of moral particularism, according to which even the haecceities of situations and possible worlds could matter from the normative perspective. But if I'm right, we shouldn't have been concerned with accommodating extreme particularism. The supervenience claim I defend does leave room for more modest versions of particularism, on which there are no snappy general principles in ethics, but we know that some kinds of non-normative properties are ethically irrelevant.

respects are *one and the same thing*! The property of *being Fred in the actual world* cannot be had by anyone other than Fred in the actual world, and, trivially, Fred in the actual world is either good or not good—he cannot be both.

Our aim is to find the weakest conceptually true supervenience thesis that is non-trivial and can serve as an important desideratum in metaethics. I claim that *repeatable* supervenience is the weakest such conceptually true supervenience thesis. There are echoes of the repeatability condition in RM Hare's (1952) original discussion of ethical supervenience.[20] However, repeatability is a far weaker constraint on the supervenience base than any previously suggested or defended.

We can state the constraint precisely. Readers not worried about this can skip to the next paragraph. For each $B \in \beta$, B is repeatable when necessarily if *x* is B then possibly there is a *y* distinct from *x*, such that *y* is B. Our account of the supervenience base should entail that it is always possible for two distinct individuals to be base-identical. In other words, we should be able to say of the B in β that renders supervenience true for a normative property N that it is the case that $\Box_m \forall x\ (Bx \rightarrow \Diamond_m\ \exists y$

[20] Repeatable supervenience is entailed by Hare's claim that moral judgments are necessarily general and universalizable. See Hare (1952, 129):

> As we shall see, all value-judgements are covertly universal in character, which is the same as to say that they refer to, and express acceptance of, a standard which has an application to other similar instances. If I censure someone for having done something, I envisage the possibility of him, or someone else, or myself, having to make a similar choice again; otherwise there would be no point in censuring him…. When we commend an object, our judgement is not solely about that particular object, but is inescapably about objects like it

Hare thought that for a value judgment to be truly universal, it cannot include proper names. See *id.* at 176-177. Repeatability allows that morally relevant non-normative properties might involve essential reference to particulars using proper names. E.g., the property of *being a kicking of Fred* is repeatable (there are many acts that are kickings of Fred), though a principle that treated the property as ethically relevant would not be suitably universal on Hare's view. What repeatability rules out is the property of *being Fred* belonging in the supervenience base for the ethical.

$(By \wedge y \neq x))$, where $\Diamond_m$ stands for 'it is metaphysically possible that.'[21] This constraint is stronger than what we need for present purposes. My argument will go through so long as there is *some* normative property, N_1 in α, such that supervenience holds for *that* property and of the B_1 which renders the supervenience claim true we can say: $\Box_m \forall x\ (B_1x \rightarrow \Diamond_m \exists y\ (B_1y \wedge y \neq x))$.[22] The supervenience phenomenon that calls for explanation, I shall argue, is the supervenience of the normative on a repeatable non-normative property.

Repeatability rules out the inclusion of haecceitistic differences in the supervenience base, of the sort discussed earlier. The base does not include *being Fred* or haecceities of situations or possible worlds, because these cannot be had by distinct individuals.[23] Other properties that repeatability rules out include exhaustive qualitative non-normative profiles of individuals of the sort that cannot be shared between distinct individuals, for identicality in terms of such particularistic non-normative profiles would entail strict identity.[24]

The first argument for repeatable supervenience is a direct argument from inconceivability. Given a lie (or a bad person, or any other badness-bearing individual), it seems impossible to conceive of something identical in non-normative terms but different

[21] There are two ways to implement this constraint. Either β, the set of base properties, is constrained so that $(\forall B \in \beta)\ \Box\forall x\ (Bx \rightarrow \Diamond\ \exists y\ (By \wedge y \neq x))$. Alternatively, we can include the constraint in the statement of supervenience: $\Box_C\ (\forall N \text{ in } \alpha)(\forall x)[Nx \rightarrow (\exists B \text{ in } \beta)(Bx\ \&\ \Box_M(\forall y)(By \rightarrow Ny)\ \underline{\&\ \Diamond\ \exists a\ (Ba \wedge a \neq x))}]$. This affirms inter-world duplication. A stronger requirement would affirm *intra*-world duplication: one might also think (for some normative properties) that $\Diamond_M(\exists a \exists b)(Ba\ \&\ Bb\ \&\ a \neq b)$.

[22] i.e. $\Box_C(\forall x)[Nx \rightarrow (\exists B \text{ in } \beta)(Bx\ \&\ \Box_M(\forall y)(By \rightarrow Ny))]\ \&\ \Box\forall x\ (B_1x \rightarrow \Diamond\ \exists y\ (B_1y \wedge y \neq x))$.

[23] One wouldn't be able to say $(\forall B \text{ in } \beta)\ \Box\forall x\ (Bx \rightarrow \Diamond\ \exists y\ (By \wedge y \neq x))$ because there exists a B—the haecceitistic property of being *this* individual—that can't possibly be had by distinct individuals.

[24] If we assume Leibniz' law: $\forall x \forall y\ ((\forall F)(Fx \leftrightarrow Fy) \rightarrow x=y$ then there will always be some qualitative profile that is identity determining (or non-repeatable). But the assumption is not necessary.

normatively. This imaginative exercise, which is the primary motivation for the conceptual truth of supervenience, standardly involves the failure to imagine that *two* non-normatively identical individuals differ normatively, not the failure to imagine two non-normatively identical individuals. Consider the distinct individuals Fred and Bob who are otherwise qualitatively identical in all non-normative respects. To make their non-normative similarity vivid, suppose that Bob and Fred are two clones created at the same time, and that they are plugged into a computer simulation that feeds them the same experiences. Suppose also that they end up behaving in all the same ways. The only non-normative difference between the two is that Fred is Fred and Bob isn't. Is it conceivable that Fred is a good person, but Bob is not? Surely not. It seems ruled out by the very meaning of the term "good." But this is evidence for the conceptual truth that goodness supervenes on repeatable non-normative properties.[25] And, it is sufficient for our purposes that there is some normative property—the ethical property of *being good*—of which it is conceptually true that it supervenes on repeatable non-normative properties.[26]

The second and third arguments for repeatability appeal to the theoretical significance of supervenience within metaethics. Even though the repeatability constraint has not been made explicit in discussions of supervenience, there are good reasons to

[25] Here's another argument for thinking that the type of inconceivability that motivates supervenience involves consideration of two genuinely distinct normative property bearers. Suppose that in trying to imagine individuals that are non-normatively identical we can only imagine a single self-identical individual. This raises the worry that what we really intuit when it seems to us that we cannot normatively distinguish individuals after fixing their non-normative features is not the truth of supervenience but the logical tautology that a single individual cannot both be N and ¬N. The imaginative exercise is probative of the truth of supervenience only if it involves imagining two genuinely distinct acts that are non-normatively identical and discovering that these distinct acts must be normatively identical.

[26] What about obligations to oneself? If I'm obligated to care about myself simply because of my identity, is that not a counterexample to repeatability? The obligation to care about oneself is not necessarily identity-dependent. The relevant non-normative fact that grounds my obligation to myself is that I am a person with interests, but *being a person with interests* is repeatable.

think it has been assumed. When we look at the uses to which supervenience has been put and ask which thesis metaethicists had in mind, it looks to be repeatable supervenience in every case.

Consider Simon Blackburn's argument for supervenience. Blackburn (1973) suggests that supervenience should be accepted because it helps explain what he calls the "ban on mixed worlds." Blackburn points out that it is numbingly obvious as anything could be that there cannot be non-normative qualitative duplicates in the actual world that differ normatively. And he points out that this is true in the case of *actual* non-normative qualitative duplicates. Two states of affairs in the actual world, one located in the northern hemisphere and the other in the southern, but otherwise non-normatively identical *must* be normatively identical. We can know this without knowing which non-normative features these two states of affairs exemplify. Blackburn argues that the conceptual truth of supervenience explains this fact: the reason *actual* distinct states of affairs that are otherwise non-normatively identically cannot be normatively different is that it is conceptually impossible for entities that are identical in this way to be normatively distinct. Blackburn must have had repeatable supervenience in mind, because his explanation of the "ban of mixed worlds" presupposes repeatability. Unrestricted supervenience cannot explain why genuinely distinct but otherwise non-normatively identical states of affairs in the actual world cannot differ normatively, because genuinely distinct states of affairs are not identical in *all* (unrestrictedly) non-

normative respects.[27] The assumption about repeatability should therefore be common ground between me and my primary interlocutors: expressivists like Blackburn.[28]

The final argument for repeatability is slightly more involved. Suppose we read in tomorrow's newspaper that David Lewis (1968) was right: there is no such thing as strict or literal identity between individuals in different possible worlds. When I consider myself in a different possible world where I am a successful musician, I am not *literally* thinking about myself, but rather a very similar counterpart. Lewisian modal realists are not the only ones who think that the notion of "trans-world identity" is incoherent.[29] But the reasons for thinking so are not our concern.[30] What is relevant is that *no one* ever understood the challenge of explaining normative supervenience in a way that turns on a proof that individuals are world-bound. The conceptual truth of supervenience was supposed to function as a dialectical lever that separates good metaethical theories from bad ones. For example, one of the main vulnerabilities of non-naturalist realism is believed to be the view's struggles when it comes to explaining supervenience.[31] But if *unrestricted* supervenience were the relevant explanadum, then everyone, including the non-naturalist, should be relieved to discover that individuals are world-bound. For we can now explain unrestricted supervenience without assuming anything about the nature of the normative. We can simply appeal to the non-repeatability of the base properties

---

[27] In other words, the actual fact that calls for explanation is $(\forall x)[Nx \rightarrow (\exists B \text{ in } \beta)(Bx \,\&\, (\forall y)(By \rightarrow Ny))$ $\underline{\&\, \exists a \exists b(Ba \,\&\, Bb \,\&\, B\, a \neq b)}]$. So, $\Box_C(\forall x)[Nx \rightarrow (\exists B \text{ in } \beta)(Bx \,\&\, \Box_M(\forall y)(By \rightarrow Ny))]$ wouldn't be explanatory without the repeatability of B.

[28] Blackburn obviously accepts it. Gibbard seems to as well as we shall see in a moment.

[29] See e.g. Heller (1998), Sider (2001) and Forbes (1982).

[30] One might think this because of a commitment to modal realism—one and the same concrete individual cannot be part of two entirely distinct yet concrete worlds. Alternatively, one might accept an *ersatz* view of modality on which only actual individuals have individual haecceities, but denizens of possible worlds (possible individuals) have no such individual essences.

[31] See McPherson (2012).

plus the law of non-contradiction.[32] The reason it is a conceptual truth that it is metaphysically impossible for two base-identical things to be normatively distinct is that base-identical things are identical *tout court* and $\forall x \forall \phi (\phi x \vee \neg \phi x)$! The problem of explaining the conceptual truth that the normative supervenes on the non-normative (a class that includes non-repeatable properties) turns out to be a rather Easy Problem indeed.[33]

The strength of this final argument lies in its showing not just that unrestricted supervenience is too weak, but that repeatability is precisely what we need to assume to arrive at the weakest non-trivial conceptual truth that can serve as an important desideratum in metaethics.[34] It is not like we can give up on the law of non-contradiction. And no one ever thought that assumptions in modal metaphysics concerning identity across possible worlds should bear on the explanation of normative supervenience. Unrestricted supervenience is too easy to explain because it allows non-repeatable properties into the supervenience base, ensuring the literal identity of base-identical individuals, which is the source of the problem.

---

[32] We need to assume that individuals are world-bound—i.e. no strict or literal identity between individuals in different possible worlds—because LNC only applies to individuals at a world. Of course, not everyone rejects trans-world identity, but the point of the argument still stands: supervenience is too easy to explain if B-identity entails strict identity. It can be explained by appeal to (1) B-identity = strict identity; (2) LNC: $\forall x \forall \phi (\phi x \vee \neg \phi x)$; (3) No trans-world identity. Any counter-part theorist or meta-ethicist who denies transworld identity for normative property bearing individuals can explain the phenomenon. Thanks to X for pushing me to make this explicit.

[33] Can the expressivist invoke LNC given that, strictly speaking, on her view the application of normative predicates does not involve the attribution of properties? The next section demonstrates that there is an expressivism-compatible response to the Easy Problem.

[34] Conceptual repeatable supervenience is clearly true in many areas. If two things are exactly alike in all repeatable underlying respects, either both are chairs or neither is. Why is that? It is somehow a conceptual truth that what makes a chair a chair is a matter of its repeatable features, including its shape, its purpose, and so on. What seems true in the case of the chair concept is also true in the case of the concept of goodness. Thanks to X for the comparison.

The strong and interesting thesis that any acceptable metaethical theory must vindicate is that the normative supervenes on repeatable non-normative base (henceforth, I'll take the repeatability condition for granted unless stated otherwise).[35] The challenge of explaining supervenience becomes considerably more difficult once we assume such a minimal and eminently plausible constraint on the supervenience base. The explanatory burden will no doubt vary depending on the details of one's account of the normative. But expressivists, too, face an explanatory challenge, one that has an important general upshot in metaethics.

II. "The Expressivist Side-Step"[36]

The expressivist attempts to explain SUPERVENIENCE with a distinctive account of normative concepts. To deploy a normative concept, according to the expressivist, is not to predicate a property of a thing.[37] On a rough-and-ready understanding of the view, there are *no* normative properties on the expressivist picture, at least not in any substantial sense.[38] Though normative judgments, or utterances involving normative

[35] The assumption, quite apart from being reasonable, is a dialectically fair one insofar as expressivists like Blackburn (1984) and Gibbard (2003) take the relevant truth to be the supervenience of the normative on the repeatably non-normative.
[36] Dreier (2015).
[37] Gibbard (1990: 7-8): "Normative talk is part of nature, but it does not describe nature. In particular, a person who calls something rational or irrational is not describing his own state of mind; he is expressing it."
[38] We can ignore quasi-realist wrinkles in what follows for ease of discussion. Quasi-realist expressivists are happy to admit talk of normative facts and properties, with a deflated account of such talk. To judge that there is a normative fact that one ought to help the poor is equivalent to judging that one ought to help the poor. Both judgments get the expressivist treatment. The argument does not turn on whether we assume a quasi-realistic version of expressivism.

terms, share the syntax of ordinary descriptive judgments, they simply express the agent's non-cognitive attitudes: her desires, preferences, or plans.

To illustrate, on Gibbard's (2003) "plan-based expressivism" the relevant non-cognitive attitude is that of a plan. Roughly, to judge that one ought to give 40% of one's income to charity just is to plan to be that generous with one's money. To *assert* that one ought to do so is to express such a plan. A plan is akin to a dominant preference or controlling desire for a specified outcome—an outcome characterized in terms of actions and circumstances.[39] Let $p\{\phi, C\}$ be a plan to ϕ in a circumstance C (entertained under some description). *ϕ-ing in C* can be described as the plan's aim. There can be grander plans, on Gibbard's view, ones that specify actions for multiple circumstances, and even a universal plan that specifies an action for every possible situation: a plan to ϕ_1 in C_1, ϕ_2 in C_2, ϕ_3 in C_3, and so on.[40]

The expressivist's key explanatory move is to focus on what it is to *judge* supervenience to be true given the special non-cognitivist semantics for normative terms (Gibbard 2003: 90-2; Compare: Klagge 1988; Blackburn 1985).[41] The judgment that it is a conceptual truth that two non-normatively identical things cannot possibly differ

[39] While planning is supposed to be an activity we all engage in and on which we have an intuitive grip, plans are, on Gibbard's construal, markedly flexible. We can plan not just for actual and contingent situations, but also for situations that it is impossible for us to be in. It is crucial for plans to be flexible in this way if expressivism is to be able to capture the full range of normative judgments. Technically, Gibbard uses sets of what he calls 'fact-plan worlds' to interpret states of mind expressed, where a fact-plan world can be thought as a pair (f, p), consisting of a state of the world f and a plan for that state. A fact-plan world (f, p) is compatible with the state of mind expressed when you judge that you ought to phi in a given situation if the 'fact' component is consistent with what you believe the world is like, and the second element—the plan—has you phi-ing. The state of mind is then characterized in term of the set of all fact-plan worlds compatible with one's view of the world and one's motivations regarding what to do. These complexities can be ignored in what follows.

[40] Gibbard (1990: 168-160). See also (1990: 48–53, 68–71).

[41] Complications arising on quasi-realistic expressivism are helpfully explored in Dreier (2015).

normatively amounts to the judgment that it follows from *what it is* to deploy a normative concept—namely, to express a certain kind of plan—that one cannot apply the concept differently to two non-normatively identical things.[42] The only thing that stands in need of explanation is why this constraint on normative concepts holds, an explanation that must be given in terms of the nature of planning.

We can state the conceptual constraint that needs explaining precisely. For ease of discussion, suppose the only normative concept on the scene is that of <being what ought to be done>, and that this concept applies to act-circumstance pairs based on their features, where an act either is or is not *what ought to be done* in the relevant circumstance. SUPERVENIENCE entails that it follows from the nature of the ought-concept that it cannot apply differently to two act-circumstance pairs that are repeatably non-normatively identical (henceforth, base-identical or B-identical).[43] One cannot coherently judge that one ought to ϕ in C_1 but not in C_2 while recognizing the circumstances as B-identical. It would be akin to judging that Tim is a married bachelor. One can of course *assert* "one ought to ϕ in C_1 but not in C_2 and the two circumstances are B-identical." But the assertion would be meaningless or unintelligible.[44] The challenge is to explain why such judgments are unintelligible.[45]

The plan-based expressivist attempts to explain their unintelligibility by appeal to constraints on planning derived from *what it is* to plan (2003: p93-8). A certain kind of

[42] Dreier (2015) aptly describes the expressivist's approach to SUPERVENIENCE as the "expressivist side step" for avoiding the metaphysical puzzle,
[43] We can ignore the precise derivation. For discussion see Klagge (1988).
[44] One can, of course, normatively judge base-identical things differently at different times, just as one can meaningfully judge one and the same person to be married and to be a bachelor at different times.
[45] There may be a related puzzle concerning norms of belief or judgment revision, but we will get to it later when discussing Sturgeon's (2009) objection against the expressivist.

plan is impossible: a plan to ϕ in C_1 and not ϕ in C_2 while recognizing that C_1 and C_2 are B-identical. If such a plan *were* possible, given the semantics, one could meaningfully judge that one ought to ϕ in one situation but not in another while recognizing the situations as B-identical. But since the plan is impossible, the corresponding judgment is incoherent.

Why is the supervenience-flouting plan impossible? Gibbard's own argument provides a helpful starting point for the analysis. First, he assumes the class of base properties is the class of scientifically observable properties or as he sometimes puts it: "prosaically factual properties" (pgs. 32, 98-99). The category includes not just properties currently recognized by the natural sciences, but "spooky properties like the properties of ghosts and gods, so long as these can be *recognized* or *observed* and thus could figure in an empirical science."[46] Second, he suggests that a plan is the type of motivational state that cannot treat empirically-indiscernible circumstances differently:

> [A] plan can distinguish between situations only in terms of their prosaically factual properties, and it can distinguish between acts only in terms of the prosaically factual properties of those acts. If two acts in two possible situations differ in no prosaically factual way, a plan cannot distinguish them, permitting one and ruling out the other. Either it will permit both or it will rule both out (2003: 91-2).

[46] The notion of recognizability is not altogether clear. Gibbard sometimes refers to the relevant properties as "prosaically factual properties" which suggests that the notion might be quite strong. For example, it might involve *phenomenological* recognizability, where if a property is recognizable then it is directly available in phenomenal experience or else impinges in some way on our sensory surfaces. But more plausibly, the category includes just those properties whose existence can be taken for granted under the norms of ordinary scientific practice.

The argument for the claim goes by quickly, but the basic idea is supposed to be that empirically identical (or E-identical) outcomes are quite simply indistinguishable from the planning perspective.[47] We cannot tell two E-identical circumstances apart. A plan to ϕ in C_1 but not ϕ in C_2 when C_1 is E-identical to C_2 would be *just like* planning to ϕ and not ϕ in what one recognizes to be one and the same situation. And it should be evident from a moment's reflection on the nature of planning that a putative plan to ϕ and not ϕ in a single situation would be no plan at all! This is because plans, constitutively, aim to guide action and a plan to ϕ and not ϕ in one and the same situation, would "preclude offering any guidance on what to do on that occasion" (p56).[48]

So, to sum up, the plan-based expressivist doesn't just think a supervenience-flouting plan is one our community of planners generally avoids or dislikes. The plan is *impossible*, and the fact of its impossibility is supposed to be knowable *a priori* from introspective awareness of one's own mental states.[49] The argument for its impossibility drives a tight parallel between the supervenience-flouting plan and a putative plan to ϕ and not ϕ in a single situation. We can make the key moves of the overall explanatory strategy explicit. The expressivist needs to explain why a judgment inconsistent with SUPERVENIENCE is unintelligible. The unintelligibility of the relevant judgment follows from the semantics plus key assumptions about the nature of the motivational state expressed through normative judgment:

---

[47] Dreier (2015) interprets Gibbard similarly: "It is of the nature of plans that the contingencies for which they are plans are features of the situation that are at least in principle recognizable."
[48] Gibbard's intriguing claims about the nature of planning have been curiously neglected. The only one to express any skepticism as far as I know, and that too in passing, is Hawthorne (2002: 173-4)
[49] I consider later and rule out the possibility that Gibbard might be understood as *stipulating* an artificial notion of a plan with the constraint built in.

1. To judge that one ought to ϕ in C_1 but not ϕ in C_2 and that C_1 and C_2 are E-identical *just is* to express a plan with ϕ-ing in C_1 and not ϕ-ing in C_2 as its aim where C_1 is E-identical to C_2. The judgment is unintelligible if the associated plan is impossible. [Plan-based Expressivism]

2. A plan that cannot guide action is impossible; hence, the impossibility of a plan with ϕ-ing & not ϕ-ing in C_1 as its aim. [Basic Constraint on Planning]

3. A plan that aims at ϕ-ing in C_1 and not ϕ-ing in C_2 cannot guide action. [No Action Guidance].

From (1), (2), and (3), we conclude that one cannot coherently judge that one ought to ϕ in C_1 and not ϕ in C_2 while recognizing that C_1 is E-identical to C_2. Call this the Argument from Planning.

III. The Hard Problem for Plan-based Expressivism

The plan-based expressivist has no trouble at all with the *Easy* Problem of Supervenience. Suppose the relevant notion of empirical detectability is such that all properties, including bare haecceitistic properties like *being this particular circumstance at this time and location…* , count as empirically detectable, and that, accordingly, it is not possible for two numerically distinct circumstances to be empirically identical. In other words, suppose the supervenience base is defined to be non-repeatable. If C_1 is E-identical with C_2 entails that C_1 and C_2 are strictly identical, then the Argument from Planning seems very compelling indeed. A plan with ϕ-ing in C_1 and not ϕ-ing in C_2 as its

aim where C_1 is E-identical to C_2 really does amount to a plan with ϕ-ing & not ϕ-ing in C as its aim. And it is very plausible that one cannot plan to ϕ *and* not ϕ in a single circumstance C. No motivational state can cause one to ϕ and not ϕ in one and the same circumstance. The impossibility of this kind of plan mirrors in the required way the unintelligibility of judging that one ought to ϕ and not ϕ in one and the same circumstance. Of course, it might be possible, to independently form distinct plans for a single circumstance that are jointly inconsistent.[50] After considering one's current situation at different times, one might form the plan to go out for dinner *and* plan not to go out. But it is one thing to have two (cognitively-isolated) plans for a single situation that are incompatible; and quite another to have a *singular* plan that aims at what the agent recognizes as an impossible end: *ϕ-ing & not ϕ-ing in C_1*. The expressivist can reasonably maintain that the impossibility of planning to ϕ and not ϕ in C_1 mirrors in the

[50] Sturgeon (2009) uses such cases to raise a challenge for Gibbard. Sturgeon suggests that a commitment to SUPERVENIENCE must involve accepting a rational norm of revising pairs of normative judgments jointly incompatible with SUPERVENIENCE. Sturgeon (2009) demands an explanation for why discovering inconsistencies in one's plans should be universally recognized as "a cause for alarm, or reform… rather than just for amusement? Expressivists needs to show (and it is uncertain whether they can) that all planners are committed to a norm of revising pairs of plans that are jointly inconsistent with SUPERVENIENCE. I am skeptical that assent to SUPERVENIENCE necessarily involves accepting the relevant norm governing combinations of plans, but my main purpose in discussing Sturgeon's objection is not to criticize it. It is to note that the cases he describes do not amount to counterexample to the Basic Constraint on Planning.

If the expressivist needs to explain why assent to SUPERVENIENCE operates as a judgment-guiding norm, it suffices to point out that once an agent recognizes that two of her judgments are jointly inconsistent, she will assert, based solely on her competence with normative concepts, that both cannot be true. Gibbard should say, in accordance with his general strategy, that the agent knows—or at least is guided by—an a priori fact about planning that renders unintelligible the judgment that normative judgments that are jointly inconsistent with SUPERVENIENCE are both true. He does not need to show that assent to SUPERVENIENCE involves being motivated to resolve inconsistencies. Admittedly, Gibbard writes as though assent to SUPERVENIENCE involves being committed to a higher-order plan of a certain sort: a plan to avoid having incompatible pairs of plans for E-identical situations.

required way the incoherence of a singular judgment that one ought to ϕ and not ϕ in C_1.[51]

While the plan-based expressivist has no trouble at all with the Easy Problem of Supervenience, the Hard Problem presents a real challenge. If we assume the supervenience base of empirically recognizable properties is repeatable, then the Argument from Planning seems far less compelling because its third premise—No Action Guidance—seems false. Whereas a plan to perform incompatible acts in a single situation may well be impossible because it cannot guide action, a plan to treat two situations differently *just because* they are distinct, even though their distinctness is undetectable, seems possible because it *can* guide action.[52]

We might compare a plan to treat Bob differently from Fred, when Bob and Fred are identical twins and there is no way to distinguish Fred from Bob. The Argument from Planning tries to establish such a plan's impossibility by appeal to the action-guiding nature of plans. Necessarily, if *x* is a plan, then *x* is a motivational state with some action-involving outcome as its aim *and x* is a state that can possibly cause or bring about the action-involving outcome. That is why a plan that aims at performing an action and not performing it in one and the same situation is impossible. No mental state can possibly cause one to both ϕ and not ϕ in C. But there *is* a motivational state that:

(1) has incompatible actions for undetectably distinct situations as its aim; and

[51] That said, one can meaningfully judge that one ought to ϕ in C_1, and independently judge that one ought *not* to ϕ in C_1; just as one can make independently meaningful judgments that Tim is a bachelor and Tim is a married man.

[52] I think that the empirical undetectability of bare identity is perhaps the most plausible (but ultimately mistaken) basis for thinking that our motivations cannot treat two situations differently *just because* they are distinct, which is why Gibbard's discussion of the issues provides a useful starting point and frame for the overall argument.

(2) can appropriately cause the realization of its action-involving aim.

After all, one can dominantly *desire* or *prefer* to perform incompatible acts in two situations just because they are distinct situations. There is no bar to dominantly desiring incompatible acts for genuinely distinct situations while recognizing the situations as empirically indiscernible. The desire may be whimsical, to be sure. But whimsical desires are not impossible. The posited desire can combine with one's beliefs in a situation to result in action. So long as one forms true beliefs about which situation one is in, one's dominant desire that one perform incompatible acts in the two situations will result in some action (possibly even the desired one). So, it *is* possible for a mental state to be both action-guiding and have as its aim: incompatible acts for distinct but empirically indiscernible situations.

If one is tempted by the thought that a plan *just is* a dominant preference/desire for an outcome characterized in terms of actions and circumstances, then one should think that the supervenience-flouting plan is possible. The view is very tempting indeed if all we're told about plans is that they are necessarily action-guiding motivational states individuated using action-circumstance pairs. But perhaps the expressivist can do better than the Argument from Planning. She might insist that a plan is not just a motivational state that can guide action. It constitutively involves a *reasonable expectation* that one will realize the outcome aimed at. The motivation to treat undetectably distinct situations differently must combine with a suitable expectation if it is to count as a plan. And it is this reasonable expectation that is missing in the case of a whimsical desire to treat two situations differently just because they are distinct situations. The only possible worlds where the desired outcome is realized involve luck—worlds where one has a randomly

acquired true belief about which of two distinct situations one is in. Plans plausibly involve evidentially supported or rational expectations for plan-realization, along with the relevant desires. [53]

This might seem like a compelling response, but it is vulnerable to two basic objections. First, we must be careful not to confuse the potential oddness or imprudence of the candidate plan for the strong sort of impossibility at issue.[54] Generally, when we plan to do something we also have some sense of how we will reliably do it, and this seems like good practice. It would be setting oneself up for disappointment to regularly leave the successful implementation of one's plans to chance. One way of guarding against frustration is making sure one has a reliable method for determining whether the situations relevant to one's plans have transpired. A plan to treat two situations differently just because they are distinct and not on the basis of any detectable difference between the two goes against common sense because it leads easily to the frustration of one's ends. The crucial question, however, is not whether such a plan is a strange or bad one, but whether it is impossible or inconceivable. It is neither. There are plenty of impudent planners who are not *failing to plan* on account of their imprudence.[55]

---

[53] Other ways that theorists have suggested plans might be different from mere dominant desires are not helpful. Bratman (1987), for example, thinks that intentions are distinct from mere desires in essentially involving a kind of resolution not to reconsider one's intention. But, of course, pairing resolve with a dominant desire to realize {(ϕ, Day1) (not ϕ, Day2)} is possible. Even if plans in Bratman's sense must be supervenience-respecting, an argument I give later concludes that plans in the sense relevant to normative judging must be a kind of supervenience-flouting motivational state.

[54] Gibbard suggests only that plans necessarily aim to guide the agent's actions (2003: 56). Nowhere is the narrower conception of action-guidance on the basis of evidentially supported beliefs defended or even made explicit.

[55] There appear to be plenty of ordinary cases of agents purporting to have plans of the sort described – that distinguish situations based on undetectable features. Gnostics believe in a God who exists beyond space and time and is not just unknown but *impossible* to know, at least by ordinary mortals. A world where the Gnostic God exists and a world where he doesn't are plausibly E-identical, but unobservably distinct. A person who follows the faith's practices will

The second and more decisive reason why the 'reasonable expectations' constraint should be rejected is that, setting aside our pre-theoretic intuitions about plans, plans in the sense relevant to normative judging *must* be the kind of motivational state that can treat undetectable differences between situations as relevant from the planning perspective. It might be true that a plan to treat situations differently just because they are distinct situations cannot involve a reasonable expectation of plan-realization. But that *cannot* be the reason for the plan's being impossible, given the role that plans play in the expressivist's semantic theory. We must keep in mind that plans, for the expressivist, do significant theoretical work: they undergird meaningful normative thought and talk. Moreover, expressivists are (and indeed should be) committed to vindicating as meaningful those of our ordinary normative judgments that involve so-called *robustly objective* uses of ought. Such usage involves judging that a person ought to do something even when she does not know and could know that it is the thing she ought to do. The person might be constitutionally disposed to doing the wrong thing.[56] Consider:

<Paranoid schizophrenics ought to get themselves some therapy>

According to the expressivist, in asserting this fact, I am, expressing my plan to *get therapy in the circumstance of being a paranoid schizophrenic*. But if I were a paranoid schizophrenic, I might be insensitive to facts about my own paranoia and have no inclination to get therapy. If my extreme psychological propensity to not ϕ in a situation because of my inability to reliably identify the situation based on my evidence were

---

treat these worlds very differently with respect to her plans. She might plan to be kind in a world where a Gnostic God exists and not be kind in the world where he doesn't. I think we have good reason to take the agent's reports about her plans at face value.

[56] See Daskal's (2009) discussion of Gibbard's view in the context of "innocent mistakes of mental constitution," and, in particular, what Gibbard can say about 'Mortis,' the ideally coherent psychopath.

enough to make a plan to ϕ in that situation impossible, the expressivist would have to give an error-theoretic diagnosis of what seems to be a clearly intelligible and quite likely *true* normative claim: that paranoid schizophrenics *should* get therapy. So, a plan to treat two situations differently just because they are distinct situations cannot be impossible on the grounds that it depends, for its success, on the eventuality of forming true non-evidentially supported beliefs about unrecognizable features of one's situation.

Similarly, it is sometimes argued that there are uses of *ought* where blameless ignorance does not undermine the truth of the ought judgment. Consider:

> <I morally ought to turn the light switch on and off repeatedly if doing so will cure cancer, *even if* it is entirely unknowable by me that it would>

While it is controversial as a matter of first order moral theory whether such objective ought claims are true, they are at least meaningful. An expressivist committed to vindicating them as meaningful *has* to think that one can plan to turn the light switch on and off in a circumstance where doing so will cure cancer, even if there is no way for one to detect this feature of the circumstance.

So, we can summarize the critical point (which would apply even if we started our discussion with a different motivational state favored by expressivists, like a stable, higher-order desire): whatever desire-like attitudes plans are supposed to be, they must be attitudes that can underwrite as meaningful robustly objective normative judgments. But this means a plan to treat two situations differently *just because* they are distinct situations cannot be impossible for treating as relevant from the planning point of view undetectable features of two situations. Plans cannot constitutively involve rational or

reasonable expectations of plan realization.[57] It must be possible to plan to treat E-identical but genuinely distinct situations differently. So, supervenience-flouting plans seem possible, if the supervenience base is defined as the category of *repeatable* non-normative properties—a category which does not include bare haecceitistic features of situations.[58]

At this point, the plan-based expressivist might be tempted to avoid the challenge by means of stipulation. She might grant that one can indeed plan to treat E-identical but genuinely distinct situations differently while insisting that only supervenience-respecting plans are ones the expression of which results in meaningful normative concept use. We can, after all, define a set of plans* which includes only those plans that do not distinguish between distinct circumstances recognized as E-identical. Normative concept use necessarily involves the expression of plans*, so repeatable supervenience-flouting judgments are unintelligible. Analogously, I could define up the concept of a desire*, which is a lot like the concept of a desire, except that one desires* to ϕ only if ϕ-ing involves maximizing happiness. It is impossible to desire* to not maximize happiness because of *what it is* to desire*.

---

[57] Note that, by contrast, an *unreasonable* expectation may be constitutively involved in planning, but that does not help the expressivist. A desire to ϕ if it is C_1 and not ϕ if it is C_2 *can* combine with an expectation that the desired outcome will be realized. For suppose one forms the evidentially unsupported belief that:

<*In each circumstance, I shall form a correct belief as to whether it is C_1 or C_2.*>

The critical question is whether this belief is consistent with taking the two situations to be E-identical, which we must hold fixed. The answer is that it is indeed consistent. We can stipulate that one doesn't expect one's beliefs about which situation one is in to be correlated in anyway with the truth.

[58] As it turns out, it wasn't even necessary for us to establish decisively that supervenience-flouting plans are possible. We have successfully raised a serious problem for the expressivist by marshaling enough reason to *doubt* and *suspend judgment on* the impossibility of supervenience-flouting plans. But suspending judgment on whether planning is constitutively constrained to bar supervenience-flouting normative judgments entails suspending judgment on whether supervenience holds as a conceptual matter if expressivism is true.

There are several reasons why brute stipulation cannot save the expressivist. Firstly, the move would render normative concepts and whether one succeeds at deploying them arbitrary. After all, while I can't plan* to act differently in situations that are E-identical, I can do something that looks a lot like planning*—namely, I can plan to act differently. It seems hardly plausible that whether I succeed at meaningfully deploying normative concepts can hang on so tenuous a thread.

More importantly, the expressivist cannot simply stipulate that normative concepts involve plans* and not plans, no more than one can stipulate a meaning for the ordinary concept of redness. The expressivist must *motivate* the claim that normative concepts involve planning* and not planning. Crucially, she cannot appeal to our commitment to SUPERVENIENCE as proof that we are planners* because an explanation of our commitment to SUPERVENIENCE was supposed to emerge from reflecting on the activity we are engaged in when we deploy normative concepts. That is, we were supposed to have a more explanatorily basic grasp on the nature of the non-cognitive attitude expressed when we use normative terms like 'good' and 'ought' than on why it is that we are committed to SUPERVENIENCE. And if there is no other justification she can give for restricting meaningful normative judgments to those that involve the expression of a plan*, the restriction will seem *ad hoc* and designed to save the view from objection.

It is worth emphasizing this final point as a kind of summary of the dialectic so far. We must keep in mind the order of explanation: our verdict on supervenience is supposed to be *explained* by our understanding of normative concepts and not the other way around. Since, on the expressivist picture, an understanding of normative concepts turns out to be an understanding, perhaps implicit, of the non-cognitive mental state we

express when we deploy normative concepts, it had better be the case that we know we are planners* and not planners (when we deploy normative concepts) independently of our knowledge of SUPERVENIENCE. I have argued that expressivists have offered us no reason for thinking that we are planners* rather than planners when we deploy normative concepts.

IV. A General Challenge

There are two other ways expressivists might try to deflect our challenge. One would be to concede that the motivational state involved in normative thought and talk *can* be supervenience-flouting, but then provide some reason for further complicating the semantics. The expressivist might try to explain *why* we only recognize normative judgments as meaningful when the motivational state expressed is supervenience-respecting (instead of simply stipulating that we do).

Compare, for example, how a non-naturalist robust realist about normativity might go about explaining features of our normative concept. The robust realist might argue that our concept of a normative property is that of a property that covaries with a repeatable non-normative property precisely because it lies in the nature of the objective property in normative reality that our concept tracks that it covaries with a repeatable non-normative property, and normative reality has in some way shaped the concepts we use. This at the very least has the form of a genuine explanation, whatever one thinks of its assumptions. The expressivist obviously cannot appeal to objective normative reality to explain why meaningful normative concept deployment involves the expression of

supervenience-respecting plans.[59] Instead, she might appeal to facts about our usage and the purposes that normative language is designed to serve. As argued earlier, what she cannot do is appeal to the truth of SUPERVENIENCE.

The problem is that no expressivist has offered a remotely plausible explanation of why, independently of our commitment to SUPERVENIENCE, only supervenience-respecting non-cognitive attitudes can underwrite normative thought and talk. Blackburn (1986), for whom the relevant non-cognitive attitude is a kind of stable, higher-order desire, suggests that we use normative vocabulary to coordinate with one another. We engage in normative discussion, and the associated expression of our motivational states, to invite others to share our higher-order desires and plans so we can harmonize our actions. If normative talk aims at coordination, this might explain our disposition to negatively evaluate those who express supervenience-flouting motivational states. That is, the coordination account would predict that we would be inclined to use normative language to express our disagreement with those who express supervenience-flouting motivations insofar as our own motivations are supervenience-respecting. But it is hard to see why the coordination aim of normative language should entail that normative claims underwritten by supervenience-flouting non-cognitive attitudes would be rejected as *meaningless*. To put it differently, if normative language was introduced to facilitate coordination amongst agents, the expression of motivational attitudes that *can* possibly violate supervenience serves just as well as the expression of those that constitutively cannot.[60]

[59] On this point, see Dreier (2015).

[60] So, when Blackburn (1993) writes:

> it seems to be a conceptual matter that moral claims supervene on natural ones. Anyone failing to realize this or to obey the constraint would indeed lack something constitutive

An alternative means of deflecting the challenge would be to suggest that plans cannot treat haecceitistic features of situations—bare identity—as relevant from the planning point of view for some reason other than the empirical undetectability of bare identity. While our previous arguments show that the fact (if it is one) that bare identity is empirically undetectable cannot be the reason for treating supervenience-flouting plans as impossible, we have not shown that there might not be some other reason.

There are two things to say about this attempted deflection. First, the burden is on the party claiming the impossibility of a kind of mental state to provide reasons for thinking it impossible. In fact, our reason for using Gibbard's discussion as a starting point was precisely that his account seemed most likely of any to produce a reason for thinking it is impossible to plan based on haecceitistic differences. But we've seen that the empirical undetectability rationale does not work. Secondly, we have strong reasons, based on the previous discussion, for supposing that the motivational states relevant for normative judging *must* be able to treat individuals— persons, acts, situations, or even possible worlds (construed as maximally described situations)—differently based purely on their identity. It is certainly true of a dominant preference or desire: I can dominantly prefer Fred over Bob even if Fred and Bob are indistinguishable twins just because Fred

---

> of competence in the moral practice. And there is good reason for this: it would betray the whole purpose for which we moralize, which is to choose, rank, approve, for bid, things on the basis of their natural properties. (p66)

he presupposes the fact that needs to be explained—namely, that the purpose behind using normative terms (or moralizing) is to choose, rank, approve and so on *specifically on the basis of the natural properties of things*. As we have seen, the coordination account cannot explain why a supervenience-flouting normative judgment would so betray the aim of coordinating to render the judgment unintelligible. Without a plausible explanation of why it must be supervenience-respecting motivational attitudes that underwrite normative claims, the expressivist should suspend judgment on whether supervenience is a conceptual truth, at least if the supervenience base is defined in terms of 'natural' or empirically recognizable features.

is Fred and Bob isn't! The only constraint on such mental states derivable from their nature that *is* plausible is a rather trivial constraint: to treat *numerically identical* individuals alike. So, perhaps a singular preference that has its object ϕ-ing and not ϕ-ing in a single circumstance C is impossible. This trivial constraint on our motivational states ensures that the expressivist can straightforwardly address the *Easy* Problem of Supervenience. But she has no answer for the *Hard* Problem of explaining why it is a conceptual truth that the normative supervenes on the repeatably non-normative.

V. Ways of Biting the Bullet

The expressivist has two options in light of the Hard Problem, neither one of which leaves the conventional wisdom about supervenience and expressivism intact. One is to concede that the conceptual fact of repeatable supervenience has no explanation. This would amount to embracing the view I described earlier as stipulative, but without pretending that it solves the Hard Problem. The expressivist can complicate her semantic story and accept as a *brute fact* about our interpretive practices *vis à vis* normative discourse that we only recognize normative judgments as meaningful if they involve the expression of motivations consistent with supervenience. There is no deeper explanation for why normative talk works this way. As far as brute facts go, this one needn't be all that mysterious. In general, there are presumably many arbitrary features of our linguistic practices that lack deep explanations. Notably, expressivism, unlike other meta-normative theories, at least renders it *plausible* to take the relevant conceptual fact to be brute. Contrast the unattractiveness of an analogous move in the case of the non-naturalist. If

the non-naturalist takes it to be a brute fact about our normative practice that we do not recognize normative judgments inconsistent with supervenience as meaningful, she confronts the question: how is it that an arbitrarily determined feature of our practice mirrors a robustly metaphysical fact in normative reality—namely, the supervenience of robustly real normative properties on the repeatably non-normative?[61] So, perhaps the right thing for expressivists to say is not that expressivism *explains why* it is a conceptual truth that the normative supervenes on the non-normative. She can say that expressivism explains why it is unproblematic to take as *unexplained* the conceptual truth that the normative supervenes on the non-normative.

But, of course, expressivists have long claimed the ability to *explain* the conceptual truth of SUPERVENIENCE.[62] Expressivist explanatory ambitions have presumably been guided by the attractive assumption that our normative concepts are not arbitrary and that a feature like the supervenience constraint calls out for explanation. While it may be true that the conceptual fact is more plausibly taken as brute given expressivism rather than, say, non-naturalism about the normative, the move remains less than satisfying for anyone who took it to be an important pre-theoretical datum that the nature of the concept demands a deeper explanation. Another problem with taking the conceptual truth as brute is that it is in tension with the fact that, by the expressivist's own lights, we find no further support for SUPERVENIENCE through greater insight into the nature of normative concepts. What we learn about the nature of normative thought and

[61] Incidentally, the non-naturalist does not have to take the relevant conceptual fact as brute, and in this non-naturalists may find basis for claiming a theoretical advantage over expressivists (and reductive naturalists), but that is a discussion for a different paper.
[62] Blackburn (1970: 120; 1985: 64): "[Supervenience] is *explained* by the anti-realist nature of moralizing." See Gibbard (2003: 96) and the discussion of Gibbard's view in Section II.

talk—namely, it's motivation-expressive function— is consistent with SUPERVENIENCE being false.

There is an alternative reaction to the Hard Problem that seems more attractive than treating SUPERVENIENCE as a brute conceptual truth. The expressivist might reject the claim that it is a *conceptual* truth that the normative supervenes on the repeatably non-normative. She might argue, instead, that supervenience is a *normative* truth rather than a conceptual one.[63] As discussed, the expressivist needn't face any significant challenge explaining why we might shun plans to treat empirically indiscernible situations differently. Such plans certainly seem like bad ones insofar as they reliably lead to the frustration of one's ends and are possibly grounded in unreasonable expectations. Embracing supervenience as a kind of normative truth would allow the expressivist to explain why it might have *seemed* to be a conceptual truth. For one, the epistemology of normative and conceptual beliefs is similar. Both types of belief appear to be based on *a priori* reasoning.[64] Moreover, this approach allows the expressivist to give a deeper explanation for the truth of supervenience, albeit a normative one, in terms of why one ought not to have supervenience-flouting motivations.

While some of the costs of adopting either one of the two non-ideal (or "unhappy-face"[65]) solutions to the Hard Problem may be perfectly general and concern the enduring

[63] There is a minority view that takes precisely this view of supervenience. See Kramer (2009); Rosen (MS).
[64] For a priorism in ethics, see Audi (2015: 61).
[65] Schiffer (2003: 5-6) calls a solution to a philosophical puzzle a "happy-face" solution if it can identify which amongst a set of incompatible intuitions is false and explain why it is false. Unhappy-face solutions simply point out that there isn't a happy face solution to the puzzle. There is instead an irreconcilable tension resulting from our pre-theoretical commitments. The problem of squaring expressivism with our intuitions about supervenience admits only of an unhappy-face solution.

plausibility of supervenience conventionally understood as a conceptual truth that demands explanation, there is a theoretical price that expressivists necessarily incur. Expressivists have long regarded it a central virtue of their account of the nature of the normative that it is both consistent with and explains the conceptual truth of normative supervenience.[66] For example, Simon Blackburn (1993: 137) writes that it is a principle virtue of expressivism that it fully explains why "it seems to be a conceptual matter that moral claims supervene. . . . Anyone failing to realize this or to obey the constraint would indeed lack something constitutive of competence in the moral practice." Once we render the relevant supervenience phenomenon precise, we find that expressivist explanations of the conceptual truth falter. Having to take repeatable supervenience to be a normative truth or an unexplained conceptual truth undercuts the expressivist's right to claim a dialectical advantage that has long been taken for granted on all sides: the advantage of having no revisionary or counter-intuitive implications with respect to supervenience. The problem of explaining the conceptual truth of normative supervenience, precisely described, turns out to be a more general problem in metaethics than conventional wisdom would have us believe.

References

Audi, R. (2015) "Intuition and Its Place in Ethics", *Journal of the American Philosophical Association*, 1: 57–77

[66] Blackburn (1984); Gibbard (2003).

Bader, R. (2016) “The Grounding Argument Against Non-reductive Moral Realism.” Oxford Studies in Metaethics.

Blackburn, S. (1971) “Moral Realism,” *in Essays on Quasi-realism*, Oxford: Oxford University Press.

_____(1984) *Spreading the Word*. Oxford: Oxford University Press.

_____(1993) “Supervenience Revisited,” *Essays on Quasi-realism*, Oxford: Oxford University Press.

Bratman, M. (1987) *Intentions Plans, & Reasons*. Center for the Study of Language and Information.

Daskal, S. (2009) “Plan-Based Expressivism and Innocent Mistakes,” *Ethics* 119: 310–35.

Dancy, J. (1981) “On Moral Properties,” *Mind*, 1981, 90: 367—385.

_____ (2004) *Ethics Without Principles*, Oxford: Oxford University Press.

Dreier, J. (2006) “Negation for Expressivists: A Collection of Problems with a Suggestion for their Solution,” in (ed. Shafer-Landau, R.) *Oxford Studies in Metaethics*, Vol. 1

_____ (2015) “Explaining the Quasi-Real,” in (ed. Shafer-Landau, R.) *Oxford Studies in Metaethics*, Vol. 10. Oxford University Press. Oxford.

Enoch, D. (2011) *Taking Morality Seriously*. Oxford UK. Oxford University Press.

Fine, K. (2002) “Varieties of Necessity,” in (Tamar Szabo Gendler & John Hawthorne, eds.) *Conceivability and Possibility*. Oxford University Press. pp. 253-281.

Gibbard, A. (1990) *Wise Choices, Apt Feelings,* Cambridge, Mass: Harvard University Press.

_____ (2003) *Thinking How to Live.* Cambridge, MA. Harvard University Press.

Griffin, J. (1996) *Value Judgment,* Oxford: Clarendon Press.

Hawthorne, J. (2002) “Practical Realism?,” *Philosophy and Phenomenological Research,* 64: 169-178.

Heller, M. (1998). “Property Counterparts in Ersatz Worlds,” *Journal of Philosophy*, 95: 293-316.

Jackson, F. (1998) *From Metaphysics to Ethics*, Oxford: Clarendon Press.

———, Pettit, P, & Smith, M (2000). “Ethical Particularism and Patterns”, in Hooker and Little (eds.) *Moral Particularism*, Oxford: Oxford University Press.

Kim, J. (1984) “Concepts of Supervenience,” *Philosophy and Phenomenological Research* 45: 153–76.

Klagge, J.C. (1988) “Supervenience: Ontological and Ascriptive,” *Australasian Journal of Philosophy* 66: 461-70

Kramer, M. H. (2009) “Supervenience as an Ethical Phenomenon,” in Kramer, *Moral Realism as a Moral Doctrine*. Oxford: Wiley-Blackwell, 304–64.

Lewis, D. (1986) *On the Plurality of Worlds*, Blackwell: Oxford.

McPherson, T. (2012) "Ethical Non-Naturalism and the Metaphysics of Supervenience." In *Oxford Studies in Metaethics* Vol 7.

Parfit, D. (2011) *On What Matters*, Oxford, UK: Oxford University Press

Rosen, G. (MS). *"*What is Normative Necessity?,*"* available at: https://www.academia.edu/9159728/Normative_Necessity.

Schiffer, S. (2003) *The Things We Mean*, Oxford University Press. Oxford UK.

Smith, M. (1994). *The Moral Problem*. Oxford: Blackwell.

_____ (2004) "Does the Evaluative Supervene on the Natural?" reprinted in *Ethics and the A priori*, Cambridge: Cambridge University Press.

Sturgeon, N.L. (2009) "Doubts about the Supervenience of the Evaluative," In *Oxford Studies in Metaethics*, Vol. 4

Zangwill, N. (1995) "Moral Supervenience". *Midwest Studies in Philosophy*, 20: 240—262.

Chapter 2. How to be Impartial as a Subjectivist

Abstract: The metaethical subjectivist claims that there is nothing more to a moral disagreement than a conflict in the desires of the parties involved. Recently, David Enoch has argued that metaethical subjectivism has unacceptable ethical implications. If the subjectivist is right about moral disagreement, then it follows, according to Enoch, that we cannot stand our ground in moral disagreements without violating the demands of impartiality. For being impartial, we're told, involves being willing to compromise in conflicts that are merely due to competing desires—the parties to such conflicts should decide what to do on the basis of a coin flip. I suggest that Enoch is mistaken in his conception of what it means to be impartial. Once impartiality is properly construed, standing one's ground in desire-based conflicts, whether or not moral values are at stake in the conflict, is consistent with being impartial. I defend a view on which impartiality can be understood in terms of features of our desiring attitudes. An agent acts impartially in desire-based conflicts whenever she is motivated by a final (i.e. non-instrumental) desire that aims at promoting the wellbeing of persons in a way that is insensitive to the identities of persons and their morally arbitrary features like their gender or skin color. Based on the account, I explain where Enoch's discussion of the argument goes wrong, as well as why responses to the argument from Enoch's critics have so far missed the mark.

Keywords: subjectivism, robust realism, impartiality, moral disagreement

Introduction

David Enoch (2011) argues that metaethical subjectivism is inconsistent with some of our basic ethical principles.[1] Here is a pared down version of his argument. On the one hand, the right thing to do in many cases of moral disagreement is to stand one's ground and refuse to compromise. On the other hand, if a disagreement between parties is merely due to their having incompatible desires, the right thing for them do is to be impartial, where being impartial involves being willing to compromise—for instance, the parties might decide what to do on the basis of a coin-flip. The subjectivist thinks that there is nothing more to a moral disagreement than a conflict in the desires of the parties involved. And so it seems the subjectivist is forced to either reject the principle that one ought to be impartial in desire-based conflicts or else reject the principle that one ought to stand one's ground in moral disagreements. Neither option seems especially attractive.

Fortunately for the subjectivist, there is a third way out of the argument. It involves showing that Enoch's conception of impartiality is mistaken. Although Enoch stipulates that 'being impartial' means compromising on one's own preference in mere desire conflicts, his principle of impartiality is only as plausible as the degree to which it reflects the underlying moral considerations: those concerning what it means to be impartial in the familiar, moral sense. And being impartial in the familiar, moral sense is not always a matter of compromising on one's own desires in desire-based conflicts, or so I hope argue. My aim is to show not only that a refusal to compromise can amount to a genuinely impartial response, but also that in *all* cases of desire-based conflict where we,

[1] While the argument is principally intended as a *reductio* of subjectivism, Enoch suggests that it easily generalizes to threaten even the expressivist. My response on behalf of the subjectivist is one the expressivist can also help herself to.

intuitively, ought to be impartial as well as refuse to compromise, we can easily do both. In other words, my aim is to bar Enoch from using any such case as the basis for an 'argument from impartiality' against subjectivism.

To that end, I offer a general account of what it means to be impartial in desire-based conflicts. I suggest that an agent always acts impartially in such conflicts when she acts on the basis of 'final' or non-instrumental desires of a certain sort. Roughly, these are final desires that aim at outcomes where the good of persons is promoted in a way that is insensitive to the identity of persons and their morally arbitrary features. Cases of desire-based conflict where one should behave impartially but also refuse to compromise are cases where one's desires that gave rise to the conflict already reflect one's impartiality. I argue that such a view is most plausible when paired with a certain, ecumenical conception of an agent's reasons for acting—that is, the considerations in light of which the agent acts.

The account offers more than just a decisive answer to Enoch's argument. Based on it, I explain why the responses that Enoch considers on behalf of the subjectivist (and rightly dismisses) seem unsatisfying, as well as why recent responses in the literature miss the mark (e.g. Katherine Manne's and David Sobel's (2014)). These responses seem to me to be worth rehearsing to emphasize a point about the dialectic— that first-order challenges to anti-realist views are not so easily dismissed.[2] Indeed, while I disagree with Enoch's conclusions, I am broadly sympathetic to his sense that the compatibility of anti-realist views with our moral practices is far too often assumed without argument. Enoch is not the only theorist to harbor doubts about the subjectivist's capacity to act

[2] On the significance of first-order challenges to meta-ethical views, see Sections III.

consistently with familiar moral principles governing our conduct,[3] and a clearer account of the relationship between morality's demands and what value resides in being impartial should go some way towards addressing such concerns.[4]

I. The Argument from Disagreement

Enoch's argument, in its most basic form, is intended as a *reductio* of 'caricatured subjectivism.' According to the caricatured subjectivist, a moral utterance of the form 'Torture is wrong' just means 'I prefer that people not torture one another,' where 'prefer' is to be understood as "picking out a simple, non-special straightforward preference" (p25). The view can be specified in terms of desires rather than preferences without affecting the argument, and I will use the terms interchangeably throughout the paper. Enoch offers the following characterization of the target view, assumed to be true for purposes of the *reductio*:

> CARICATURED SUBJECTIVISM: Moral judgments report simple preferences, ones that are exactly on a par with a preference for playing tennis or for catching a movie.

The caricature, Enoch thinks, resides in attributing to the subjectivist the claim that moral and non-moral preferences are 'on a par'—a claim we might refer to as the parity thesis.

---

[3]Although Fantl (2006, p. 30) does not put the point in terms of impartiality, he argues that given the subjectivist's meta-ethical beliefs, her participation in our ordinary (presumably justified) practice of interfering violently when persons intend to cause great suffering seems morally inappropriate.

[4] The objections to my view that I consider towards the end of the discussion allow me to comment on the relationship between morality and the good in acting impartially as well as on the general relevance of an agent's meta-ethical beliefs to the moral evaluation of her practical attitudes.

It is not entirely obvious how Enoch interprets the claim. For now, we can limit ourselves to noting that Enoch presents subjectivists who endorse it as being most vulnerable to the argument.

The argument relies on a crucial normative principle that Enoch calls IMPARTIALITY and takes to bear on interpersonal conflicts involving competing preferences.

> IMPARTIALITY: In an interpersonal conflict, we should step back from our mere preferences, or feelings, or attitudes, or some such, and to the extent the conflict is due to those, an impartial, egalitarian solution is called for. Furthermore, each party to the conflict should acknowledge as much: Standing one's ground is, in such cases, morally wrong (p19).

The principle might be motivated by appeal to cases of the following sort. Suppose you and I are trying to decide whether to spend our afternoon watching a movie or playing tennis. Whereas I would prefer to watch a movie, you would rather play tennis, and our preferences happen to be irreconcilable. It seems we ought to feel some moral pressure to resolve our dispute impartially, perhaps, by flipping a coin. Enoch thinks that the pressure stems from an obligation you and I share to treat each other as moral equals. As he puts it, "each one of us should acknowledge that we are equally morally important," and "that our preferences should—other things being equal—count equally" (p.18). IMPARTIALITY purports to capture the relevant moral considerations as they bear on our conflict, those concerning what it means to be impartial in the true and moral sense.

Enoch is quite explicit in recognizing that it is far from obvious that moral considerations, having to do with impartiality or otherwise, always demand that we settle

interpersonal conflicts involving incompossible preferences in an egalitarian fashion as his principle commands. Since being impartial is not the only value one can serve through one's actions, there will be plenty of cases where compromise is *not* the response that is, all things considered, called for. Nevertheless, he thinks IMPARTIALITY is relevant often enough to be worth taking seriously and that it captures an important value at stake in interpersonal conflicts. For our purposes, we can take for granted that there are indeed many situations involving preference conflict—and the movies-v-tennis case may be one of them—in which considerations of impartiality demand a conciliatory solution.

Combining IMPARTIALITY with CARICATURED SUBJECTIVISM has unacceptable consequences. According to CARICATURED SUBJECTIVISM, moral disagreements amount to conflicts in mere preference, and IMPARTIALITY demands that we settle such conflicts in an egalitarian, conciliatory vein. But clearly there are many cases where it is not only morally permissible but quite obligatory that we stand our ground and refuse to compromise with the wickedly intentioned. For example, it seems that I ought to feel free (perhaps even obliged) to guide my actions in the light of my judgment that torture is wrong even if a sadist judges things differently and wishes to torture me. Caricatured subjectivism appears to have objectionable moral implications, given that it demands compromise in many cases of interpersonal conflict where moral steadfastness appears to be the right response.

Borrowing directly from Enoch, the argument from impartiality might be formalized as follows:

(1) CARICATURED SUBJECTIVISM. (For Reductio.)

(2) If CARICATURED SUBJECTIVISM is true, then interpersonal conflicts due to moral disagreements are really just interpersonal conflicts due to differences in mere preferences. (From the content of CARICATURED SUBJECTIVISM.)

(3) Therefore, interpersonal conflicts due to moral disagreements are just interpersonal conflicts due to differences in mere preferences. (From 1 and 2.)

(4) IMPARTIALITY, that is, roughly: when an interpersonal conflict (of the relevant kind) is a matter merely of preferences, then an impartial, egalitarian solution is called for, and it is wrong to just stand one's ground.

(5) Therefore, in cases of interpersonal conflict (of the relevant kind) due to moral disagreement, an impartial, egalitarian solution is called for, and it is wrong to just stand one's ground. (From 3 and 4.)

(6) However, in cases of interpersonal conflict (of the relevant kind) due to moral disagreement often an impartial solution is not called for, and it is permissible, and even required, to stand one's ground.

(7) Therefore, CARICATURED SUBJECTIVISM is false. (From 1, 5, and 6, by Reductio.) (p. 25—26).

Although Enoch thinks that a subjectivist with any shot at resisting the argument needs to deny premise 1 (specifically, the parity thesis), the approach I recommend involves rejecting premise 4—that is, Enoch's IMPARTIALITY principle. For clarity's sake, it might be worth distinguishing two separate claims made in premise 4:

(4a) when an interpersonal conflict (of the relevant kind) is a matter merely of preferences, then an impartial, egalitarian solution is called for.

(4b) when an interpersonal conflict (of the relevant kind) is a matter merely of preferences it is wrong to just stand one's ground (because an impartial, egalitarian solution is called for).

4a seems correct—an impartial, egalitarian solution is indeed called for in interpersonal conflicts of the relevant kind.[5] I will argue that the subjectivist should take issue with 4b, as neither implied by 4a nor correct on independent grounds, at least if 'impartial' is taken in its actual, moral sense. The demands of impartiality do not preclude standing one's ground in cases of preference conflict, and may very well require it.

Before moving on to evaluating the argument, we might briefly consider Enoch's remarks regarding why the objectivist is not vulnerable to a similar argument. For the objectivist, moral disagreements do not reduce to conflicts in mere preference. When you and I disagree about whether it is OK to torture people, we purport to be representing how things really are in the world, independently of our preferences. The truth about whether torture is wrong provides one of us with an impartial standard to appeal to in defense of a refusal to conciliate. I do not offend against the equal moral worth of persons when I stand my ground on a moral issue if, by doing so, I am merely standing up for the truth—or so Enoch thinks. He motivates the claim by appeal to non-moral factual disputes. When you and I are trying to disable a bomb and we disagree about whether to

[5] Later I will suggest that 4a is correct *for the most part*. That is, there may be special cases of desire-based conflict where we should all agree (realists and subjectivists alike) that we ought not to be impartial (See section IV). Indeed, Enoch recognizes that the principle is one that need not hold without exception. All he needs is for the principle to hold often enough to make trouble for the subjectivist.

cut the blue wire or the red wire, and one of us knows the truth about which wire to cut, that person need not feel obliged to conciliate for reasons of impartiality. Whether or not Enoch's explanation of what goes on in factual disputes is persuasive, we can safely ignore it, at least for the time being.[6] What interests us at present is not whether the argument from impartiality reveals a problem that subjectivists *uniquely* face, but whether the argument presents subjectivists with any kind of problem in the first place.

II. The Nature of Impartiality

Enoch's argument rests on a false assumption—namely, that the moral demands of impartiality preclude standing one's ground in mere preference conflicts.[7] This section introduces on behalf of the subjectivist, an account of what it takes to stand one's ground impartially. I deduce from this account an alternative to Enoch's IMPARTIALITY principle, and present some of the reasons why it ought to be preferred. Applying the principle to both moral and non-moral preference conflicts, I explain why it offers the subjectivist a decisive response to Enoch's argument. In later sections, I explain why the response I develop here is to be preferred to various alternatives in the literature (Section III), and address some of the objections that might be raised against the meta-ethically neutral conception of impartiality on which the response depends (Section IV). The rationale for postponing discussion of likely objections until later in the paper is that it should make

[6] We will return, in section IV, to the purported virtues of an account of impartiality specified in terms of belief in response-independent truth.
[7] Throughout the remainder of the discussion I will use 'impartial' to refer to the quality of an act or person that the relevant moral considerations in fact favor; the quality that Enoch's principle purports to capture. I will use IMPARTIAL to express Enoch's stipulated meaning.

exposition easier, and I am optimistic that once the overall account is clearly laid out, its plausibility should be readily apparent.

Let me begin by clarifying some of the terminology I will be using in what follows. Desiring has a two-part structure, one that involves a state of affairs desired, always under some description or mode of presentation, and the attitude of desiring it. It is a familiar enough fact that a state of affairs may be desired as a means to the realization of some other state— that is, instrumentally; or *for its own sake*, in which case it is desired finally. The 'essential description' of a finally desired state of affairs includes all and only those of its features that sustain one's desire for it. A single state of affairs may be represented in many ways, each picking out a different set of salient features. Only some of those features will be such that one's desire for the state would diminish if it did not have them. So, for instance, if I desire to stop sadists from torturing my pet dog, it is an essential feature of the state I desire that it involves the cessation of pain caused to a sentient creature. It need not be an essential feature of the state that it involves *my* dog as opposed to my neighbor's. Undoubtedly, it may take a fair bit of introspection to determine the essential description under which one finally desires a state of affairs.

On the view I will be defending, whether or not a person acts impartially in a given case can be determined by the essential content of the final desires that motivate her to action. Consider, for instance, a way for one's final desires to be evidence of one's *partiality*. Suppose that my final aim in standing my ground during a preference conflict is simply the hedonic upshot associated with satisfying my own first order desires— desires which, let us say, happen to be directed at outcomes familiarly regarded as morally good. If my refusal to compromise on my own preference (when others prefer

differently) is motivated entirely by a desire to promote my own happiness, my inflexibility falls short of the requirements of impartiality. In being inflexible, I am manifestly privileging my own interests over the interests of others. Even if I am generally inclined to finally desire the gladness associated with satisfying my own desires, it seems important that, if I am to act impartially, this self-focused final desire of mine not be what motivates me to stand my ground during a preference conflict.[8] Of course, the state of affairs my desires finally aim at when I stand my ground may be *describable* as one involving the satisfaction of my own (moral) desires over someone else's (immoral) ones. But it need not be finally desired under a description that makes essential reference to facts about my own (or anyone else's) desire satisfaction.[9]

The point might be put in terms of the agent's reasons for acting (in the deliberative and not merely causal sense) as indicative of the description under which she finally desires a state of affairs.[10] An agent's reasons are the considerations or features of a situation in light of which she acts, those that present to her as relevant/justificatory in her practical deliberations about what to do. It is admittedly controversial what the state of a feature 'striking' a person as practically relevant amounts to. But I do not think that the dispute between the realist and the subjectivist on this question is likely to be relevant here. The point I am about to make draws on what should be common ground: namely, that agents typically do attend to (prosaically naturalistic) features of the world in

[8] For a discussion of how virtue, generally, involves the situational silencing of desire, see McDowell (1998) and Seidman (2004).
[9] A single state of affairs can be desired under different descriptions, and I know of no argument that shows that when I finally desire some outcome, I must finally desire it under every possible description.
[10] A fact can causally explain why the agent acted without being the agent's reason for acting in the sense at issue.

deciding how to act and sometimes the features they attend to are part of the explanation for why they acted the way they did.[11] That is, sometimes the considerations the agent takes as justificatory are in fact what motivate her, in the sense that they correspond to the content of her motivating desires. If her desire to ϕ finally aims at states of affairs with feature F, her reason for ϕ-ing might be that ϕ-ing promotes states of affairs with that feature. Of course, an agent's reasons and her motivating desires can come apart, and everyone needs an account of what such states involve.[12] But we can focus on cases where the agent's reasons and the content of her desires align in the sense outlined—where her reasons are *indicative* of what in fact motivated her.[13]

The 'selfishly steadfast' agent described above seems partial partly because, insofar as she is transparent to herself, her reason for sating her desire for the morally good

[11] Admittedly, the view of an agent's reasons being presupposed here is not entirely uncontroversial. But as far as I can tell, the presupposition is dialectically fair. Notably, Enoch is happy to regard an agent's reasons (unlike *normative* reasons) as ordinary features of actions and states of affairs: "The reasons for which we act – that it contains vitamin C, that she needs help, that he's charming, that it's so expensive, that I really want to – these can be perfectly ordinary, naturalistically respectable things" (2011, p.219).
In fact, the subjectivist might say a good deal more about what is involved in a consideration striking one is practically relevant (or, in other words, it *seeming* to one that the consideration is relevant). It involves *experiencing* a desire with a particular aim and, perhaps, forming appropriate beliefs about one's desire. To judge that the fact that ϕ-ing promotes states of affairs with feature F is a reason to ϕ is to judge that one desires to realize states of affairs with feature F. For further discussion on this point, see footnotes 13 and 14.

[12] Although the details of the subjectivist's view are not entirely relevant, here is what I think she can say. Sometimes the agent can be mistaken about her reasons in the sense that the considerations that strike her as relevant do not in fact motivate her. It might *seem* to me that the fact that ϕ-ing maximizes happiness is a reason to ϕ, when in fact I only ϕ because it promotes my own interests (my motivating desire to ϕ finally aims at my own happiness). Nevertheless, for it to *seem* to the agent that F is a reason to ϕ, the agent must, at the very least, be in a state that seems, phenomenologically, like a motivating desire state with the appropriate content. As far as I can tell, nothing in Enoch's argument rules out the subjectivist's commitment to such theses.

[13] The focus on such cases is reasonable because, as I go on to suggest, her reasons are relevant to her impartiality only insofar as they are indicative of what motivates her.

outcome is the prospective pleasure to be gained from doing so.[14] But there is no need for the subjectivist to suppose that whenever an agent acts her reasons are self-regarding in this way. The temptation to think otherwise may stem from the following rather suspect line of reasoning. Suppose the subjectivist takes the fact that ϕ-ing will maximize happiness to be practically relevant—to, say, militate, decisively, in favor of ϕ-ing. By her own lights, this cannot amount to much more than the judgment that she desires to maximize happiness.[15] After all, the judgment that F militates decisively in favor of ϕ-ing (is a decisive reason to ϕ) seems like a *normative* judgment, which must be analyzed in terms of desires. In a straightforward sense, then, according to the subjectivist what makes the fact that ϕ-ing will maximize happiness practically decisive is the fact that she desires to maximize happiness. According to a view that Mark Schroeder (2007, p.24) in a slightly different but related context refers to as the "No Background Conditions View," anything

[14] I say partly because it is not enough to be considered impartial to not take self-regarding considerations as ones reasons. For one might still be motivated by self-regarding final desires. In fact, on the view I am proposing the agent's impartiality is essentially determined by the content of the final desires that motivate her, and insofar as the agent's reasons are relevant at all it is because the considerations she takes as practically significant are (generally) indicative of what motivates her. Even the falsidical state of being mistaken about one's reasons (in the sense that the considerations one regards as relevant are not what motivate one) might reveal something about the agent's motivational set. This is because, as I understand the falsidical state, it is due to the agent's being in a non-motivating desire or 'desire-like' state. Roughly, for it to *seem* to an agent that the fact that ϕ-ing will promote states of affairs with feature F is her reason to ϕ, she should experience a state that is phenomenologically a lot like a motivating desire for states of affairs with feature F. Ideally, an agent's desires (whether or not they motivate her) should not be in any way self-regarding; and impartiality may be, on the view I am proposing, a matter of degree—in particular, the degree to which an agent's motivational states are at all self-regarding. These finer details of the view I have kept out of the main discussion because they are not essential to countering Enoch's argument. They might be useful, however, in explaining why one might have the intuition that an agent's reasons bear on her impartiality. An agent's reasons have a kind of derivative (evidential) significance and only because of the significance of her actual desires—a fact that should become vivid once the overall conception of impartiality is fully fleshed out.

[15] Of course, a less caricatured subjectivist might give a richer account of the 'taking' state, but I am embracing the caricature for purposes of the argument.

invoked to explain *why* something is practically relevant must itself be practically relevant.[16] A commitment to NBCV entails that the subjectivist must always regard the fact that she desires to ϕ as practically relevant or, in other words, her reasons for acting must always refer back to her own desires. And if her reasons for acting are always self-regarding in this way, it raises the worry that her actual motivations are as well.

Fortunately for the subjectivist, there are good reasons for rejecting the NBCV. For one thing, the combination of subjectivism and the NBCV is inconsistent with the phenomenology of practical deliberation—we do not ordinarily reflect on facts about our own desires in deciding what to do (see Smith and Pettit 1990). [17] For another, the distinction between what some thing is and the things that must be cited in a complete explanation of why it is that thing is a perfectly good one in other contexts (see, e.g., Schroeder's example of the conditions that explain why Barack Obama is the president of the United States and what it is to be the president of the United States). In any event, it would be dialectically fair and reasonable for the subjectivist to reject NBCV. If pressed to say more, she should say that an agent's reasons—the considerations in light of which she acts / those that present to her as practically relevant—are typically what her desires

[16] Schroeder states the view in terms of normative reasons. The view is that anything involved in the explanation of what makes it the case that something is a normative reason is itself part of the reason. I have deliberately eschewed talk of normative reasons to avoid giving the impression that my response on behalf of the subjectivist illicitly relies on robustly realistic meta-ethical commitments. Of course, the subjectivist is entitled to speak in terms of normative reasons. I suspect that the response being offered is more illuminating when put in terms of desires and the agent's reasons.

[17] "We are no more inclined to think that the deliberating agent always considers his desire-states than we are to imagine that he always considers his states of belief. In deliberating, and more generally in inference, the agent will consider alleged facts such as that p or that q without considering the fact about himself, if it is a fact, that he believes that p or that q. He may take cognizance of the fact that he has this or that passion or yen or hankering. But such self-concern seems to be the exception, not the rule." Smith and Pettit (1990). See also Enoch (2011, p221).

aim at, not the desires themselves; and that the distinction between a desire and its aim is a perfectly good one. Even when her desire to ϕ is phenomenologically salient during practical deliberation—in the sense that she is aware that she is in a state with a distinct motivational character (the desiring to ϕ state) as she deliberates about what to—this does not make her desire to ϕ part of her reason for ϕ-ing.[18] The only case in which her desire to ϕ becomes part of her reasons for ϕ-ing is when she thinks something like 'ϕ-ing has the property of promising to satisfy my desire to ϕ, *so I should* ϕ,' a judgment that might indicate that she is motivated by a self-regarding final desire.

So let us suppose that the subjectivist who stands her ground in a moral disagreement is not motivated by facts about her own desire-satisfaction; what must be true of her motivations for her conduct to qualify as impartial? Another way to put the question is as follows: in virtue of which features of her actual motivations might she be regarded as impartial. The general lesson to take away from the example of the selfishly steadfast is that if I am to act impartially in standing my ground, my final desires should not be responsive to the interests of particular persons over the interests of others merely in virtue of their identities. Let us call final desires 'identity-independent' if they aim at states of affairs under a description or mode of presentation that does not make essential reference to any particular person. When my desire for a state of affairs is identity independent it means that the features of the state that sustain my desire for it—the features of the state in virtue of which the state presents as inherently attractive to me—

[18] Smith and Pettit (1990) sometimes seem to suggest that it suffices for an agent's desire to be part of her reasons (in their terms: 'deliberatively foregrounded') for the desire to be phenomenologically salient during deliberation. On my view, this is not the case. For the desire to be part of her reason, the agent must think something like 'ϕ-ing has the property of promising to satisfy my desire to ϕ, *so* I should ϕ.'

can be characterized without using any indexical terms or proper names. The identity-independence condition implies that, in order for me to qualify as impartial, the motivations underpinning my steadfastness during a disagreement cannot finally aim at a state of affairs in which the interests of particular persons (like myself) are catered to simply in virtue of who they are.

There may be conditions other than identity-independence that my final desires would have to satisfy before it would be reasonable to describe me as impartial. One possibility is that my desires may need to also be 'non-arbitrary.' Final desires count as non-arbitrary if they are insensitive to morally arbitrary features of persons like their gender or skin color.[19] Note that it is just the essential content of my final desires that cannot refer to the relevant features. I might desire to promote race-based affirmative action, but my final desires would not thereby be *essentially* responsive to anyone's race. The non-arbitrariness condition will be relatively ignored in the ensuing discussion because the kind of partiality mainly at issue in Enoch's argument is the partiality involved in having identity-*dependent* desires—that is, the subjectivist is portrayed by Enoch as favoring herself over others by standing her ground in preference conflicts. A further reason to be less concerned with the non-arbitrariness condition is that it might be possible to describe the failure of impartiality involved in racism and other forms of discrimination simply in terms of the having of identity-*dependent* desires (specifically, in terms of desires that aim at privileging one's in-group).

Consider, then, the significance of standing one's ground in a conflict by acting on a final desire to promote the wellbeing of persons in a way that is both non-arbitrary

[19] Which features count as morally arbitrary will, of course, depend on the correct moral theory.

and identity-independent.[20] For present purposes, 'wellbeing' refers to the good of persons, non-normatively construed in familiar ways. We can rely on the 'objective list' theories of Scanlon (1993) and Hurka (1993), which characterize wellbeing in terms of such things as happiness, self-respect, and knowledge. While the concept will have to remain under-specified, that does not mean it is entirely lacking in content. We know, for instance, that a person's wellbeing, so conceived, is not just a matter of her desires being satisfied.[21] My motives would be consistent with the above characterization if, in standing my ground, I were to simply desire the maximization of aggregate wellbeing. A final desire to maximize aggregate wellbeing would clearly satisfy the stated conditions. Alternatively, I might finally desire the maximization of wellbeing subject to familiar side-constraints, so long as my interest in respecting those side-constraints does not reflect an identity-*dependent* final desire or a final desire essentially responsive to morally arbitrary features of persons. For instance, I may desire that aggregate wellbeing be maximized subject to a prohibition against violating the fundamental rights of persons or subject to a requirement that persons who harm others are themselves harmed.

Might it not be enough to make a person's impartiality vivid to note that she acts on the sort of final desire described above? Indeed, final desires characterized in terms of identity-independence, non-arbitrariness, and well-being promotion are all the subjectivist needs in order to motivate an alternative and considerably more intuitive conception of the relevant moral intuitions (those concerning what it means to be

[20] I do not mean to suggest that the impartial perspective has to be cashed out in terms of desires that aim at wellbeing promotion. The characterization in terms of wellbeing promotion is largely illustrative, as I go on to explain.

[21] Pure desire-satisfaction theories of wellbeing are being ruled out for being overly reductive. There are well-known objections to these theories that I need not discuss here.

impartial in the true, moral sense) than Enoch's. According to this conception, the following principle governs inter-personal conflicts:

> IMPARTIALITY*: In interpersonal conflicts that are merely due to competing preferences, an impartial solution is called for, which means acting on the basis of a final desire for states of affairs whose essential mode of presentation is identity-independent and non-arbitrary. An impartial solution entails, among other things, that one cannot act on the basis of a desire that particular persons enjoy greater wellbeing simply in virtue of their identities. Acting on the basis of an identity-independent, non-arbitrary final desire for the promotion of wellbeing is a paradigm case of acting impartially.

Note that, like Enoch's IMPARTIALITY principle, the one on offer here does not purport to capture *all* of the relevant moral considerations that bear on interpersonal conflicts. The principle attempts to capture only what is called for (the response favored) by what value exists in being impartial.[22]

Here are reasons for accepting IMPARTIALITY* and even preferring it to Enoch's principle.[23] To begin with, IMPARTIALITY* seems consistent with the rationale Enoch offers for his own principle. Recall that Enoch motivates IMPARTIALITY by appeal to the *equality of persons*. Realists and subjectivists alike should want to go in for something

[22] A further clarification: like Enoch's principle, the one on offer holds *for the most part*. I explain in section IV that there may be special cases of desire-based conflict where we should all agree that one ought not to be impartial.

[23] Here is Enoch's principle for reference: "In an interpersonal conflict, we should step back from our mere preferences, or feelings, or attitudes, or some such, and to the extent the conflict is due to those, an impartial, egalitarian solution is called for. Furthermore, each party to the conflict should acknowledge as much: Standing one's ground is, in such cases, morally wrong."

like IMPARTIALITY, he argues, because "each one of us should acknowledge that we are equally morally important" (p. 18). Now, surely, being motivated to promote the wellbeing of persons in a way that does not privilege or disfavor any particular person on the basis of their identity or their morally arbitrary features is a way of showing persons equal respect. In other words, having the kind of motivational profile that IMPARTIALITY* demands is supported by the considerations that speak in favor of treating persons equally. By contrast, it is not at all clear that the same considerations always favor having the motivations required by Enoch's IMPARTIALITY principle—which, recall, bluntly calls for a willingness to compromise whenever there is a (mere) preference conflict and an equal number of preferences on either side of the conflict. In cases where being willing to compromise on one's own preference entails giving up on one's desire to promote the interests of persons equally, following Enoch's principle would be inconsistent with treating persons with equal respect (understood as promoting their interests equally). The point is that it seems more plausible that the impartial way to behave in a world where subjectivism is true—where moral disagreements (including over whether we should treat persons' interests equally) are just due to competing preferences—is as IMPARTIALITY* demands rather than IMPARTIALITY.

Moreover, Enoch's principle excludes a wide range of responses to non-moral preference conflicts that seem consistent with treating persons with equal respect. Consider the movies-v-tennis case. In that case, we are invited to suppose that going to the movies would contribute to my wellbeing just as much as playing tennis would contribute to yours. The supposition seems reasonable and, so, a conciliatory solution to our dispute seems favored—we should decide on the basis of a coin flip, seeing as it best

promotes our interests without privileging either one of us. However, whether we should conciliate is not entirely *settled*, as Enoch's IMPARTIALITY principle seems to imply, by the existence of a preference-conflict. There are ways of further specifying the case so that a non-conciliatory response begins to seem just as impartial as flipping a coin—a fact that only IMPARTIALITY* can capture. Suppose that while you and I disagree about whether to watch a movie or play tennis, I know your preference to be capricious. I know that you will lose your desire to play tennis and find yourself wishing you had gone to the movies the moment we start playing. There is at least an argument to be made that instead of ceding to your current, yet fleeting preference for tennis, I should insist on a movie because it would best promote your wellbeing which, one might reasonably think, is a function of your diachronic happiness. My conduct might seem objectionably paternalistic. It may even be morally disfavored, all things considered. But as far as the specific reasons of impartiality go, they do not disfavor my insisting on a movie, so long as I am not ultimately motivated by my own interests—I would be willing to play tennis, if I knew it would make you lastingly happy. Given my intention in standing my ground to promote our wellbeing interests in an identity-independent, non-arbitrary manner, it would be unfair to accuse me of *partiality*, whatever else I might be justly accused of. Of course, a conciliatory response might also be consistent with acting impartially.[24] The point of the case is just to show that whether we should conciliate in ordinary cases of non-moral preference conflict for reasons of impartiality is far less settled by the existence of a preference-conflict than Enoch's principle makes it out to be. I submit that

[24] If I were to decide to conciliate, perhaps because I prioritize respecting personal autonomy over promoting happiness, I would qualify as impartial so long as my final desires remain identity-independent and non-arbitrary.

IMPARTIALITY*, by leaving open a wider range of responses to non-moral preference conflicts than Enoch's principle, offers the more precise characterization of what acting impartially in such cases really amounts to.

Enoch suggests in response that, here, his principle does not apply because the preferences of the parties are not 'equal' in the sense that one is capricious and the other not. But nothing in Enoch's IMPARTIALITY principle suggests that such a move would be kosher. For his principle does not distinguish between desires according to their stability across time. It states that an egalitarian solution is called for whenever there is nothing more to a conflict than the conflicting preferences of the parties involved. Moreover, it seems evident that the case should *not* be an exception to the principle—that is, his principle should have something to say about the case. After all, it is a case involving conflicting desires where one ought to be impartial. The fact that Enoch's principle needs to be refined even in response to such an ordinary case, should raise serious doubts about whether it even approximately captures the relevant moral considerations bearing on how we should behave in desire-based conflicts.

Might Enoch appeal to objective moral truth (realistically understood) as a distinguishing factor here? The reason, he might say, that my insisting on my preference for watching a movie would be impartial is that the moral facts militate in favor of acting in ways that maximize (diachronic) happiness. In other words, the conflict over whether to watch a movie or play tennis as I have described it is *not* just due to desires in Enoch's sense. In response, we might stipulate that the case is one where, although morality recommends acting impartially, it does not settle whether I should insist on the movie or flip a coin. There are, after all, good reasons to pursue either option (as noted above).

Quite plausibly, objective moral truth—supposing that there are such truths—does not settle whether one should insist on the movie or not. If so, then by standing my ground I couldn't just be doing as objective moral truth recommends I do, even if I am a robust realist. If the only thing the realist can say in defense of insisting on the movie is that morality permits her to act in this way, why cannot the subjectivist say the same: namely, that in some conflicts that are just due to desires in the sense that objective moral truth (realistically understood) does not settle what desire one should have, sticking by one's desires is perfectly permissible/impartial? These considerations are not meant to be decisive, of course. But they should give us some reasons for being skeptical of IMPARTIALITY.

Endorsing IMPARTIALITY* allows the subjectivist to agree with Enoch that there is value in being impartial when parties have competing preferences while also disagreeing with him that that value necessarily conflicts with standing one's ground in moral conflicts if subjectivism is true. The principle is consistent with acting on one's own preference during mere preference conflicts, so long as one's preference is itself reflective of one's impartiality. Indeed, in paradigmatic cases of moral disagreement, subjectivists and realists alike should recognize that standing ones ground and frustrating the preferences of immoral people best serves the goal of impartially maximizing wellbeing. The person vulnerable to torture, for example, clearly has a greater wellbeing interest at stake in the moral disagreement than the would-be torturer.[25] In sum, we can

[25] In a review of *Taking Morality Seriously*, Sepielli (2012) argues, in a similar vein, that the subjectivist should explain why standing one's ground in such a case does not offend against impartiality by appeal to the intuition that some tastes are simply offensive and so do not warrant our concern. Enoch might reasonably ask why we shouldn't think that in disregarding the torturer's tastes the subjectivist is simply revealing her partiality towards herself and her own preferences. In my terms, the disregard of offensive tastes may reflect a side-constraint on the

account for our practices of compromise and non-conciliation during preference conflicts in terms of a fairly familiar sort of desire—the desire to promote the wellbeing of persons without privileging particular persons in virtue of who they are or arbitrary features of them. And what more could we be possibly asking of a person than that he act on such a desire, when we demand that he act impartially? (I will turn this rhetorical question into something closer to an actual argument in section IV… although, as I said before, I expect the fleshed out proposal to be inherently plausible.)

Clearly, a morally steadfast subjectivist is not prevented from acting on the kind of final desire that I claim makes for impartial conduct (unless she endorses a kind of motivational egoism—see discussion above on the No Background Conditions View). Admittedly, the subjectivist may run the risk of polluting what motivates her to stand her ground by having one thought too many about her own desires and the prospective pleasure to be gained from satisfying them.[26] But an unhealthy (impartiality-threatening) focus on her own motivations is certainly not forced upon her by her meta-ethical beliefs. The prospect of any private gains from standing her ground can be motivationally 'silenced' by the strength and character of her final desires for the identity-independent, non-arbitrary promotion of wellbeing.[27]

---

promotion of wellbeing. So long as the subjectivist's refusal to take offensive tastes into account does not reflect bias or prejudice towards particular persons or sensitivity to arbitrary features—that is to say, so long as she finally desires outcomes under identity-independent, non-arbitrary descriptions—her conduct qualifies as impartial. See the discussion in section IV for additional objections Enoch might raise against such a view and how they might be answered.

[26] I am not entirely sure about the degree to which the subjectivist, in light of her meta-ethical beliefs, may be psychologically prone to what Mark Johnston (2001) describes as the "pornographic attitude," one involving a "change of attentive focus from the appeal of other things and other people to their agreeable effects on us." I discuss the relevance of this issue briefly at the end of section IV.

[27] For a general discussion of the relationship between motivational silencing and virtuous action, see McDowell (1998).

III. Why Alternative Responses are Unsatisfying

The preceding discussion was guided by the thought that impartiality appears to be a feature of a person's attitude towards other people, an attitude reflected in her motivations (and, typically, her reasons for acting). The challenge for the subjectivist was to describe the motivations of a person willing to stand her ground during moral disagreements in sufficient detail so as to make vivid the person's impartiality quite apart from her meta-ethical beliefs. I suggested that the challenge can be straightforwardly met by identifying the final desires of the person as identity-independent, non-responsive to arbitrary features of persons, and aiming at wellbeing promotion. The characterization in terms of wellbeing promotion, in particular, was largely illustrative and part of an attempt to accurately depict the intentions behind our general willingness to frustrate the goals of wickedly intentioned individuals.[28]

Bearing in mind that impartiality is reflected in what a person's desires aim at, we can see why the responses that Enoch himself rejects do not seem quite right. In his original discussion of the argument, Enoch suggests that the only way for the subjectivist to resist his argument is to insist that, in cases of moral disagreement, the party with the desire for what is familiarly regarded as morally good is entitled to give her own desire greater weight in deciding how she should behave, precisely because it is an inherently

[28] An alternative way of characterizing the impartial agent's final desires may be in terms of outcomes that are familiarly regarded as morally good and are consistent with the identity-independence and non-arbitrariness conditions. I discuss alternative characterizations in section IV.

better desire than her rival's (who let us assume desires the morally bad outcome).[29] Enoch puts the point thusly:

> [The subjectivist] can certainly say that some preferences are more important than others, or that some moral views (or, for that matter, some preferences) are better than others, thereby (purportedly) reporting yet more preferences, and rejecting the supposed parity between the preferences to which morality is reducible and preferences for playing tennis or for catching a movie (where presumably she does not think that some preferences are better than others). And if some preferences are better than others, then why be impartial among them?

Although he does not make his reasons entirely explicit, Enoch feels that a response along these lines would be "objectionably ad hoc" (p.33) and unsatisfying as an explanation for why the impartial agent compromises in the movies-v-tennis case but not in the moral case. Note that the objection isn't the dialectically unfair one that the subjectivist's normative claims, about some desires being fundamentally better than others, seem implausible when heard in the subjectivist's intended register (that is, as claims involving the self-attribution of certain higher-order preferences). The objection is that the subjectivist's normative explanation, for why the morally steadfast agent acts in accordance with the demands of impartiality, is insufficiently explanatory even before it is cast in any particular meta-ethical light.[30]

---

[29] In other words, the subjectivist needs to deny the parity thesis—that all preferences/desires are "on a par" (p27).

[30] Indeed, the subjectivist is granted, if only for the sake of argument, use of ordinary normative vocabulary. She is allowed to say things like 'A is inherently better than B' or that 'A ought to be preferred to B.' Of course, the subjectivist has a precise understanding of what such claims amount to—namely, that they involve the self-attribution of preferences or desires. For example,

Enoch thinks that the explanation is unsatisfying because it lacks a crucial dimension—one that involves appeal to robustly real normative facts. In fact, the problem with the explanation lies not in its meta-ethical ecumenism, but rather in it's being *implausible* as a normative explanation. The explanation as stated suggests, rather implausibly, that acting impartially in interpersonal conflicts is necessarily a matter of weighing the fundamental normative differences between the desires of the parties involved. The reason, we're told, for our disparate ways of behaving in non-moral preference conflicts, on the one hand, and moral preference conflicts, on the other, is that we find the desires of the parties to be equally good in the one case and inherently unequal in the other. The explanation seems implausible because we do not ordinarily expect individuals in moments of interpersonal conflict to act in the light of judgments of comparative worth across desires. Before refusing to let a sadist torture animals, I do not find myself examining the normative differences between what I desire and what the sadist does. My desires and deliberation seem entirely focused on protecting creatures from harm—that is, the state of affairs I would likely realize if I were to interfere with the sadist's goals.[31] The ordinary explanation for why we act differently in the two kinds of preference conflicts and why most of us succeed in respecting the demands of impartiality as a matter of course is thus unlikely to refer to comparative judgments made about desires. It makes for a far more intuitive response to the argument to locate the

---

the claim that preferring A is inherently better than preferring B simply amounts to the self-attribution of a higher-order non-instrumental preference: 'I finally prefer preferring A to preferring B.' Enoch's line of attack involves questioning the plausibility of the subjectivist's normative explanation quite apart from her meta-ethical commitments. The approach is dialectically fair and we should embrace it.

[31] A realist might insist that the relevant aspect of my deliberation, as far as my impartiality is concerned, is that it is guided by some belief about the response-independent moral truth about what we ought to do. I argue against this proposal in section IV.

ordinary agent's impartiality (along with the explanation for her practices of compromise and non-conciliation) in her rather ordinary first-order desires. The agent is motivated by final desires that aim at promoting the wellbeing interests of all of the parties to a disagreement, without privileging particular persons in virtue of who they are or arbitrary features of them. These attitudinal features of the agent straightforwardly underwrite her impartiality, features that have little to do with normatively comparing motivational states.[3233] Once the subjectivist responds as I have suggested, it should be apparent that no further (metaphysically deeper) explanation is needed. At any rate, I will discuss further in the section to follow whether my own proposal on behalf of the subjectivist requires a further layer of explanation.

Before moving on, it might be worth considering one more alternative response to Enoch's argument, in part because it should make more vivid the advantages of the present proposal. Kate Manne and David Sobel (2014) rightly question Enoch's account

---

[32] In saying that her first-order desires reflect an impartial attitude, we need not be attributing fundamental better-making features to her motivational states that militate in favor of giving them added weight in the decision making calculus. In fact, her rival's desires may be just as impartial, on the present definition. It is not as though the fact that her first-order desires—desires, say, to prevent the torturer from harming persons—reflect her impartiality counts as a further reason for her to act on them on top of such favoring considerations as the fact that by acting on them she might be helping those in most need.
The response I have offered is thus one that even a caricatured subjectivist can sign on to. The subjectivist can grant, if only for the sake of argument, that insofar as essentially psychological facts about the preferrings of the parties involved in a disagreement have any *inherent* normative significance for how one should behave, all preferrings count equally regardless of content. In that precise sense, the preferences, construed as psychological phenomena, are indeed "on a par," and to be given equal weight (if any) in deciding how to act. Accepting the parity thesis so construed does not mean that the subjectivist cannot impartially act on her own preference for morally good outcomes in cases of disagreement.

[33] Similar problems beset other responses on behalf of the subjectivist discussed by Enoch; e.g. the responses he attributes to Stephen Finlay and Ronald Dworkin (fn.310). Both are implausible because they locate the normative difference maker in a feature of the moral case that seems largely irrelevant (how deeply rooted moral desires are, in the case of Dworkin, and their strength, in Finlay's case). For a related response that attempts to distinguish cases based on how 'well-informed' the preferences happen to be and Enoch's discussion of it, see Enoch (2014).

of impartial action, and offer the following alternative. They define impartiality in terms of a notion they take to be primitive—that of the 'importance' of getting one's way in a preference conflict. On their view, impartiality involves compromising when getting one's way is not all that important and refusing to conciliate otherwise. Manne and Sobel argue that in cases of moral disagreement, it is generally quite important that the person with a desire for morally good outcomes get her way, although, the reasons why this is so are left unspecified; and that, therefore, the morally steadfast subjectivist behaves consistently with impartiality's demands. Notably, their account of impartiality does not seem to turn on the actual motivations behind a person's intention to stand her ground. Yet it would be a mistake to focus exclusively on the nature of the outcome a person realizes by standing her ground to the neglect of her motivations in realizing it, as I have tried to show. To see this consider a case where a stranger is at my door seeking refuge from a deadly blizzard. One might say that it was 'important' for me to grant the stranger refuge. However, the fact that the outcome was important to achieve wouldn't settle whether I acted impartially in allowing the stranger in to my home. To settle the question, we would need to know whether, in letting him in, I was motivated to promote our interests equally, or whether I only did it because I needed help completing a household task. It seems wrong to describe me as having acted impartially if what ultimately motivated me was the self-centered desire rather than the identity-independent and characteristically impartial one.

Furthermore, their account appears to yield counter-intuitive results. While the concept of 'importance' remains under-analyzed, it does appear to require a threshold amount of good (or whatever quality bears on importance) to be at stake before standing

one's ground is permitted by the standards of impartiality. The result seems to be that we are forbidden from standing our ground, even for altruistic reasons, where the good to be done by being steadfast is slight. Such a case might arise, for example, when we disagree about which charity to support using our collective funds, even though all of the options under consideration are good ones. On their view, it would be inappropriate for me to insist that we support my favored charity because it is not all that important that I get my way. Whichever charity ends up receiving our funds, a lot of good will be done, and so I should be willing to resolve the conflict by flipping a coin. In response, Enoch (2014) argues (plausibly I think) that, whatever might be said for conciliating in such a case, standing one's ground does not seem prohibited by considerations of impartiality, and, so, Manne and Sobel get the wrong verdict. By contrast, the view I've defended illuminates why it is permissible to stand one's ground even in what might be deemed 'low importance' cases. When I insist that we support what I feel is the best charity despite the fact that your preferred charity seems a very good one, I do not violate impartiality so long as my insistence sincerely aims at the identity-independent, non-arbitrary promotion of wellbeing. Even if I am open to criticism on other grounds, perhaps for being stubborn or too sure of myself, it is not in virtue of any partiality on my part. This, I submit, is closer to the truth than a view that construes impartiality as involving a willingness to conciliate whenever the amount of good to be done by standing ones ground fails to meet some obscure threshold.

These alternative responses discussed by Enoch and others seem to me to be worth rehearsing in order to emphasize a point about the dialectic. In mounting a first-order challenge against the subjectivist, Enoch presents the latter's meta-ethical

commitments as being inconsistent with our ordinary ethical beliefs (namely, the belief that we can and ought to be morally steadfast while also respecting the demands of impartiality). And so it behooves the subjectivist to avoid committing herself to counter-intuitive ethical claims in response to the challenge. A revisionary account of when and why persons act consistently with the demands of impartiality ought to be avoided, especially if a more intuitive response to the challenge is readily available. Otherwise, the subjectivist will be susceptible to the charge of having unfairly "introduce[d] the normative input needed to get the right normative output" (p33), further fueling Enoch's suspicion that the realist has the meta-ethical resources to offer a more satisfying account of why being morally steadfast is consistent with being impartial. Of the accounts available to the subjectivist, I have offered what seems to me to be the most intuitive, and, as I try to argue in the next section, realist alternatives do not enjoy any special advantage over it

IV. Objections and Replies

I address, in this final section, two objections that might be raised against the proposal. According to the first, my view implies, perhaps counter-intuitively, that acting morally in certain special circumstances entails being partial. According to the second, my account of impartiality only seems plausible when we ignore the fact, according to subjectivism, that moral disagreements are *just due to desires* in an important sense.

Starting with the first objection (as it is easier to dispense with) consider a case where three children are at risk of drowning, one of whom happens to be my daughter, and I am in possession of a single life vest. The parent of the other two children thinks I

ought to save one of his, whereas I would prefer to save my own. It seems that I might be morally entitled—perhaps even obliged—to stand my ground and act on the basis of an identity-*dependent* desire to save *my* child. Let us suppose that one is indeed morally entitled to favor one's own when significant benefits to one's loved ones are at stake in a conflict. Despite the fact that the parent at risk of losing both his children and I have competing preferences regarding what I should do, I am, on the view we are supposing, morally entitled to act on my own preference. Based on what I have argued, I would be acting *partially* in standing my ground and following my own preference. My account of impartiality seems to imply that in certain cases of moral disagreement we are morally obliged, or at the very least entitled, to stand our ground despite the fact that doing so entails behaving in a partial manner.

I think what we should say about such a case, if it indeed turns out that morality obliges us to take special care of our own, is that sometimes the value in being impartial conflicts with the demands of (all things considered) morality. This seems to me to be a far more natural thing to say than what the objection assumes we should—namely, that I act *impartially* in saving my child despite doing so simply because she is *my* child; that is, despite the fact that what I care about is for *my* child to survive, as opposed to anyone else's.[34] After all, the value in being impartial consists in treating everyone's interests

[34] Note that it would be unrealistic to suppose that I might be acting out of an identity-*in*dependent desire to act consistently with morality's demands. Acting out of a sense of moral duty instead of affection for one's child may even be inappropriate in a case like this: see Williams (1981, p18). Thus, even if we, suppose, with the realist, that we act impartially whenever we aim to act in conformity with the objective truths about morality, it would not allow us to say that I behave impartially in saving my child from drowning, supposing that my impartiality (or lack thereof) is a function of my *actual* motivations.

equally, as we noted earlier, and I am plainly failing to treat individuals as equal when I would rather promote my child's wellbeing than the wellbeing of others.

In fact, such cases highlight why we cannot simply identify impartiality with having and acting on first-order desires that aim at ends familiarly regarded as morally sanctioned/required; and, relatedly, why the subjectivist cannot simply advert to the fact that her preferences aim at a morally sanctioned/required outcome in response to Enoch's argument. Sometimes the morally sanctioned outcome involves privileging one's own interests over the interests of others. Impartiality, I have argued, consists in having and acting on first order desires that aim at morally sanctioned/required outcomes *of a certain sort*—namely, where the interests of persons are treated equally. If my account of impartiality is correct—if all it takes to be partial is to act on the basis of an identity-dependent desire—then it is a fact, whether or not one embraces subjectivism, that in some special cases of desire-based conflict, standing one's ground as morality requires involves being partial. We can safely say that in such cases of interpersonal conflict, where agent-relative values are at stake, we are *not* required to be impartial. Enoch's argument is not about such cases.[35] It purports to be about cases where we intuitively ought to be impartial *and* morally steadfast at the same time.

Moving on, consider the second objection, which goes to the heart of the present proposal. Fix on the fact that despite all that I have said, moral disagreements, according to the subjectivist, are *just due to desires* in the sense that the fact that parties disagree is nothing over and above the fact that they have conflicting desires. Isn't there some plausibility to the notion that conflicts that are just due to desires in *that* sense are ones

[35] Even his principle demanding that we compromise during desire-based conflicts was supposed to hold *for the most part* and not in every case. See discussion in Section I.

where the parties ought to conciliate?[36] That it does seem plausible is, at any rate, what the realist might insist. I hope that the above discussion has gone some ways towards undermining the notion, but it is perhaps worth saying something more about why it might seem plausible. There is a sense in which conflicts that are 'just due' to conflicting desires are ones where the parties ought to compromise. But it is not the sense that the objection presupposes. Perhaps once we clarify the way of a conflict being just due to desires that is relevant to impartiality, it will become even clearer that the subjectivist does not face the problems Enoch thinks she does.

Sometimes, as was noted earlier, it seems natural to pursue a course of action for no other reason other than the fact that we desire to do so. The principal consideration that strikes us as practically relevant in such cases—the consideration that ultimately motivates us—is that by acting we would be sating our own desires. Conflicts can, of course, arise when parties have such self-regarding aims that cannot be jointly satisfied. In some such cases, the only good to be achieved from acting one way as opposed to another is that either person's desires will be fulfilled. Such a conflict is 'just due to desires' in the sense that the only possible motivation/reason that the parties might have for acting one way as opposed to the other is that their desires will be fulfilled. Moreover, either course of action will realize the very same good: the good of a person's desires being satisfied. When a conflict is *just due to desires* in that sense—that is, what makes the opposing courses of action at all worth taking is the fact that the parties have the

[36] Relatedly, the realist might suggest that that my account of impartiality only seems plausible if we suppose that there is a response-independent fact of the matter about whether we should promote wellbeing in a manner that is non-arbitrary and identity-independent. Perhaps once it is made clear that, according to the subjectivist, wellbeing is just a condition of persons that the morally steadfast happen to have a strong preference for promoting, it seems less than plausible that acting on such a preference, when others prefer differently, suffices for acting impartially.

conflicting wants that they do—compromise is often what is called for. In particular, compromise is called for when the happiness/satisfaction to be gained by parties from having their desires sated is equal on either side of the conflict. For a large part of being a morally attuned agent involves recognizing that insofar as one's own satisfaction is worth promoting, so is the satisfaction of the desires of others. When one sees that one's reason for wanting to ϕ is just that ϕ-ing will sate one's desire to ϕ, one should compromise once one notices that someone else wants one to not-ϕ (and other things are equal). In such cases, the agent confronts a practical conflict not just because two people want opposing things.

Such conflicts occur quite frequently--- a fact that both Enoch and the subjectivist can recognize. When you and I have conflicting movie preferences, arguably the only relevant considerations bearing on which movie we should watch concern our respective desires and the prospective pleasure to be gained by each of us from the different possible outcomes. However, many disagreements between people are not like that. In particular, moral disagreements are generally of the sort where a wide variety of goods (if 'goods' sounds in a worryingly realist register, substitute with 'finally desired ends familiarly regarded as morally good') are at stake other than mere desire satisfaction, goods whose realization is not dependent on the parties' desires being sated.[37] Take, e.g., the good of ensuring that people's fundamental rights are not violated. Achieving that good does not necessarily entail sating my desire for it. Moral conflicts are not 'just due to desires' in

[37] It remains true that, for the subjectivist, taking some feature of a situation to be normatively significant or good simply amounts to having a desire to promote it (or some such), but that should not detract from the point.

the sense that desire satisfaction is rarely the only motivating consideration that gives rise to them. The only plausible normative principle in the vicinity of Enoch's is that one should be conciliatory in a conflict when the only good at stake in acting one way or the other is the good of sating desires, with an equal number of desires on either side of the conflict. By contrast the overbroad principle that one should be conciliatory whenever conflicts are ultimately grounded in competing desires is one we should reject.

What if the realist digs in her heels and insists that without belief in response independent truth, a person's final desires and her reasons for acting, no matter how identity-independent, non-arbitrary, and directed at the general wellbeing, cannot be sufficient to guarantee her impartiality? Are there any positive arguments that might be given to show that realist commitments would make no difference to the impartiality displayed by someone whose actions are motivated by final desires of the relevant sort? The chief things to say in favor of the present proposal have already been discussed—its fit with our moral intuitions about what ought to be done in particular cases, its power to explain our practices of compromise and non-conciliation, and its plausibility given the considerations we attend to in deliberating about what to do in cases of interpersonal conflict. If there is any hope of persuading those who remain unconvinced about the irrelevance of response-independent truth, perhaps it lies in reflecting on cases of the following sort.

First, there is the example of the 'mistaken realist,' which helps show that a person's impartiality does not turn on whether her desires are, in fact, consistent with a mind-independent moral truth (supposing, for the moment, that there are such truths). We can imagine a realist who endorses, say, a purely hedonistic conception of wellbeing and

acts to maximize pleasure in the world without privileging particular persons. Presumably, we should be willing to describe her as impartial, and even if the response-independent normative facts turn out to be such that the good of persons consists in more than just pleasure. A realist mistaken about the response-independent normative facts seems impartial, as long as she is well-enough intentioned—that is, as long as she is interested in promoting the interests of persons equally and non-arbitrarily. Hopefully, even Enoch can agree with this verdict. He may insist, however, that it matters that the mistaken realist *believes* that her desire is backed by a relevant normative truth.

The example of the 'unreflectively decent' suggests that even the belief in a mind-independent moral truth is unlikely to be relevant to the agent's impartiality. We can easily imagine a person who has never considered meta-ethical questions, has no beliefs either way about the objectivity of morality, and yet has strong moral convictions. In cases of moral disagreement, he acts just like our resolute subjectivist. He follows his desire to promote the wellbeing interests of parties equally. I doubt we would regard such a person as partial. His identity-independent, non-arbitrary desire to promote the good of persons seems enough to confirm his impartiality. Enoch may counter that the unreflectively decent are at the very least disposed to endorse realist claims, and it is this unexpressed disposition that contributes to the impartiality of their conduct. To put the point somewhat differently, the subjectivist's explicit belief that there are no response-independent moral truths is what renders her conduct, unlike the conduct of the unreflective, distinctly partial. This thought seems false to me, but I do not currently have much in the way of argument to offer in criticism. I can only speculate about the reasons one might be tempted to think it.

It may be that an agent's meta-ethical beliefs bear on her impartiality *indirectly*—that is, by influencing the desires she ends up acting on—and perhaps this indirect connection tends to be mistaken for a more direct one. Realists are fond of suggesting, after all, that belief in morality's response-dependence is likely to weaken our moral resolve. [38] If, for some reason, this turns out to be true—that is, believing a view like subjectivism makes individuals more likely to be self-interested—then we might have some basis for thinking that subjectivists act partially even when they realize morally good outcomes. For morally good outcomes pursued out of self-interested motives mire the agent in partiality, even on my account. That said, I find it hard to see why the meta-ethical commitments of subjectivism should lead, invariably, to the motivational profile of a thoroughgoing egoist. Even if some such 'egoistic decline' were found to be inevitable, an agent's impartiality would still be grounded in her desires, and not her meta-ethical beliefs. After all, the egoist seems partial precisely because all of her final desires aim at promoting her own interests. We do not need to know anything about her meta-ethical beliefs in order to know that she falls short of the standards of impartiality.

Despite all of this, Enoch may insist that an agent's disbelief in response-independent moral truths undercuts her impartiality in some more direct way, and not just by influencing her motivations. There is not much left for me to say other than that a more direct connection seems unlikely to me. I can only hope the reader shares my sense

[38] Although the claim is rarely defended explicitly, its proponents sometimes appeal to, on the one hand, an alleged first-order normative principle that states something like 'if morality is response-dependent, then anything goes,' and, on the other, a psychological claim that 'if anything goes' then we are more likely to be selfish than other-regarding. Kit Fine (2001) attributes a principle of this kind to Ronald Dworkin. Jonas Olson (2010) thinks that something like the principle is ordinarily assumed and underlies a widely-felt discomfort about expressivism. For discussions of the worry that a denial of moral objectivity entails giving up on our moral commitments, see Smith (1989) and Foot (1978, p. 167).

that once it is revealed that an agent is drawn to states of affairs where the good of persons is promoted equally and without arbitrariness, any questions regarding her impartiality have already been settled. If I am right, we have no basis for questioning the subjectivist's commitment to being impartial in desire-based conflicts, at least absent an argument showing that subjectivists cannot be motivationally drawn to the good of all persons. As far as I can tell, there are no such arguments, and the subjectivist gets to be impartial like everyone else.

Acknowledgments. Special thanks to David Enoch, Erin Miller, and Sarah McGrath for their feedback on earlier versions of the paper and for pushing me with hard objections. Thanks, also, to Johann Frick, Elizabeth Harman, Eric Hubble, Sebastian Koehler, Michael Smith, and Nat Tabris for their comments.

References

Enoch, D. (2011). Taking Morality Seriously. Oxford: Oxford University Press.

Enoch, D. (2014). In Defense of Taking Morality Seriously: reply to Manne, Sobel, Lenman, and Joyce. Philosophical Studies, 168(3), 853—864.

Fantl, J. (2006). Is Metaethics Morally Neutral? Pacific Philosophical Quarterly, 87, 24–44.

Foot, P. (1978). Virtues, Vices, and Other Essays in Moral Philosophy. Berkeley, CA: University of California Press.

Fine, K. (2001). The Question of Realism. Philosopher's Imprint, 1(2), 1—30.

Hurka, T. (1993). Perfectionism. Oxford: Oxford University Press.

Johnston, M. (2001). The Authority of Affect. Philosophy & Phenomenological Research, 63(1), 181—214.

Manne, K., & Sobel, D. (2014). Disagreeing about How to Disagree. Philosophical Studies, 168, 823—834.

McDowell, J. (1998). The Role of Eudaimonia in Aristotle's Ethics. In Mind, Value, and Reality. Cambridge, MA: Harvard University Press.

Olson, J. (2010). The Freshman Objection to Expressivism and What to Make of It. Ratio, 23(1), 88—101.

Pettit, P. & Smith, M. (1990). Backgrounding Desire. The Philosophical Review, 99(4), 565—592.

Scanlon, T.M. (1993). Value, Desire, and Quality of Life. In Nussbaum, M.C., & Sen, A. (Eds.), The Quality of Life. Oxford, UK: Clarendon Press.

Schroeder, M. (2007). Slaves of the Passions. Oxford, UK: Oxford University Press.

Sepielli, A. (2012). Review of David Enoch: Taking Morality Seriously. Notre Dame Philosophical Review. http://ndpr.nd.edu/news/29022-taking-morality-seriously-a-defense-of-robust-realism

Seidman, J. (2004). Two Sides of 'Silencing.' Philosophical Quarterly, 55, 68—77.

Smith, M. (1989). Dispositional Theories of Value. Proceedings of the Aristotelian Society, 63, 89—174.

Williams, B. (1981). Moral Luck. Cambridge, UK. Cambridge University Press

Chapter 3. On Ground as a Guide to Realism

Abstract. According to Fine (among others), a nonbasic factual proposition must be grounded in facts involving those of its constituents that are both real and fundamental. But the principle is vulnerable to several dialectically significant counterexamples. It entails, for example, that a logical Platonist cannot accept that true disjunctions are grounded in the truth of their disjuncts; that a Platonist about mathematical objects cannot accept that sets are grounded in their members; and that a colour primitivist cannot accept that an object's being scarlet grounds its not being chartreuse. The Finean might try to defend these implications, but it generates further problems. Instead, the principle should be rejected. An important upshot is that the principle cannot be relied on to distinguish robust realism from anti-realism about a propositional domain, for the principle obscures ways of taking features to be both real and fundamental.

Keywords: Grounding, Fundamentality, Realism, Anti-Realism, Moral Non-Naturalism

I. A Finean Principle of Ground

Grounding is an explanatory relation standardly introduced by way of examples, together with an account of its theoretical role. To ask what grounds p is to ask what it is *in virtue of* that p holds: 'if the truth that p is grounded in other truths, then they account for its truth' (Fine, 2001, p. 15). While there may be various modes of explanation

(causal, logical, normative), grounding explanation distinctively invokes a metaphysical relation of dependence that, among other things, entails necessary connections between distinct facts (Rosen, 2010; Audi, 2012; Dasgupta, 2014).

Within this framework, it is natural to think some true propositions may be ungrounded (or basic). Perhaps certain microphysical facts do not obtain in virtue of anything else. Say that a propositional constituent is 'fundamental' if it occurs in a true, basic proposition (Rosen, 2010. p. 112).[1] If <electrons have negative charge> is true and basic, then, *inter alia*, the property *being an electron* is fundamental.

Some nonbasic propositions contain a fundamental constituent. A question in the theory of ground is whether there are general principles governing the grounding of such propositions. Kit Fine (2001; 2007) proposes the following principle (call it PRIMITIVE):

> Whenever a constituent occurs in a true, basic, factual proposition and also occurs essentially in some true, factual proposition, then any [full] ground for the latter proposition must contain the constituent.

A constituent occurs *essentially* in proposition p if substituting it with something else can change p's truth value (2001, p. 18). To regard a true proposition as *factual* is to be a realist about it. PRIMITIVE does not apply if a propositional domain is understood in anti-realist terms. Later we shall consider in greater detail what anti-realism about a propositional domain amounts to, but we'll be working primarily with paradigmatically realist views on which the truths in question have all the trappings of robust factuality (e.g. mind-independence). So, to illustrate, if <electrons attract protons> is a true, factual, nonbasic proposition that essentially involves the property *being an electron*, and this

[1] This is intended to be stipulative. Whether fundamentality should be defined in terms of ground is controversial. See Mehta *forthcoming*.

property is a fundamental, fully factual constituent of reality, then, according to Fine, facts involving the property must feature in the grounds.[2]

PRIMITIVE does important theoretical work. It undergirds Fine's criterion for distinguishing an anti-realist characterization of a propositional domain from a realist characterization. Fine's criterion has proven influential in meta-ethics. For example, Dreier (2004) uses it to distinguish quasi-realism from full-blooded realism about moral properties. But, I argue, the criterion is problematic because it relies on a principle that is vulnerable to several dialectically significant counterexamples.

II. The Grounding Principle Should be Rejected

Certain canonical examples of grounding appear in virtually every introduction to ground. These include the fact that true disjunctions are grounded in the truth of either of their disjuncts and that true conjunctions are grounded in their true conjuncts (Correia, 2010, § 6; Rosen, 2010, p. 117; Fine, 2012, §7; Kment, 2014, p. 165). Taking the relata of the (full) grounding relation to be facts, and writing '[p]' for 'the fact that p,' we can state the following principles:

(I) If p then: [p] grounds [p $\vee$ q].

(II) If p and q then: [p] and [q] ground [p $\wedge$ q].

These grounding claims are intended to be uncontroversial. But PRIMITIVE entails that they must be rejected by a logical platonist who takes logical operators to be fundamental

[2] Pautz (2016, p. 485) describes a closely related 'congruence constraint' on ground: 'if a fact involves a certain real item, then the facts which ground that fact also involve that item.'

entities.[3] A platonist who thinks a logical law involving disjunction, say [(p ∨ q) → ¬(¬p ∧ ¬q)], is a true, basic, robustly factual proposition containing disjunction as an essential constituent can't accept (I) if PRIMITIVE is true. For PRIMITIVE requires that she invoke a fact involving disjunction to ground [p ∨ q]. Fine accepts (I) as a canonical example of grounding, and canonical examples should not rule out substantive views in the metaphysics of logic. PRIMITIVE is therefore vulnerable to counterexample by Fine's own lights.

The Finean might respond: it is not clear that [(p ∨ q) → ¬(¬p ∧ ¬q)] can be taken as basic. Suppose we view the logical laws as equivalent to their corresponding disjunctions. The example proposition becomes [¬(p ∨ q) ∨ ¬(¬p ∧ ¬q)]. Such disjunctive propositions are grounded in their true disjuncts. So, the logical platonist who does not dispute the Finean criterion for being fundamental and who embraces standard views about the grounds of logically complex propositions has not yet produced a clear example of a proposition containing disjunction that is true, factual, and clearly basic.[4]

There are several things to say in response. First, nothing in Fine's (2012) pure logic of ground forces the Platonist to accept the formulation of logical law statements as particular disjunctive facts (or even as mere universal generalizations over such facts). Fine's logic of ground does not say what the grounds of a claim involving the logical consequence relation are supposed to be. More importantly for our purposes, it does not matter that there is *a* possible view in the philosophy of logic concerning the logical laws on which the example does not work. There are views that take the logical laws to be

[3] On logical platonism generally, see Rush 2014.
[4] Thanks to an anonymous referee for raising this worry.

strong metaphysical laws, ones involving a primitive notion of logical consequence (cf. Tahko, 2014, p. 242).[5] We can stipulate that a proponent of such a view tells us that the logical laws involving disjunction should be formulated as nomic generalizations: e.g. $\forall\phi$ $\forall\psi$ $\{(\phi \vee \psi)$ logically implies $\neg(\neg\phi \wedge \neg\psi)\}$. Rather than being grounded in their instances, these strong laws explain *why* particular facts like $[(p \vee q) \rightarrow \neg(\neg p \wedge \neg q)]$ hold, while being themselves metaphysically basic (cf. Armstrong, 1983). No basic principle of ground should rule out such a view of the logical laws; nor, as we shall see later, should it classify such views as *anti-realist* about logic simply based on their account of the grounds of particular disjunctive facts.[6]

Suppose, then, the Finean takes a different tack and insists that the Platonist should reject the standard view regarding the grounds of particular disjunctive facts because, by her own lights, the explanation is incomplete: it needs to be supplemented by some fact involving disjunction. Instead of saying [p] grounds $[p \vee q]$ all by itself, we should say that [p] together with, say, $[p \rightarrow (p \vee q)]$, grounds $[p \vee q]$. But this seems implausible and goes against standard views on grounding. Following Fine (2012) and Rosen (2010), it is tempting to explain why [p] grounds $[p \vee q]$ by appeal to the nature of disjunction—someone who understands disjunction's nature should be able to see that the connection holds. So, the explanation of the grounding fact [[p] grounds $[p \vee q]$] may

[5] See also Berto's (2006) 'ontological' formulation of the law of non-contradiction.
[6] There are facts other than those concerning the logical laws that platonists might reasonably regard as fundamental. Tieszen (2011, p. 97), exploring the views of Frege, Husserl, and Gödel, suggests that a key feature of platonistic rationalism about logic is the claim that knowledge of logical truths is object knowledge, and moreover that the 'givenness' of logical objects in rational intuition is not susceptible to further explanation. For such a platonist, [I conceive $p \vee \neg p$] is either itself fundamental or grounded in a fundamental fact involving disjunction. A full-scale defense of such claims and platonistic rationalism more broadly is not presently required.

involve an essential truth about disjunction. But the platonist seems just as able as anyone else to give this type of explanation for (I)'s truth. Moreover, there are costs to expanding the grounds of [p ∨ q]. If a further fact involving disjunction is required to make it the case that p ∨ q— say [p → (p ∨ q)]—then it would seem that [(p ∧ (p → (p ∨ q))) → (p ∨ q)] should *also* be included in the grounds, and more complex logical principles *ad infinitum* (Carroll 1895).[7] Later we shall consider what basis the Finean might have for insisting that the platonist accept the unwelcome consequences of denying (I).

The fact that sets are grounded in their members is another grounding claim often presented as canonical. The existence of the singleton set containing Socrates—i.e., [∃x, x = {Socrates}]—is grounded in the existence of its sole member, or [∃x, x = Socrates] (Shaffer, 2009, p. 375; Fine, 1995, p. 271). Canonical examples should not render trivially false substantive views in metaphysics. But as before, when combined with Fine's principle, the claim rules out mathematical platonism. If the set forming operation {} is a fundamental element of reality, and, accordingly, features as a constituent of some true, basic, factual propositions—say, the axioms of Zermelo-Fraenkel set theory—then PRIMITIVE forces the mathematical platonist to invoke some fact involving {} to ground

[7] Shouldn't the logical Platonist admit that [[p] grounds [p ∨ q]] is itself grounded in some fact involving disjunction, a logical law or a truth about the nature of disjunction? She should, but we still have a violation of PRIMITIVE because the Platonist does *not* invoke disjunction to ground [p ∨ q], a non-basic true factual proposition involving disjunction. See also discussion in section 3.

the existence of Socrates' singleton.[8] In other words, PRIMITIVE places the mathematical Platonist squarely outside of the consensus on ground.[9]

One might be tempted to treat the grounding of facts involving mathematical and logical constituents as a special case, and to think Fine's principle remains safe to use in general theory. But the problems with PRIMITIVE go deeper.

It is very tempting to think [*x* is scarlet] grounds [*x* is not chartreuse]. As before, we assume our essentialist diagnostic for the completeness of grounding explanations: given the natures of the constituents of [p] and [q], does explaining [q] in terms of [p] leave an explanatory gap? Intuitively, someone who understands the nature of the determinate colour properties—*what it is* to be scarlet/chartreuse—and also understands the nature of negation should require no further metaphysical explanation for why [x is

[8] These examples suggest a recipe for generating counterexamples to the principle. There are general ontological principles of the form: <For any entities (that meet condition C), there is some F that is R-related to these entities>. For example, <For any entities (that meet condition C), there is a set containing all and only those entities>; <For any entities (that meet condition C), there is an entity that is their mereological sum>. Given such principles, one might say [there is an F that is R-related to x, y, z, ...] is fully grounded in [x, y, z, ... exist], perhaps together with [x, y, z, ... meet condition C]. And, yet, if either the property F or relation R is fundamentally real, then such views violate PRIMITIVE.

[9] The only other challenge to the principle, as far as I know, is due to Horwich (2007) who provides several purported counterexamples that are dialectically far more controversial. Horwich offers [Mars exemplifies the property of rotating] is grounded in [Mars is rotating]; and [the number of unicorns are 0] is grounded in [there are no unicorns]. But, as Fine (2007, p. 18) points out, it is not clear that a factual and basic constituent has actually been eliminated from the grounds in these examples: '[i]n the first case, the property of rotating occurs 'nominally' in the grounded proposition and only 'predicatively' in the grounding proposition. But in changing its mode of occurrence, it is not thereby eliminated.' Horwich's third case involving an artificial predicate relies on a grounding principle concerning universal generalizations that Fine rejects (p. 19).

The examples discussed here have the advantage of relying on grounding claims that Fine and others accept as canonical. Moreover, the examples presuppose platonist views which ensure that the constituent eliminated from the grounds is both factual and fundamental. The point of raising these cases is precisely that PRIMITIVE renders canonical grounding claims parochial and substantively controversial.

not chartreuse] holds given [x is scarlet].[10] That a thing cannot both be (solidly) scarlet and (solidly) chartreuse is not some further fact that one might fail to grasp despite being fully aware of the nature of *being scarlet* and the nature of *being chartreuse.* So, the grounding explanation seems unimpeachable.

But now consider the colour primitivist.[11] The primitivist (or simple objectivist) treats the colours as fundamentally real monadic properties of things. While there may be standard causal or psycho-physical explanations for *why* a thing is chartreuse, in terms of surface reflectance properties of objects, wavelengths of light, our brain chemistry, and so on, none of these facts provides a complete *metaphysical* explanation of the colour fact. The view is often motivated by appeal to what we know simply from reflecting on the natures of properties like *being chartreuse*. Someone who knew everything there is to know concerning micro-physical and other non-colour facts would still be in the dark about *why* when such non-colour facts obtain they give rise to an object's being chartreuse as opposed to, say, being scarlet. Accordingly (or so the primitivist argues), [y is chartreuse] is plausibly basic or lacking a full metaphysical explanation.

But taking the colours as fundamental and real in no way bars the colour primitivist from thinking that the natures of *being chartreuse* and *being scarlet* preclude a thing's being both. In other words, there is no conflict in thinking both that (1) [y is chartreuse] is robustly factual and lacks a complete metaphysical explanation, and (2) [x is scarlet] grounds [x is not chartreuse]. This is a straightforward counterexample to

[10] There might be open questions involving other modes of explanation (e.g. causal explanation). On the differences between causal and metaphysical explanation, see discussion in fn21.
[11] On primitivist-realism about colour, see Gert (2006).

PRIMITIVE.[12] Whether one accepts colour primitivism or not, no basic principle of ground should rule out such a plausible combination of metaphysical commitments.[13]

III. The 'No Ground-Essence Mediation' Reply

A means of disarming the various examples might seem tempting at this point. The examples all rely on essentialist principles mediating grounding explanations.[14] [[p] grounds [p ∨ q]] is underwritten by or is itself grounded in an essential truth about disjunction. The fact that [∃*x*, *x* = Socrates] grounds [∃*x*, *x* = {Socrates}] is underwritten by the nature of the set-forming operator. We know [*x* is scarlet] grounds [*x* is not chartreuse] because of the natures of *being chartreuse*, *being scarlet*, and *negation* taken together. Suppose one insists that these essentialist truths that allegedly underwrite the first-order grounding claims must be themselves included in the grounds. So, we should say [*x* is scarlet] together with various essentialist facts about the colours and negation jointly ground [*x* is not chartreuse]. Including the essentialist facts in the grounds, we're told, results in a tighter explanation, and the requirement conveniently forces the realist to invoke the features she regards as fundamentally real in explanation.

---

[12] To make things explicit: a true factual proposition is grounded without mentioning one of its constituents—*being chartreuse*—which is fundamentally real (assuming colour primitivism) and which occurs essentially (replacing *being chartreuse* with *being scarlet* in <x is not chartreuse> changes the proposition's truth value).

[13] As before, it shouldn't be necessary to fully motivate colour primitivism for present purposes, though I return to questions of plausibility in section IV.

[14] On essentially mediated grounding explanations, see Rosen (2010, p. 131) and Fine (2012, p. 75). For further discussion of mediating principles, see Bennett (2011), deRossett (2013), and Kment (2014).

Some problems with this strategy were considered earlier in the case of disjunctive facts. But there are more general problems. First, it is perfectly standard to assume, as Fine himself does, that essentialist facts can at least in some cases underwrite grounding explanations (Fine, 2012, p. 75; Rosen, 2015, pp. 130-132). Paradigmatic examples of grounding that involve mediation by bridge principles, essentialist or otherwise, are widely endorsed. Because this reply denies our right to an assumption standardly made in theorizing about metaphysical explanation, it seems *ad hoc*.

Second, it is not at all clear that invoking the essentialist facts as grounds in the cases discussed above results in a tighter explanation. Consider the case of the colours. That a thing cannot both be (solidly) scarlet and (solidly) chartreuse is not some further fact it is possible to be in the dark about despite grasping [*x* is scarlet] and [*x* is not chartreuse]. [*x* is scarlet] is a fact that is partly constituted by *being scarlet*. It is hard to imagine grasping [*x* is scarlet] without grasping at least some of *what it is* to be scarlet, including its chartreuseness-excluding nature. [15] That is why there is no explanatory work for the relevant essentialist fact about *being scarlet* to do that is not already done by [*x* is scarlet] in which scarletness appears.

Third, the reply undercuts the principle's theoretical usefulness. Fine and others rely on PRIMITIVE to distinguish the realist from the quasi-realist (or 'nonfactualist') about a propositional constituent. The Finean test for realism proceeds as follows. Take a nonbasic proposition which involves a constituent whose full-blown reality is disputed

[15] Of course, it is not always the case that an item's essence wholly figures in a fact involving that item. [*x* is water] is grounded in [*x* is h20] and it lies in the nature of *being water* that if a thing is water, it is h20. Knowledge of the relevant essentialist fact about water requires knowing some chemistry. But the colour case is special because plausibly the colour properties have at least partly manifest essences directly involved in colour facts.

but that both sides acknowledge is fundamental—say, the moral property of *being wrong*.[16] Ensure that the chosen proposition's overall factuality is not disputed by the realist and quasi-realist. For example, both might accept <Sam believes lying is wrong> is factual.[17] Now ask: what grounds the truth of this nonbasic factual proposition? Fine's thought is: the realist, and only the realist, invokes facts involving the disputed constituent to ground the target proposition (2001, pp. 17-18).

Fine's test for realism relies in part on PRIMITIVE, which forces the realist to invoke a constituent she regards as fundamentally real in grounding non-basic propositions involving the constituent. But grounding-essence mediation also plays a key role in generating the explanatory asymmetry on which the test depends by allowing the quasi-realist to ground the target proposition *without* invoking the disputed constituent. Fine thinks the quasi-realist (unlike the realist) can fully ground a non-basic proposition involving *wrongness* without mentioning *wrongness*, because the quasi-realist takes the constituent to be ontologically lightweight (2001, p. 18).[18] It is notoriously difficult to say

[16] To simplify matters, we can assume the realist and quasi-realist both regard the feature whose full-blown reality is disputed as fundamental—that is, as featuring in some true, basic propositions.

[17] According to Fine (2001, pp. 16-18), it is critical that we find a proposition that both sides agree is not just *true* but also factual—e.g. Sam believes lying is wrong. The quasi-realist and realist disagree about the factuality of a constituent of the proposition—namely, *wrongness*. But they agree about the proposition's overall factuality. The reason this is important is that if the target nonbasic proposition is one that the quasi-realist regards as *nonfactual*, she *will* invoke the contested feature in explaining it. For example, the quasi-realist might say that [fraud is wrong] is a nonbasic nonfactual true proposition, and is grounded in something like the fact that fraud involves deception and *deception is wrong*. Fine isn't entirely clear on the rules that govern the grounding of nonfactual propositions, but he is clear in requiring that the target proposition that is supposed to generate the grounding disagreement be one whose overall factuality is not under dispute.

[18] The quasi-realist might ground [Sam believes lying is wrong] in [Sam desires not to lie]. Quasi-realism in ethics pairs a deflationary view of moral properties and moral facts with non-cognitivism (or 'expressivism') about moral belief. There is some debate over the coherence of such a position, but we can assume coherence. For our purposes, 'quasi-realism' is meant to be a catch-all term for sophisticated and viable anti-realist positions.

precisely what such talk of ontological lightness amounts to. But roughly, the quasi-realist is happy to say fundamental moral properties and facts exist, but only in some deflated sense—the relevant entities are less than fully real (Blackburn, 1993). For our purposes, the thing to focus on is this: by Fine's own lights, a fact about the nature of the (quasi-real) constituent explains why it can be excluded from grounding explanations of facts involving it.

In other words, an alleged essentialist fact about *being wrong*— namely, it's ontological lightness or less-than-full factuality (whatever this means)—underwrites the quasi-realist's full metaphysical explanation of facts about moral belief in non-moral terms. One might ask the quasi-realist: why does *wrongness* appear in a non-basic proposition that is grounded in facts not involving *wrongness*? This seems like a meaningful explanatory question. And the quasi-realist should advert to an essentialist fact about *wrongness*: once one understands the kind of property *being wrong* is, one sees that one needn't invoke it in explaining factual propositions involving it.[19] It is therefore crucial to the test's effectiveness that the quasi-realist be allowed to exclude from grounding explanations relevant essentialist truths.

So, not only is it contrary to conventional wisdom to suppose that essentialist facts underwriting grounding explanations must feature as grounds themselves, it is not a theoretically viable option for the Finean.

[19] One way to put it is that the quasi-realist invokes moral properties in *higher-order* metaphysical explanation. She appeals to what such properties are like in explaining why they needn't be invoked in grounding certain nonbasic facts involving them that don't also involve the relation of grounding.

IV. Does the Principle Remain Theoretically Useful?

Perhaps PRIMITIVE and Fine's test for realism remain defensible in individual domains. Fine's test has been especially popular in metaethics, inspiring several 'explanationist' approaches to distinguishing moral realists from quasi-realists based on whether a theorist invokes moral properties in explanations of non-basic facts involving such properties.[20] None of the discussed counterexamples to PRIMITIVE involve moral properties. So, perhaps its theoretical relevance is not fully undermined.

But the fact that the test relies on a questionable principle of ground should give us pause. We cannot take for granted that a moral realist who takes *wrongness* to be fundamentally real *must* invoke the property in grounding nonbasic propositions involving it. The general structure of the previous counter-examples to PRIMITIVE suggests a worry. As we have seen, grounding facts like [[p] grounds [q]] can in turn be grounded in or derived from essential truths involving the constituents of p and q without these essential truths featuring in the grounds of [q]. So, suppose a moral realist takes on the following commitments:

> It lies in the nature of *wrongness* that a person *P* is *R-related* to <*x* is wrong> if *P* is disposed to treat x in certain ways: C(*P*, *x*).

Being R-related to <*x* is wrong> may involve *believing x is wrong*. But the nature of the relation is not what matters. The crucial point is that on such a view, it lies in the nature

[20] See Dreier (2004, pp. 36-39). Critics of Fine's test in meta-ethics have wondered whether the quasi-realist should embrace the notion of metaphysical explanation on which the test depends (Asay, 2013). But there is a deeper problem with the test—it relies on a questionable principle of ground.

of *wrongness* that a person can be related to it simply by satisfying a non-moral behavioral condition, C(*P*, *x*), one that never mentions *wrongness*. For example, if *x* is the act of lying, the non-moral behavioral condition might involve *P* avoiding lying or acting resentfully towards those who lie. Accordingly,

(1) [R(Sam, <lying is wrong>)]

a non-basic factual proposition involving *wrongness* can be grounded in

(2) [C(Sam, lying)]

a fact that never mentions *wrongness*, while the grounding fact

(3) [[C(Sam, lying)] grounds [R(Sam, <lying is wrong>)]]

is underwritten by an essential truth about *wrongness*— namely:

(4) It lies in the nature of *wrongness* that $\forall P\ \forall x$ (C(*P*, *x*) $\supset$ R(*P*, <*x* is wrong>)).

The essentialist truth does not entail that *wrongness* is definable as, say, *being the property one is R-related to in condition C*. For all we have said, *wrongness*' nature may still be *sui generis* and robustly factual—the property may be causally efficacious, for example.[21] Nor does the essentialist claim force the realist to deny the fundamentality of facts of the form: *x* is wrong. In general, that it lies in *F*'s nature that *F*-ness is *G* does not

[21] To make her realism about *wrongness* vivid, suppose our imagined theorist thinks [lying is wrong] is what *causes* the non-moral dispositions in agents that result in their being *R-related* to <lying is wrong>. The causal preconditions of some fact do not generally belong in its grounds. [The workers are on strike] may be fully grounded in [the workers are picketing outside their workplace] even if there is a complex causal explanation for why the workers are striking (See Rosen, 2010: § 7; Bernstein, *forthcoming*). Similarly, our imagined realist needn't invoke moral facts involving *wrongness* in grounding [R(Sam, <lying is wrong>)] even if she thinks they are causally related.

entail that facts like [*x* is *F*] obtain *in virtue o*f facts involving *G*.[22] Hence, nothing said so far rules out the possibility that *wrongness* has a robustly factual nature and occurs in some metaphysically basic facts. Fine's criterion would classify such a view as quasi-realist simply for grounding [R(Sam, <lying is wrong>)] without mentioning *wrongness*. But the classification seems mistaken and, at the very least, under-motivated.

Perhaps the described view of *wrongness* is too implausible to be taken seriously. Our imagined theorist who rejects PRIMITIVE is committed to several essentialist facts about *wrongness*: (i) it has a *sui generis* and causally efficacious nature; (ii) its obtaining of some act-types is not to be explained in virtue of anything else; and (iii) its nature is such that persons are related to it simply by behaving in ways characterizable in non-moral terms. No moral realist, as far as I can tell, has explicitly defended such commitments. And this fact can be leveraged in Fine's defense. Fine (2001, pp. 20-21) admits that his test turns on considerations of what is *plausible* for the realist (or quasi-realist) to accept. No argument forces realists to accept the constraints entailed by PRIMITIVE. Rather, Fine's thought is that it would be *implausible* for realists not to accept the constraints. So, it will not do for purposes of refuting the principle merely to state a position ostensibly at odds with it. The candidate view needs to be developed in sufficient detail so we can be sure it is plausible enough to warrant accommodation in general metaphysical theorizing.

It is certainly not my aim to defend the plausibility of a moral realist position with the stated commitments. However, it is worth noting that several self-proclaimed moral

[22] It lies in the nature of {2} that {2} is a set, but [∃x, x = {2}] isn't grounded in facts involving sets. Similarly, it lies in *F*'s nature that *F* and *G* are coextensive does not entail that when a thing is *F*, its being *G* is *what makes it F*. See discussion in Rosen (2015).

realists have resisted being classified as *quasi*-realists under the Finean test.[23] The discussion suggests a direction these theorists may wish to go in rendering their view precise using the machinery of ground and essence.

More importantly, whatever we end up saying about the described view of moral properties, the earlier examples involving logical and mathematical platonism and colour primitivism rely on defensible and paradigmatically realist views in metaphysics. Take the colour case. Although moral realists don't usually characterize their view in terms of essences, this *is* one of the standard ways of characterizing colour primitivism. It is natural for the colour primitivist to say that it is facts about the manifest natures of colour properties like *being chartreuse* that explain why such properties are both real and fundamental. It is equally natural for the primitivist to say that [[*x* is scarlet] grounds [*x* is not chartreuse]] is explained by the manifest natures of *being scarlet* and *being chartreuse*. This is a view about the colours that should be taken seriously and cannot be ignored if we are trying to understand realist positions in metaphysics generally.[24] It is precisely the kind of view that principles of grounding and a criterion for realism need to be tested against.

So, for all I have shown, Fine's criterion for realism may correctly classify standard *moral* views as realist or anti-realist. But the counter-examples to PRIMITIVE reveal that the criterion fails as a *general* test for realism. It fails because paradigmatic realists in various domains can plausibly reject PRIMITIVE.

---

[23] T.M. Scanlon (2014, fn. 62) observes in an intriguing footnote that he would be wrongly classified as a quasi-realist by the Dreier-Fine test. He does not explain why he rejects the classification, however. See also Wedgwood (2007).

[24] For what it's worth, I think the view is not just plausible, there are good reasons for taking it to be true.

ACKNOWLEDGMENTS. Many thanks to Gideon Rosen for feedback on the paper. Thanks, also, to Alisabeth Ayars, Boris Kment, Daniel Kranzelbinder, Sarah McGrath, Michael Smith, Nat Tabris, and an anonymous referee, for their thoughtful objections and suggestions.

References

Armstrong, D. (1983). *What Is a Law of Nature?* Cambridge: Cambridge University Press.

Asay, J. (2013). Truthmaking, Metaethics, and Creeping Minimalism. *Philosophical Studies*. 163, 213—232.

Audi, P. (2012). Grounding: Toward a Theory of the In-Virtue-Of Relation. *Journal of Philosophy*. 109, 685—711.

Bennett, K. (2011). By Our Bootstraps. *Philosophical Perspectives*. 25, 27—41.

Bernstein, S. Forthcoming. Grounding is not Causation. *Philosophical Perspectives*.

Berto, F. (2006). Meaning, Metaphysics, and Contradiction. *American Philosophical Quarterly*. 43: 283—297.

Blackburn, S. (1993). *Essays in Quasi-Realism*. Oxford: Oxford University Press.

Carroll, L. (1895). What the Tortoise Said to Achilles. *Mind*. 4, 278—280.

Correia, F. (2010). Grounding and Truth-Functions. *Logique et Analyse*. 53, 251—279.

Dasgupta, S. (2014). On the Plurality of Grounds. *Philosophers' Imprint*. 14, 1—28.

deRossett, L. (2013). Grounding Explanations. *Philosophers' Imprint*. 13, 1—26.

Dreier, J. (2004). Meta-ethics and the problem of creeping minimalism. *Philosophical Perspectives*. 18, 23—44.

Fine, K. (2001). The Question of Realism. *Philosophers' Imprint*. 1, 1—30.

— (2007). Response to Horwich. *Dialectica*. 61, 17—23.

— (2012). Guide to Ground. In Correia, F., & Schnieder, B. (Eds.). *Metaphysical Grounding: Understanding the Structure of Reality*. 37–80. Cambridge: Cambridge University Press.

Gert, J. (2006). A Realistic Colour Realism. *Australasian Journal of Philosophy*. 84, 565—589.

Horwich, P. (2007). The Quest for Reality. *Dialectica*. 61: 6—16.

Kment, B. (2014). *Modality and Explanatory Reasoning*. Oxford: Oxford University Press.

Mehta, N. Forthcoming. Can Grounding Characterize Fundamentality? *Analysis*.

Pautz, A. (2016). Propositions and Properties. *Philosophy and Phenomenological Research*. 93, 478—486.

Rosen, G. (2010). Metaphysical Dependence: Grounding and Reduction. In Hale, B., & Hoffmann, A. (Eds.). *Modality: Metaphysics, Logic, and Epistemology*. 109—136. Oxford: Oxford University Press

— (2015). Real Definition. *Analytic Philosophy*. 56, 189—209.

Rush, P. (2014). *The Metaphysics of Logic*. Cambridge: Cambridge University Press

Scanlon, T.M. (2014). *Being Realistic about Reasons*. Oxford: Oxford University Press

Schaffer, J. (2009). On What Grounds What. In Chalmers, D., Manley, D., & Wasserman, R. (Eds.). *Metametaphysics*. 347—383. Oxford: Oxford University Press

Tahko, T. (2014). The Metaphysical Interpretation of Logical Truth. In Rush, P. (Ed.). *The Metaphysics of Logic: Logical Realism, Logical Anti-Realism, and All Things In Between*. Cambridge: Cambridge University Press.

Tieszen, R. (2011). *After Gödel: Platonism and Rationalism in Mathematics and Logic*. Oxford: Oxford University Press.

Wedgwood, R. (2007). *The Nature of Normativity*. Oxford: Oxford University Press.

Chapter 4. The Normative Distinction between Moral Realism & Quasi-Realism

Introduction

In the previous chapter, I argued that the leading view on what distinguishes quasi-realism from robust realism about ethical properties is mistaken. "Real" features are not distinguished based on their explanatory role, contrary to the proposals of Fine (2001), Dreier (2004), and others. We need a different account of the distinction. The problem of distinguishing realism from quasi-realism is an urgent one, for understanding the difference between the two views is necessary for determining whether quasi-realists can deliver on their promise of being able to vindicate ordinary factualist discourse about morality without taking on the realist's objectionable metaphysical commitments.

One way forward is to say precisely what minimalism about metaphysical concepts like factuality, truth, property, and proposition amounts to. We can then distinguish realism as a negative claim: the relevant concepts do not work the way the quasi-realist says they do. I am optimistic about this approach to the problem but will not pursue it here. Work by Schiffer (2003), Thomasson (2014), and others provides a reasonably clear account of the minimalist strategy which can be adopted by the quasi-realist.[25]

Here, I explore an alternative means of distinguishing quasi-realism from realism about normative properties. The fact that it may help solve what Dreier calls 'the problem

[25] But see Cuneo (2013).

of creeping minimalism' (roughly, the problem that the realist and quasi-realist seem to make all of the same metaphysical claims about normativity which makes it hard to distinguish their views) is not the only reason for being interested in the distinction.

I. The Distinguishing Claim: REALIST ANGST

The distinction between robust realism and quasi-realism should be cashed out in normative terms. The views can be distinguished based on whether they accept as meaningful a type of assertion understood as a paradigmatic normative claim. Here are some examples of paradigmatic normative claims:

"Lying is wrong"

"People shouldn't steal"

"If you're wealthy, you ought to donate to charity"

Such claims may or may not be distinguishable from:

"Tom believes that lying is wrong"

"According to the law, people shouldn't steal"

"If utilitarianism is true, you ought to donate to charity."

Whether the claims under the two lists differ in their meaning in ways relevant to meta-normative theory is controversial. But it will be important in what follows to leave open the possibility that the two sets of claims are quite distinct, and that only the ones in the first list amount to bona fide normative judgments—that is, judgments involving normative terms being used in the standard normative sense. Accordingly, we shall focus on the claims under the first list. It seems reasonable to suppose that claims like 'lying is

wrong' are paradigmatic normative claims. Plausibly, both robust realists and quasi-realists will treat 'lying is wrong' as a paradigmatic normative judgment. We can set aside worries about whether there is a deeper normative explanation for why lying is wrong. If there is, we can take as paradigmatically normative the fundamental normative claims, the truth of which is not susceptible to explanation.

I defend the view that there is a type of normative assertion that only the robust realist recognizes as a paradigmatic normative judgment, with the quasi-realist insisting that interpreting the claim as normative results in incoherence. The quasi-realist might be open, however, to interpreting the claim as a disguised non-normative descriptive claim. I suggest that it is perfectly sensible for both quasi-realist and realist to insist that there are contexts where normative terms do not make their usual contribution to the meaning of a sentence. What distinguishes quasi-realists from realists is that only the realist recognizes the *standard normative* reading of the relevant type of assertion.

The relevant type of assertion is a conditional claim with a meta-normative antecedent:

REALIST ANGST "If there are no normative facts, then there is nothing that one ought to do"

Similar claims—e.g., 'if there are no normative facts in the *robustly realist* sense, then nothing matters'—have been dubbed the 'Freshman Objection to Expressivism' (Olson 2010).[26] REALIST ANGST looks like a counterfactual assertion. What is expressed seems to be the tautologous proposition that in the (nearest) possible world where are there are no

[26] Similarly, Enoch (2011) suggests that there are metaphysical-normative bridge principles that undermine expressivism. See chapter 2.

normative facts, there is nothing that one ought to do. Our initial question concerns whether the quasi-realist can make sense of REALIST ANGST.

II. Quasi-Realist Accounts of Realist Angst

There is a general puzzle concerning how to make sense, given quasi-realist expressivism, of conditional claims with normative & meta-normative antecedents. Consider: "if lying is wrong, then fraud is also wrong." That looks like a bona fide normative judgment, one that patterns with "lying is wrong." But it isn't obvious what motivation is being expressed when one asserts the former. One option is to say that the motivation expressed is that of disfavoring (being averse to) the state of favoring one but not the other of two non-normatively characterized act-types: lying and committing fraud. Another option, especially relevant for present purposes, is to be a 'contextualist' and insist that normative terms aren't making their usual semantic contribution given the sentential context. Rather than playing a motivation expressing role, the conditional simply amounts to a description of a set of commitments. This is more easily illustrated via the following claim:

"According to utilitarianism, you ought to maximize happiness."

The quasi-realist should not be tempted to go in for a motivation-expressive account of this claim. This is because it is manifestly the case that one needn't be in any motivational state to meaningfully assert the claim. But this does not amount to an objection to expressivism. For the expressivist can reasonably insist that the use of normative terms does not always involve motivation-expression. The idea that, in

general, terms in ordinary language behave differently in different sentential contexts is well-motivated (Burge 1979, Chalmers 2011). So, it would not be ad hoc for the expressivist to say something similar about normative language. The natural thing to say is that the aforementioned claim amounts to a *description* of a set of normative propositions characterized as utilitarian.[27] The thing being described is a particular set of propositions. the utilitarian set, which, we're told, includes the proposition <one ought to maximize happiness>.[28]

We can call such a view CONTEXTUALIST EXPRESSIVISM and the principal thing to be said in its favor is that it preserves the simplicity and intuitiveness of expressivism. It does not try to offer an expressivist semantics for claims involving normative language that do not seem to involve motivation-expression. What marks the view as recognizably expressivist is that the relevant descriptive claims made using normative language do not involve any reference to mysterious irreducibly normative entities in the world.[29] To say that Tom accepts a sentence containing a normative predicate or that he believes a

[27] I am presupposing that the quasi-realist has earned the right to talk of normative propositions. For doubts, see Cuneo (2003). There is a normative proposition, P, just in case one can meaningfully assert "P," and to say that there is a normative proposition P *just is* to say that one can meaningfully assert "P." This does not mean that talk of normative propositions or meanings of "P" is motivation-expressive or that it amounts to talk of a speaker's motivations. Talk of normative propositions and meanings is entirely descriptive—it involves reference to real abstract entities. See Schiffer (2003).

[28] More precisely, drawing on the work of Lewis (1973), Kratzer (1977), and others, the expressivist can say that "ought" is a function that ranks sets of propositions, with the context providing both the 'ordering source' (roughly, the basis for the ranking), and the relevant set of possibilities against which the propositions are ranked, with ought_c(A) being true just in case the proposition A is ranked highest in all the relevant possibilities. In some contexts, the ordering source is just some person's beliefs (with the proposition ranked high if the person believes it in all the relevant scenarios) or determined through membership in some stipulated set (e.g., the set of all propositions true according to utilitarianism). In other cases, the order is determined by the person's own motivational states, where ought_c(A) just in case one plans to A in all the contextually relevant possible scenarios.

[29] See discussion *supra* note 27.

normative proposition does not involve committing oneself to obscure irreducibly normative entities.[30]

Regardless of how quasi-realists choose to solve the problem of conditionals with normative antecedences and whether or not they accept the view I've called CONTEXTUALIST EXPRESSIVISM, there are strong reasons for them to deny that REALIST ANGST can be interpreted as a paradigmatic normative assertion, akin to "lying is wrong." Quasi-realists should insist that the statement isn't *meaningful* if the normative terms are understood as making their usual motivation-expressive semantic contribution.

Admittedly, what counts as a meaningful normative judgment given quasi-realism is not entirely perspicuous. On familiar descriptivist accounts of meaning, a sentence is meaningful if it expresses a proposition, and in order to express a proposition, the various constituents of the sentence need to combine in the right way consistent with and according to their rules of application. E.g. "6 is red" is plausibly meaningless because the rules governing the use of numerical terms do not specify whether color predications involving numerical entities are kosher. This is a controversial example, but the point is just to illustrate a general approach to meaning.

The quasi-realist can and should adopt this standard approach. For she has a natural story to tell concerning why REALIST ANGST interpreted as a paradigmatic normative judgment is meaningless, by appealing to the distinctively expressivist rules governing the meaningful use of normative terms. Consider, e.g., Alan Gibbard's (2003) brand of quasi-realist expressivism. According to Gibbard, "ought" takes act-situation

[30] At most, it involves a commitment to normative propositions, which are like propositions in general, except that they relate to normative language.

pairs described in *non-normative* terms as arguments. The predicate is meaningfully applied to such pairs when the agent expresses a plan to perform the relevant act in the relevant situation. It is crucial to Gibbard's view that plans cannot distinguish act-situations pairs that are non-normatively identical. It follows from *what it is* to have a plan, that one cannot plan to treat two non-normatively identical act-situation pairs differently (see chapter 1; and Gibbard 2003, p91-2).[31]

REALIST ANGST cannot be interpreted as a meaningful normative assertion, because it cannot involve the expression of an appropriate plan. There is no plan that distinguishes circumstances based on their normative features. Accordingly, there is no plan directed at a situation involving there being no normative facts or properties. Contrast: "if you're wealthy, you ought to donate to charity." This is a meaningful normative judgment because it expresses a possible plan: a plan to donate to charity in the circumstance of being wealthy. REALIST ANGST does not pattern with such judgments, given Gibbardian expressivism.

The quasi-realist should suggest that the assertion is analogous to the following:

"If you ought not to do anything, then you ought not to do anything"

But this is not a paradigmatic normative judgment. It does not involve normative terms being employed to express motivation. It does not involve the expression of a plan not to do something in the event of having a plan not to do it. Neither does it involve the

[31] A qualification: Gibbard speaks of identicality in terms of all *prosaicaly factual* features, but I've interpreted him charitably as meaning all non-normative properties. For problems with defining the supervenience base for the normative in terms of prosaic factuality, see chapter 1.

expression of a plan to either plan or not plan to not do anything.[32] Contextualists, at the very least, should deny that the assertion is motivation expressive. It seems far more plausible that no motivation is expressed at all and that this is a rather uninteresting descriptive statement: a set of normative propositions which includes the proposition that you ought not to do anything includes the proposition that you ought not to do anything.

III. Realist Accounts of REALIST ANGST

We can expect the realist, unlike the quasi-realist, to insist that REALIST ANGST, and other claims like it, represent bona-fide paradigmatic normative judgments, roughly, to the effect that *mattering* or *obligation* is dependent on the existence of certain entities. I clarify and defend this line in what follows.

One complication for the realist is that the antecedent seems to describe an impossible state of affairs. Normative facts, whatever they are, seem necessary. If lying is wrong, it is necessarily so. But if normative facts are necessary, then the situation of there being no normative facts is impossible.

There are several means of addressing this issue. First, not all non-naturalist realists believe that normative truths are necessary in the strongest sense. One approach to realism suggests normative truths are *normatively* necessary, but there are metaphysically possible worlds where the relevant truths do not hold (Rosen MS). Normative necessity turns out to be a kind of relative necessity, where, roughly, some P

[32] Even if the expressivist insists on giving an implausible motivation expressive account of the assertion, we can expect the realist to deny that that it is the plan she would be expressing *if* she were a quasi-realist, as I argue in the next section. In other words, the realist will deny that the relevant plan associated with REALIST ANGST is a plan to either plan or not plan to not do anything

is normatively necessary just in case P holds in all possible worlds with the same fundamental normative truths as ours.[33] Scanlon (2014), whose brand of realism is often front and center in debates about the difference between realism and quasi-realism, seems to endorse something like this view. A realist of this sort can perfectly well accept the meaningfulness of the *normative* judgment REALIST ANGST, which is just as normative as "torturing people is something that one ought not to do." It describes the fact that in worlds where there are no normative facts, there is nothing one ought to do.[34]

Even if the realist thinks normative truths are metaphysically necessary, she might stop short of saying that worlds with no normative truths are *inconceivable*. REALIST ANGST would then amount to the judgment that there is nothing that one ought to do in the conceivable (but metaphysically impossible) world where there are no normative facts, which is just like judging that a world where one tells a lie, one does something that one ought not to do.

It is less clear how a realist is to make sense of the claim if she thinks a world where there are no normative facts is inconceivable, akin, perhaps, to worlds where the rules of first order logic do not hold. If such impossible worlds are especially anarchic, this may cast doubt on their being objects of thought or their being systematic ways of thinking about them. For what it is worth, use of (even) counter-logical worlds in

[33] In order to make this view work, it is important to distinguish impure or 'mixed' normative propositions, which involve non-normative descriptions, e.g., <In lying to Tom, Bob did something wrong>, from pure normative propositions, like <lying is wrong>. On the distinction between pure and mixed claims, see Scanlon (2014).

[34] The fact that the claim is tautologous does not mean that the normative terms aren't being used in the standard normative sense. This is critical. The claim *can* be truly asserted without needing to understand the meaning of the normative terms that feature in it on account of its being tautologous. But, crucially, this does not entail that the claim cannot also be asserting using the terms in the standard normative sense.

philosophical theorizing is rapidly increasing, and perhaps the realist will be able to adopt the developing frameworks for thinking about such worlds (Nolan *forthcoming*). Alternatively, the realist might use non-modal metaphysical vocabulary to articulate the relevant thought expressed by REALIST ANGST. E.g., she might say that there being something that one ought to do *normatively depends* on their being (robust) facts of the matter about what one ought to do, where this normative dependence relation is primitive. If the realist does choose to go this route, she should again draw a parallel with other kinds of normative claims: e.g., that donating to charity being what one ought to do *normatively depends* on one's earnings surpassing some threshold.

The relevant point for present purposes is that the realist sees a parallel with standard normative claims that the quasi-realist does not. The quasi-realist cannot say that the judgment that there being something that one ought to do *normatively depends* on their being normative facts is relevantly like the judgment that donating to charity being what one ought to do normatively depends on one's income surpassing some threshold. The latter judgment plausibly involves the expression of a motivational state—perhaps a plan to donate to charity in the circumstance of earning more than the relevant threshold. By contrast, there is no plan that takes as its object a situation described in irreducibly normative terms: there being normative facts.

IV. Objections & Replies

Doesn't the quasi-realist hope to earn the right to assert all of the same normative/metaphysical claims as the realist; and, if so, shouldn't the quasi-realist try to vindicate REALIST ANGST based on her expressivist semantics as a bona fide normative judgment? The answer is no. The point of quasi-realism is not to be able say everything

the realist says without the realist's ontological commitments. Rather, it is to establish that *ordinary* claims about morality involving the concepts of truth, property, proposition, and the like are amenable to deflationary treatment, so that one can correctly assert such claims as 'there are moral facts' without taking on heavy ontological commitments. The vindication of *every* intuition purportedly had by robust realists would be an odd theoretical goal to have as a quasi-realist.

Moreover, there is a perfectly sensible interpretation of the realist available to the quasi-realist that does not involve accepting that REALIST ANGST as stated expresses a meaningful normative judgment. Suppose the realist has genuine (non-normative) properties in mind as un-instantiated in certain possible scenarios, when she judges that in those scenarios there is nothing that ought to be done. Call the relevant properties "Q-properties." The realist's mistake might be to think Q-properties are *normative* properties. Accordingly, she may well be expressing a robust normative intuition – that Q properties are normatively significant – in embracing REALIST ANGST. However, she would be mistaken in trying to use normative vocabulary to refer to such properties, if quasi-realist expressivism is true. A world where no Q-properties are instantiated may be a world where nothing is good. The quasi-realist can accept, reject, or suspend judgment on whether the realist is right about Q-properties mattering. What she shouldn't do and needn't do despite the concession to Q-properties possibly mattering from the normative point of view is think REALIST ANGST can be meaningfully read as a paradigmatic normative assertion, based on the standards meanings of the terms.

Throughout, I have assumed, with Gibbard and others, that if quasi-realist expressivism is true then motivations directed at situations described in irreducibly

normative terms are impossible. This assumption was part of the defense of the claim that the quasi-realist sees an asymmetry between standard normative claims (which are motivation-expressive) and REALIST ANGST. But the assumption might reasonably be questioned. After all, the quasi-realist is happy to speak of normative properties and propositions as real and irreducible entities, ones that are in some sense guaranteed once we have normative language.[35] Why can't we have desires or plans that are sensitive to such entities?[36] I'm inclined to think that minimalism about properties and propositions rules out their being admissible objects of our motivational states. But suppose it turns out to be possible to be motivationally sensitive to such entities. Fortunately, the overall argument need not rely on the assumption that such motivational states are impossible. The relevant point is that quasi-realists, like Gibbard, have denied that the expression of such motivational states would suffice to render meaningful the associated (pseudo) normative claims; and it is very plausible as a quasi-realist to think that such entities *shouldn't* be the objects of desire even if they can be. In other words, even if it turns out that it is compatible with quasi-realism for there to be motivations expressed via the judgment 'there is nothing that one ought to do if there are no normative properties' or 'there being something that one ought to do normatively depends on their being normative propositions,' a sensible quasi-realist should deny the truth of such judgments. The wrongness of torturing someone for the fun of it does not *normatively depend* on there being normative propositions or properties. It normatively depends on the nature of

[35] See discussion *supra* note 27.

[36] The relevant kind of sensitivity involves a *de dicto* rather than *de re* desire to do what's right, or a desire that is *based on* the property of being right. The quasi-realist, at the very least, should deny that one ought to *de dicto* desire to do what is morally right. See discussion that follows. But see, Lerner *forthcoming*. Lerner is mistaken in thinking that the cases he discusses provide evidence for a rational *de dicto* desire to do what is right.

pain and the nature of individuals who can feel pain. So, if it is indeed the case that motivational sensitivity to quasi-real entities is possible, then the quasi-realists should reject the truth of REALIST ANGST. To sum up, either the quasi-realist denies that REALIST ANGST is meaningfully interpreted as a paradigmatic normative claim or she denies its truth. Either way, REALIST ANGST offers a convenient means of distinguishing quasi-realism from robust realism.

Conclusion

I have argued that realists and quasi-realists can be distinguished based on whether they recognize a normative difference between judging "if you're wealthy, then you ought to donate to charity" and "if nothing has the property of being what ought to be done, then there is nothing that one ought to do." This way of distinguishing realists from quasi-realists is normatively interesting. For one, it reveals that according to quasi-realism the property of *being good* is not itself something that we can meaningfully judge to be good. Normative judgments do not involve normative appraisal of situations described in irreducibly normative terms. But it may well be a distinctive realist commitment to include *being good* in the extension of "good."

References

Burge, T., 1979, "Frege and the Hierarchy," *Synthese,* 40: 265—281.

Chalmers, D., 2011, “Propositions and Attitude Ascriptions: A Fregean Account,” *Noûs*, 45(4): 595-639

Cuneo, T., 2013, “Properties for Nothing, Facts for Free? Expressivism's Deflationary Gambit,” *Oxford Studies in Metaethics,* 6: 223-51.

Fine, K., 2001, “The question of realism,” *Philosopher’s Imprint*, 1(2), 1–30

Gibbard, A., 2003, *Thinking How to Live*, Cambridge, MA. Harvard University Press

Kratzer, A., 1977 “What ‘must’ and ‘can’ must and can mean,” *Linguistics and Philosophy*, 1:337-355.

Lerner, A., *forthcoming*, “The Puzzle of Pure Moral Motivation”. Oxford Studies in Metaethics 13.

Lewis, D., 1973, *Counterfactuals*, Blackwell: Oxford.

Nolan, D., *Forthcoming*, “Impossibility and Impossible Worlds,” *in* (Otavio Bueno & Scott Shalkowski, eds.,) *The Routledge Handbook of Modality*, Routledge Press: New York.

Olson, J., 2010, “The freshman objection to expressivism and what to make of it.” *Ratio*, 23(1), 88–101

Rosen, G., (MS), “What is Normative Necessity?,” available at: https://www.academia.edu/9159728/Normative_Necessity.

Scanlon, T.M., 2014, *Being Realistic about Reasons*, Oxford: Oxford University Press

Schiffer, S., 2003, *The Things We Mean*, Clarendon Press

Thomasson, A.L., 2014, *Ontology Made Easy*, Oxford University Press

Part 2. Meta-Normative Foundations for Law

Chapter 5. Juridical Obligations at the Edge of Legality

Abstract. Judges decide cases by appeal to rules of general application they deem to be law. If a candidate rule resolves the case and is, *ex ante* and independently of the judge's judgment, the law, then the judge has a legal obligation to declare it as such and follow it. That, at any rate, is conventional wisdom. Yet the principle is false—a rule's being law or the judge's believing it to be law is neither necessary nor even sufficient for a judge being legally obliged to follow it. The principle's falsity is especially apparent in so-called hard cases, where the line between legal and non-legal rules is obscure. Moreover, judges have authority to disregard law in hard cases not because moral (or non-legal) obligations trump legal obligations. Rather, the law itself circumscribes its own authority. The implications for legal philosophy are significant; for one, a theory of juridical norms can be developed independently of the precise boundaries of legality.

Keywords: Legality, Judicial Norms, Hard Cases, Positivism, Anti-positivism, Epistemic Indeterminacy

Introduction

Suppose a judge were to become persuaded that the central debate within analytic jurisprudence—the debate between positivists and anti-positivists about the nature of law—should be decided in favor of the positivist (the choice of theory is irrelevant). Should this make any difference to how she should rule in legal cases? It is tempting to think so based on the following sort of argument. Positivists and anti-positivists disagree about how to draw the line that separates legal from non-legal rules and there are cases where their competing theories come apart. As between two incompatible rules that a judge might rely on to decide a case, say one that prohibits capital punishment as cruel and unusual and another that permits it, which rule happens to be law independently of the judge's judgment turns on whether the positivist or anti-positivist is right. It is platitudinous that judges have at least a professional duty (a duty deriving from what the law requires of judges) to remain faithful to the law—that is, to follow a rule and declare it to be law if it *is* in fact law.[1] Accordingly, what judges legally ought to do turns on the resolution of the philosophical debate.

[1] HLA Hart, *Essays on Bentham: Studies in Jurisprudence and Political Theory* (Clarendon Press, 1982) 158-59:

> When a judge of an established legal system takes up his office he finds that though much is left to his discretion there is also a firmly settled practice of adjudication, according to which any judge of the system is required to apply in the decision of cases the laws identified by specific criteria or sources. This settled practice is acknowledged as determining the central duties of the office of a judge and not to follow the practice would be regarded as a breach of duty one not only warranting criticism but counter-action where possible by correction in a higher court of appeal.

This quick argument and others like it often crop up in legal theorizing, but they rest on a fundamental misunderstanding of judicial duty.[2] That judges have a duty to remain faithful to the law is indeed platitudinous, but the platitude has been misinterpreted. Judges are *not* obliged to follow the law in all cases, even if there is a fact of the matter about what the law is *ex ante*. While it is a familiar enough notion that judges may have authority to create new law in cases where the law is undetermined (in other words, a rule's being law, *ex ante,* may not be a necessary condition for its being declared as such by a judge), the claim that judges have legal authority to *ignore* determinate law or rules they believe to be law in individual cases rings of heresy.[3] Nevertheless, it is true that judges have such authority and precisely in the sort of cases where competing philosophical theories of law come apart. Moreover, judges have this authority not because, as some have suggested, *moral* obligations (or some other non-legal species of obligations) trump legal obligations. Rather, the law itself circumscribes

[2] Lon Fuller finds common ground with Hart in thinking that it is the province of general jurisprudence to clarify what 'fidelity to law' amounts to with a precise characterization of law's nature: 'if we do not mend our ways of thinking and talking [about the nature of law] we may lose a "precious moral ideal," that of fidelity to law…. one of the chief issues is how we can best define and serve the ideal of fidelity to law.' LL Fuller, 'Positivism and Fidelity to Law: A Reply to Professor Hart' (1958) 71 Harvard Law Review 630, 630-2. It is easy to multiply examples of this common sentiment. See discussion in section III. Our principle question is whether *perfect* fidelity to law could even be a *legal* ideal, let alone a precious moral one. As it turns out, fidelity to law is not the kind of legal (or moral) ideal that is best served by demarcating the precise boundaries of legality.

[3] That judges have authority to make new law where there are 'gaps' in the law is a popular notion amongst legal positivists. See, e.g., J Raz, 'Legal Principles and the Limits of Law' (1972) 81 Yale Law Journal 823, 847–8. Timothy Endicott has argued that when the law runs out, judges are legally authorized to decide the case (and plug gaps in the law) by appeal to extra-legal, moral considerations. See T Endicott, 'Raz on Gaps – The Surprising Part,' in LH Meyer, SL Paulson, and TW Pogge (eds), *Rights, Culture and Law* (OUP 2003). The claim defended here goes further and should prove more controversial: in cases where there is *no* gap in the law but the law is uncertain given the evidence, judges can lawfully ignore what they deem to be law.

its own authority.[4] In the relevant class of cases, how a judge should rule simply does not turn on what the law is. The implications for legal philosophy are significant; for one, a theory of judicial obligation can and should be developed independently of the precise boundaries of legality.[5]

The dispute within analytic jurisprudence is an example of one where rival theories of law conflict in just the range of cases where legality's normative significance for judges is slim to non-existent. But there are other, more parochial disputes that similarly involve competing claims about *what the law is* in cases where legality is not what matters—for instance, the dispute over the precise legal significance of original intent in American constitutional jurisprudence.[6] My argument generalizes to these other disputes: their resolution does not bear on how a judge legally ought to rule.[7] I explore the generalized version of the argument in other work. The present focus is on showing that the debate between positivists and anti-positivists, on its own terms, entails the untethering of judicial duty from legality.[8]

---

[4] The argument relies on various ecumenical assumptions about the nature of distinctively legal obligation. See discussion in fn. 60. Roughly, if an agent, A, is legally obliged to perform action-type φ in circumstance C, then (at the very least) there is a rule with the property of being law that obliges A to φ in C.

[5] The argument relies on the idea that the legal concept is vague at the margins for epistemic reasons, even if it has precise applications conditions as theorists standardly assume. On epistemic theories of vagueness, see T Williamson, *Vagueness* (Routledge 1994); R Sorensen, *Vagueness and Contradiction* (Oxford University Press 2001).

[6] Compare S Scalia, 'Originalism: The Lesser Evil' (1989) 57 Cinn. L. Rev. 849, with J Balkin, *Living Originalism* (HUP 2011) ch1.

[7] There are good debates to be had about what judges should do in the relevant range of cases. But these are not debates about what the law is. See discussion in sections III and IV.

[8] Analytic jurisprudence provides a convenient starting point for the analysis because it is easy to motivate the non-obviousness of law in cases where philosophical theories of law come apart, and law's non-obviousness plays a critical role in the overall argument. See discussion in L Murphy, 'Concepts of Law' (2005) 30 Austl. J. Leg. Phil. 1; B Leiter, *Naturalizing Jurisprudence: Essays on American Legal Realism and Naturalism in Legal Philosophy*, (Oxford University Press 2007), 180-4. A further reason that the analytic dispute provides a useful argumentative lens is

In section one, I defend a standard account of analytic jurisprudence as aimed at articulating the precise application conditions of the concept expressed by predicates like "is law" and "is legal." Positivists and anti-positivists disagree about the general conditions that a rule must satisfy in order to fall under the legal concept.

In section two, I argue that even card-carrying positivists and anti-positivists should acknowledge that cases where the theories come apart in their implications for which rules count as law are 'hard' cases, in the sense that it is highly non-obvious what the law is in such cases—the line that separates the legal from the non-legal rules is obscure given the evidence. To use an idiom familiar from first-order legal theory, *reasonable persons can disagree* about the law in such cases. I offer two arguments for why our confidence in our judgments about legality in the relevant range of cases should be low, one from persistent theoretical disagreement about law and the other based on the totality of facts that determine application conditions for semi-technical concepts.

In section three, which forms the bulk of the discussion, I argue that judges are not legally obliged to follow preexisting law in hard cases, and, moreover, that they are not so obliged from any other normative perspective (say, that of morality). My general strategy is to show that a rule requiring strict conformity to the law in hard cases could not *itself* be law, whether positivism or anti-positivism is true. The rule is neither conventionally embraced in modern jurisdictions nor is it a morally good rule for a legal system to adopt. The rule's non-conventionality follows from a *de re* as opposed to *de dicto* interpretation of the platitude that judges should strictly follow the law and from the actual content of judicial oaths and constitutional rules. The rule's moral inferiority

---

that the central claim can be motivated using positivistic as well as anti-positivistic theories of law.

follows from the evaluative irrelevance of a rule's legality when it's legality, though genuine, is reasonably uncertain. I conclude that judges can and should decide hard cases based entirely on features of rules other than their legality. The conclusion is a thoroughly practical one with ramifications for the actual practice of judging. It entails, for example, that judges can, consistent with their professional duties, ignore rules they *believe* to be law in hard cases, even if their beliefs turn out to be correct.

In section four, I address possible objections to my view, including that it misguidedly (a) affords judges the authority to misrepresent what the law is; (b) overlooks the virtues of a principle of judicial decision-making that demands strict conformity to preexisting law; and (c) permits judges to frustrate the reasonable expectations of persons held accountable under legal rules.

In section five, I conclude by commenting on the argument's implications for legal philosophy generally and on the light it sheds on Ronald Dworkin's classic but ultimately misleading discussion of hard cases.[9]

I. Reasonable & Persistent Disagreement About the Concept of Law

Analytic jurisprudence has traditionally been characterized as an investigation into the application conditions of a concept—the concept expressed by juridical

[9] R Dworkin, 'Hard Cases' (1975) 88 *Harvard Law Review* 1057. Dworkin uses 'hard case' to refer specifically to difficult cases that arise before courts involving plaintiffs and defendants, whereas I use the term more generally to refer to any situation where a judge has to decide what the law is on some question, but the law is obscure given the evidence.

predicates like "is law" and "is legal."[10] Just as we judge certain rules to be those of etiquette or morality, we classify some as "legal rules" or "rules of law."[11] The philosophical task is to specify those general features of rules that determine whether a rule falls under the legal concept. It might be helpful to think of concepts as akin to abilities.[12] To possess the concept <law> is to be able to discriminate between entities in the world that differ in some respect—whichever respect it is that distinguishes all legal from non-legal rules.[13] We exhibit this ability in our dispositions to attribute legality to rules.

We take for granted that the legal concept has a public character in that it is shared by different agents. At the very least, judges and other important legal actors are assumed to possess a singular concept of law. This is a significant methodological assumption that might reasonably be questioned.[14] It may turn out that individuals use

[10] As Raz puts it, 'It is part of the self-consciousness of our society to see certain institutions as legal. And that consciousness is part of what we study when we inquire into the nature of law.' J Raz, *The Authority of Law* (Oxford University Press 1979) 221; J Raz, 'Two Views of the Nature of the Theory of Law' in J Coleman (ed), *Hart's Postscript: Essays on the Postscript to The Concept of Law* (OUP 2001): 'our aim is to explain the concept as it is, the concept that people use to understand features in their own life and in the world around them'; B Leiter 'Naturalism and Naturalized Jurisprudence' in B Bix (ed), *Analyzing Law: New Essays in Legal Theory* (Clarendon Press 1998); R Alexy, 'Legal Certainty and Correctness' (2015) 28 Ratio Juris 441. S Shapiro, *Legality* (Harvard UP 2011) 16-22, offers a useful discussion of the importance of conceptual analysis.

[11] We can construe rules as functions mapping circumstances to outcomes or actions. See also discussion in fn. 60.

[12] M Dummett, *Seas of Language* (OUP 1993). The argument does not turn on whether concepts are construed instead as abstract objects or mental representations. For alternative views of concepts, see J Fodor, *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (MIT Press 1987); C Peacocke, *A Study of Concepts* (MIT Press 1992).

[13] Writing '<P>' for the 'concept expressed by 'P'.'

[14] See, e.g., B Leiter, 'Legal Realism, Hard Positivism, and the Limits of Conceptual Analysis' in J Coleman (ed), *Hart's Postscript: Essays on the Postscript to The Concept of Law* (Oxford University Press 2001) 355. For a critical take on conceptual analysis as a general philosophical methodology, see M Johnston and SJ Leslie, 'Concepts, Analysis, Generics and the Canberra Plan' (2012) 26 Philosophical Perspectives 1.

juridical terms to express different concepts and, so, fail to converge in their judgments of legality. The possibility of conceptual divergence gains traction from widespread disagreement about law amongst experts.[15] One response to this type of worry emphasizes that persons can be mistaken about their own concepts. It is one thing to have a concept, to deploy it in thought and talk, and quite another to have true beliefs about it. For instance, one might have the concept <good> as evidenced by one's ability to reliably distinguish good acts from bad ones, while having false beliefs about the features of acts one responds to in deploying the concept. One might falsely believe that all good acts are happiness maximizing, when in fact they have some other property in common (say, that of being defensible from an impartial perspective).[16] The legal philosopher can reasonably take people's self-understanding and their dispositions to call rules "law" with a grain of salt, while nevertheless treating these dispositions as *prima facie* evidence of the contours of a shared concept.

Assuming a shared concept of law, the theoretical task is to state in general terms the conditions that a rule must satisfy to fall under it. Specifying the concept's application conditions is an importantly different task from saying which rules in individual cases and jurisdictions satisfy the relevant conditions. The concept <bachelor> applies to individuals who are male and unmarried, but knowing the concept's application conditions is not by itself informative as to whether any given individual is a bachelor.

---

[15] On the pervasiveness of legal disagreement, see R Dworkin, *Law's Empire* (Harvard University Press 1986) 42-46; B Leiter, 'Explaining Theoretical Disagreement' (2012) 76 University of Chicago Law Review 1215.

[16] On conceptual analysis in ethics, see F Jackson, *From Metaphysics to Ethics* (Oxford University Press 2012).

What unifies positivists about law is the view that the legal concept's application conditions are entirely social—the concept applies to a rule if and only if it has a certain social property, roughly, one having to do with what people have historically said, believed, done, or intended to do in a community.[17] On John Austin's view, for example, legal rules are those that have been prescribed by a sovereign (or other official) who is habitually obeyed in the community and whose commands are backed by the threat of sanction.[18] H.L.A Hart famously proposed an alternative and significantly more complex characterization of the social property essential to law.[19] For Hart, a rule's legality consists in its being part of a broader system of hierarchically organized rules that are habitually followed in the community: a system that includes 'primary' rules which govern conduct in particular circumstances and 'secondary' rules which specify methods for making primary rules. Hart's theory remains thoroughly positivistic insofar as it rejects any moral preconditions on a rule's legality. The relevant property of legal rules is a complex social property that refers, in part, to a broader system of rules habitually obeyed. More recently, Scott Shapiro has advanced a view on which legal rules are ones that have been incorporated into a collective plan adopted by members of a community

[17] For a related characterization of the social property relevant to positivism, see M Greenberg 'How Facts Make Law' (2004) 10 Legal Theory 157. We can ignore the difference between inclusive and exclusive positivism in what follows. Inclusive positivists allow that moral features of rules sometimes play a role in determining whether the rule is law but only if some social fact independently makes the moral features relevant: e.g. a convention of treating punishments as legal only if they are humane. See discussion in S Shapiro, 'The "Hart-Dworkin" Debate: A Short Guide for the Perplexed' in A Ripstein (ed), *Ronald Dworkin (Contemporary Philosophy in Focus)* (Cambridge University Press 2007) 22-55.

[18] J Austin, *The Province of Jurisprudence Determined* (first pub. 1832, Cambridge University Press 1995).

[19] HLA Hart, *The Concept of Law* (2d ed., Clarendon Press 1994) 99. For a helpful discussion of Hart's view, see Shapiro (n10) 84-85.

with the aim of solving large-scale problems believed by the planners to be morally important but that need not in fact have moral significance.[20]

Anti-positivists oppose a purely social characterization of the legal concept's application conditions. Their distinctive claim is that rules fall under the concept only if in addition to their social properties, they also have some *moral* or broadly *normative* features. Ronald Dworkin's brand of anti-positivism is perhaps most well-known.[21] Dworkin argued that applying the concept <law> to a rule necessarily involves *interpreting* social practice, and interpretation involves more than just figuring out which rules community members follow.[22] Social interpretation for purposes of applying the concept consists partly in moral judgment. It involves characterizing individuals and their practices in a way that casts them in their morally best light even if the characterization is not entirely faithful to actual intentions and conduct. In other words, recognizing a rule as law involves appreciating not just that it is *actually* followed by community members but that it is a morally good rule to follow in light of community practices.[23] More recently, Mark Greenberg has suggested that legal rules are just those morally good rules whose moral status depends in part on our social practices.[24]

---

[20] Shapiro (n10) 118- 213. Shapiro describes the 'moral aim' of legal systems as 'the rectification of the moral defects associated with the circumstances of legality.' ibid 214.
[21] Dworkin (n15) 40-46. Dworkin's 'semantic sting' argument against positivism should be construed as suggesting that the legal concept's application cannot simply be a function of linguistic usage and social facts. The argument's upshot remains a distinctively anti-positivistic account of the concept's application conditions.
[22] 'Interpretation of works of art and social practices, I shall argue, is indeed essentially concerned with purpose not cause. But the purposes in play are not (fundamentally) those of some author but of the interpreter. Roughly, constructive interpretation is a matter of imposing purpose on an object or practice in order to make of it the best possible example of the form or genre to which it is taken to belong.' ibid 52.
[23] One way to make sense of the view is that the Dworkinian thinks the closest morally good approximation of the rule actually followed by members of the community is law.
[24] M Greenberg, 'The Moral Impact Theory of Law' (2014) 123 Yale Law Journal 1118.

It should be noted that no one seriously doubts that positivists and anti-positivists have identified genuine worldly phenomena that our concepts could be tracking. There are rules with the relevant social properties (we might call them "S-rules") as well as rules with social and moral properties (call them "M-rules").[25] What Hart and Dworkin disagree about is whether a rule falls under the concept <law> if and only if it is an S-rule as opposed to an M-rule.

While we have put the central issue in terms of the concept of law, some writers prefer to frame the philosophical debate in terms of the property of *being law* that some rules have and others do not.[26] The central question for these theorists is whether a rule's social properties alone or its social and moral properties together explain *why* it instantiates the legal property.[27] The benefits of this alternative construal are somewhat obscure and one reason to favor the conceptual framing is it is broadly consistent with how legal philosophers have construed their own project.[28] Moreover, it simplifies the

[25] There are skeptics about rules, functions, and other abstract objects, but their skepticism is not particular to the legal domain. This paper assumes, quite harmlessly, that there such abstract entities.

[26] G Rosen, 'Metaphysical Dependence: Grounding and Reduction' in B Hale & A Hoffmann (eds), *Modality: Metaphysics, Logic, and Epistemology* (Oxford University Press 2010) 109-136; A Marmor, 'Farewell to Conceptual Analysis (in Jurisprudence)' in W Waluchow & S Sciaraffa (eds), *Philosophical Foundations of the Nature of Law* (OUP 2013).

[27] Rosen (n26) 110: 'One of the aims of jurisprudence is to identify in general terms the facts in virtue of which the legal facts are as they are. One distinctive claim of legal positivism is that the grounds of law are wholly social…. Antipositivists typically maintain that pre-institutional moral facts often play a role in making the law to be as it is.' I have no objection to this construal so long as properties are abundant, with a corresponding property for every well-functioning predicate.

[28] See sources cited (n10). A more radical take on the dispute portrays anti-positivists, in particular, as concerned not so much with characterizing the ordinary concept of law, but with motivating its revision. There is an available reading of the later-Dworkin as a revisionist. R Dworkin, *Justice for Hedgehogs* (Harvard UP 2011) ch19. For a positivistic brand of revisionism, see F Schauer, 'The Social Construction of The Concept of Law' (2005) 25 Oxford Journal of Legal Studies 493. Of course, it wouldn't be a Hart-Dworkin *debate* if while positivists were trying to characterize a pre-fixed concept, anti-positivists were in the business of trying to modify

overall argument considerably to frame issues in terms of the legal concept. The argument can, however, be reformulated in terms of the legal property, and so there is no need to settle the question of ideal framing for present purposes.[29] The alternative construal is acceptable so long as one concedes that what practical reasons there might be to follow or obey legal rules are wholly explained by the social and/or moral features of such rules—the features in virtue of which a rule gets to be legal.[30] Happily, the claim that legality's normative significance is derivative in this way is by and large embraced by those who have defended the property-based view.[31]

II. Competing Theories of the Concept Diverge Extensionally in Hard Cases

It is easy to come up with versions of positivism and anti-positivism that end up saying very different things about which rules in a jurisdiction count as law. But *plausible* versions of positivism and anti-positivism of the sort we shall focus on end up

---

our conceptual repertoire. For this reason, I take revisionist characterizations of the debate to define away the debate. I return to this issue in section V.

[29] Although our argument can be reformulated in terms of the property-based view, there are several reasons for preferring the conceptual framing. For one, it would be a mistake to think that the property of *being law* has a robust nature like the property of *being scarlet* or *being H20*. On this point, see B Bix, 'Conceptual Questions and Jurisprudence' (1995) 1 Legal Theory 465, 468: 'The problem is that talk of ''essences'' and the ''nature'' of items does not fit as comfortably with human artifacts and social institutions as it does, say, with biological species or chemical elements.' There is some sense to be made of legal property talk on an abundant conception of properties, where all it takes for there to be a property is for there to be, corresponding to the property, a well-functioning predicate in a language.

[30] A chief reason for avoiding talk of a distinctive property of being law is that it is potentially misleading. There is a temptation to think that when rules acquire the property they acquire a *sui generis* normative significance that isn't fully explained by the rule's social and/or moral features. Scott Hershovitz in his recent work has warned against this error. S Hershovitz, 'The End of Jurisprudence' (2014) 124 Yale Law Journal 882.

[31] It is the rare theorist who denies the reductive claim. On Hans Kelsen's non-reductive view, for example, the property of being law appears to be an irreducible normative property with its own distinctive normative significance. H Kelsen, *Pure Theory of Law* (2nd edn, University of California Press 1960). See discussion in A Hagerstrom, 'Kelsen's theory of law and the state' in K Olivercrona (ed), *Inquiries into the Nature of Law and Morals* (1953) 267.

agreeing for the most part on the legal concept's extension. For instance, such views will commonly acknowledge the U.S. Constitution as well as properly enacted statutes and judicial decisions as sources of law in the United States.[32]

There are several reasons for this theoretical convergence. To begin with, the social properties that interest positivists tend to be morally significant due in part to their social character. If we plan to drive on the right side of the road, the rule prescribing driving on the right has the social property of being a rule we plan to follow. But it also happens to be one we morally ought to obey and precisely because of our planned conformity to it. On a larger scale, the plans, intentions, and practices of community members have moral significance for how individuals should behave.[33] And so, it is not surprising that the social properties of rules that interest positivists and the moral properties that interest anti-positivists—like that of being a morally good rule to follow—tend to 'co-travel.' In other words, S-rules are often enough *also* M-rules. Since everyone agrees that <law> is a morally important category, any plausible version of positivism will ensure that the concept tracks a social property that is at least in general (although not necessarily) morally significant in the sense of providing those who are subject to the law some moral reason to comply with the law's demands.[34]

On the flip side, any plausible version of anti-positivism will make sure not to wed legality too close to what is morally ideal. Legal rules that fall well short of being morally best are all too familiar. The highly retributive criminal laws of the United States

---

[32] See Greenberg (n17) 162-163 describing paradigmatic examples of law as common ground between the positivist and anti-positivist.
[33] On how coordinating conventions give rise to normative obligations, see D Lewis, *Convention* (Harvard University Press 1969); GI Mavrodes, 'Conventions and the Morality of War (1975) 4 Philosophy & Public Affairs 117.
[34] See discussion in Shapiro (n10) 181-186.

are markedly unjust.[35] Yet the injustice inherent in a rule that prescribes life-imprisonment based on a third criminal conviction regardless of its gravity did not prevent it from being one of the state laws of the United States.[36] Anti-positivists can accommodate the moral inferiority of law by, among other means, lowering the threshold of moral acceptability that socially embraced rules must meet to qualify as legal rules. By making a rule's legality turn on a more fine-grained moral property that comes in degrees, like the axiological property *of being good to degree n*, as opposed to binary deontic ones, like *being what ought to be obeyed*, the anti-positivist can make the moral preconditions on legal rules as weakly demanding as required to render the view extensionally adequate. Alternatively, anti-positivists accommodate the occasional immoral law by making the necessary moral preconditions of legality applicable to entire *systems* of socially embraced rules based on the overall good they do rather than to individual legal rules within such systems.[37]

What are the sorts of cases where the theories come apart? They involve rules that bear all the social markers that interest positivists but are so debased as to not even meet minimal standards of moral acceptability—cases involving grotesquely bad rules or systems of rules. The Fugitive Slave 'laws' of antebellum America which required marshals to return runaway slaves to their owners and the putative laws of Nazi Germany

[35] On the harshness of American criminal law, see JQ Whitman, *Harsh Justice: Criminal Punishment and the Widening Divide between America and Europe* (Oxford University Press 2003).
[36] Cal. Penal Code § 667 (West 1994).
[37] See discussion in R Alexy, 'Legal Certainty and Correctness' (2015) 28 Ratio Juris. 441, 444-445: 'not every injustice, but to be sure extreme injustice is not law'; and more generally in G. Radbruch, 'Vorscule der Rechtsphilosophie' in A Kaufman (ed), *G. Gesamtausgabe* (C.F. Muller, 1990) 154.

are often offered as examples of such debased yet apparently legal rules.[38] The positivist will insist that the Nazis had laws even though their scheme of social organization failed to meet minimal standards of moral decency. By contrast, the anti-positivist will suggest that while it is true that German legal officials might have believed at the time that the Nazi state's pronouncements were law, they believed falsely. She might accuse the positivist of confusing what individuals mistakenly called "law" with genuine law.[39] After all, the laws of most well-functioning states seem to be a force for moral good. Since we learn the concept through examples of legal systems that are plausibly believed to be morally well-functioning, it is far from obvious that moral conditions are not built into our legal concept.[40]

What seems true of such cases that separate positivists from anti-positivists is that they are *hard* cases in the sense that it is non-obvious whether the rules in question do (or did) count as law—for instance, the rule requiring the return of a runaway slave. This is because neither the positivist nor the anti-positivist line on morally wicked 'legal' rules seems decisive; and, so, as Liam Murphy writes, most people who have spent time

[38] On the question of Nazi laws, see F Haldemann, 'Gustav Radbruch vs. Hans Kelsen: A Debate on Nazi Law' (2005) 18 Ratio Juris 162. See also Shapiro's discussion of the fugitive slave act, (n10) 23, and the laws of the Soviet Union, *ibid* 16, 49.

[39] Additional arguments favoring anti-positivism are discussed in section V. The folk understanding of law is perhaps closer to the positivist line. But compatibility with the folk conception is hardly decisive, given that the folk conception of law is generally error-prone. It comes apart wildly from the *juridical* concept of law. For example, it only takes a semester of law school to disabuse students of the commonplace that all laws are to be found in statute books and constitutional texts.

[40] By analogy, consider Quine's example of learning the term 'gavagai' as a non-native speaker of the language arunta. WVO Quine, *Word and Object* (MIT Press 2013) 23–72. When a native speaker asserts 'gavagai' while pointing to a rabbit, they could be picking out a whole host of entities, including the rabbit, various rabbit parts or stages, food, an animal, and so on. We learn the legal concept by way of its application to S-rules that also tend to be M-rules, and it is hard to be certain that the public concept does not track moral features of rules.

thinking about the question 'feel the pull of both ways of thinking about the boundary between law and morality.'[41]

Even if one accepts one of these views as correct, one should still take the class of cases where the theories diverge to be hard cases. In other words, one should think that neither the positivist nor the anti-positivist line on morally inadequate rules is obviously correct (and, accordingly, that neither positivism nor anti-positivism is obviously correct).[42] There are at least two arguments for why our confidence in our judgments concerning the legality of morally inadequate rules should be low. The first is an argument from disagreement. Experts who have thought long and hard about whether the concept of law applies in borderline cases (as in the case of Nazi 'law') continue to disagree. Such persistent disagreement is sometimes thought to reveal that the concept of law is underdetermined. There might be grounds for a pessimistic meta-induction from the fact that no account of the necessary and sufficient conditions for the concept's application has persuaded all competent users of legal terms to the conclusion that there are *no* such conditions.[43] But, more plausibly, persistent disagreement about the concept may be explained *not* by its indeterminacy but by the non-obviousness of facts that

[41] Murphy (n8) 7.

[42] A common reaction amongst those first exposed to the dispute is to find the positivist line more compelling. The ensuing discussion offers several reasons to doubt one's instinctive reactions to such cases.

[43] The concept <law>, the pessimists suppose, is simply indeterminate when it comes to the sort of cases philosophers argue over. Leiter (n8). In this, the legal concept might be like the concept <bald>. There is no fact of the matter concerning whether persons midway through the process of losing hair count as bald or not, and this is because we have not even tried to achieve consensus on how to use the predicate "is bald" in borderline cases. However, the analogous conclusion in the legal case strikes one as over-hasty (and overly pessimistic). The legal case is not like the case of baldness given that there is no analogous disagreement about what baldness consists in or how it applies in borderline cases. On the 'conviction' that there are determinate principles governing law's application, see HLA Hart, 'Definition and Theory in Jurisprudence' in HLA Hart, *Essays in Jurisprudence and Philosophy* (OUP 1983) 21.

govern its application in hard cases.[44] Indeed, anyone who takes the concept of law to be sufficiently regimented for there to be a fact of the matter as to whether positivists or anti-positivists are right about law—and, accordingly, a clean and determinate edge of legality—is under pressure to admit the hardness of cases where the theories come apart. For one needs an explanation for persistent theoretical disagreement amongst epistemic peers, and the elusive character of the relevant conceptual constraints looks to be the only one available.

The other argument for taking cases that separate positivism and anti-positivism to be hard cases invokes controversial yet defensible assumptions concerning the determinants of the legal concept's application conditions. Similar assumption underwrite Timothy Williamson's epistemic theory of vagueness.[45] Williamson suggests that even putatively indeterminate concepts like <bald> or <thin> have determinate application conditions, and their apparent indeterminacy in borderline cases is only due to our inability to know the complex totality of facts that determine the precise application conditions.[46] The application conditions of <bald>, for example, are highly sensitive to the overall pattern in the dispositions of speakers to classify persons using "bald" across a range of cases. Whatever one thinks of Williamson's proposal generally, it is very tempting to think in the case of a semi-technical concept like <law>, one that has been

[44] The literature on the significance of peer disagreement is extensive. For an overview, see D Christensen, 'Disagreement as Evidence: The Epistemology of Controversy' (2009) 4 Philosophy Compass 756. The argument presented here sides with those who think peer disagreement provides significant reason to lower one's credence in a disputed proposition. A Elga, 'Reflection and Disagreement' (2007) 41 Noûs 478, 486–487. But see T Kelly, 'Peer Disagreement and Higher Order Evidence' in R Feldman and T Warfield (eds), *Disagreement* (Oxford University Press 2010), 111–174.
[45] T Williamson, *Vagueness* (Routledge 1994).
[46] ibid 231-37.

self-consciously understood by judges and other experts as applying determinately in hard cases, that its application conditions are determined not just by any particular individual's usage of "law" but by our *collective* dispositions (or the dispositions of a broad range of experts) to classify rules using legal vocabulary in epistemically ideal situations—that is, given perfect information about the non-legal features of rules.[47] If the concept is sensitive in this way to the judgment of competent speakers across many different cases, and it certainly seems to have been understood by legal experts in this way, then any single individual's understanding of law is likely to be inchoate for one has no way of surveying and knowing the overall pattern of linguistic dispositions in all its details.[48] We have reason to be wary of our conceptual intuitions, especially in cases where competent users of legal terms disagree, for these are precisely the kind of cases where the overall pattern of linguistic dispositions is likely to make a difference to which concept is expressed by legal predicates.[49]

---

[47] Plunkett offers an account of how the legal concept might be fixed by patterns in our linguistic dispositions drawing on work by Frank Jackson and David Chalmers. D Plunkett, 'A Positivist Route to Explaining How Facts Make Law' (2012) 18 Legal Theory 139, 181-186.

[48] Admittedly, it is controversial that the legal concept's application conditions are sensitive in this way to patterns in linguistic dispositions. For instance, ES Anderson and RH Pildes suggest an alternative 'expressivist' account of the concept in 'Expressive Theories of Law: A General Restatement,' (2000) 148 University of Pennsylvania Law Review 1503, on which applying the concept might involves expressing a positive non-cognitive attitude towards a rule, like a desire to follow it. While it is true that the current argument presupposes a view of concepts that is controversial, a version of the argument goes through without this commitment. So long as one thinks that the meta-semantic question about how the legal concept's application conditions are fixed is itself *hard*, as one should, and that its resolution bears on the positivism vs. anti-positivism dispute, then one should think that cases where the theories come apart are hard cases.

[49] The sensitivity of conceptual application conditions to slight differences in linguistic usage is the basis for Williamson's claim that whatever we might *believe* about the boundaries of vague concepts, we cannot have *knowledge* of the boundaries, because knowledge must satisfy a 'safety' requirement: if S knows that p, then there are no nearby possible worlds in which S believes p, but p is false. Given a concept's sensitivity to slight variations in usage, there are nearby possible worlds where one's beliefs about the boundaries are false. Williamson (n45) 230-4.

While the above arguments needn't be decisive, especially given that the second turns on a view of conceptual determination one might reject, they provide reasonable grounds for taking cases where positivism and anti-positivism come apart to be hard cases.[50] Notably, this is a charitable line to take towards the philosophical dispute. It does not beg the question by assuming there is no way of resolving the dispute. The very weak assumption is that neither positivism nor anti-positivism is obviously true. It may be reasonable to adopt either theory or suspend judgment on which is correct. Alternatively, as Bas van Fraassen has argued in the case of competing empirically adequate scientific theories, some attitude that falls short of full-blown belief (like 'acceptance') may be appropriate towards any one of a range of plausible theories of the precise boundaries of legality.[51] For our purposes, we need not settle the question of what to *believe* given the law's obscurity in hard cases. Our concern is solely with the *legal* ramifications of our epistemic situation with respect to legality.

III. Judges Have Legal & Moral Permission to Ignore Legality in Hard Cases

Let us trace the dialectic so far. The anti-positivist and positivist disagree about whether the concept <law> tracks the positivist's preferred purely social property of rules, something like the property of *being conventionally embraced* (call it the S-property and rules that have it S-rules) or the anti-positivist's moral-cum-social property,

[50] Of course, hard cases of law arise not just due to its being non-obvious which general theory of the legal concept is correct. A case can also be hard for more mundane reasons, like semantic ambiguity, conflicts in rules, and evaluative complexity. See discussion in sections IV and V.
[51] B van Fraassen, *The Scientific Image* (Oxford University Press 1980).

say that of *being morally good to follow given social conventions* (call it the M-property and rules that have it M-rules). Since S-rules are often enough also M-rules, such as when the relevant social conventions involve conformity to morally good rules, the theories by and large converge in their accounts of the concept's extension—that is, on which rules count as law in a jurisdiction and which do not. Cases where they come apart are *hard* cases: it is not obvious whether the rules in question count as law.

It is tempting to think that figuring out precisely whether the legal concept tracks M-rules or S-rules matters for judicial decision making. After all, judges bear responsibility for saying what the law is. To know how she should decide a hard case where S-rules are not also M-rules, a judge needs to understand the legal concept's application conditions. Echoing this line of reasoning, the opening chapter of Shapiro's defense of positivism begins with the promise that addressing the philosophical question has implications for how judges should decide cases.[52] Hart himself was explicit in his hopes that a clearer understanding of the legal concept might inform judicial practice.[53]

---

[52] Shapiro suggests that 'many of the most pressing practical matters that concern lawyers' turn on the philosophical dispute. Shapiro (n10) 25-9. Shapiro goes on to claim that analytic jurisprudence bears on the correct theory of constitutional interpretation. ibid 220. Farrell argues that Shapiro ultimately fails to deliver on his claims because his theory entails considerable downstream disagreement about the content of law. IP Farrell, 'On the Value of Jurisprudence. Book Review: Legality' (2011) 90 Texas Law Review 187, 221-223. Both sides presuppose law's basic normative significance for judges.

[53] Hart, 'Definition and Theory' (n43) 21-2; Hart, 'Essays on Bentham' (n1) 158-9; Hart (n19) 209: 'If we are to make a reasoned choice between these concepts, it must be because one is superior to the other in the way in which it will assist our theoretical inquiries.' There are places where Hart seems to suggest his theory of law is not going to be helpful for judges, but Hart's reasons for pessimism concern his belief that the law is full of gaps and a theory of judicial discretion in cases where the law is silent should obviously be developed independently of enquiries into law. See HLA Hart, 'Discretion' (2013) 127 Harv. L. Rev. 652, 657. As I go on to argue, the central issue is epistemic not metaphysical—it is not necessarily the actual indeterminacy of law that renders legality insignificant in hard cases, it is our uncertainty about law from our evidential standpoint.

Meanwhile, Dworkin defended various controversial developments in American constitutional law by appeal to his anti-positivism.[54]

The suggestion that the philosophical enterprise might be relevant to judges is at least prima facie plausible, but it has seen opposition.[55] Legal practitioners often express disbelief that philosophical insight into law's nature could have any impact on case outcomes. But the skepticism tends to be question begging.[56] Richard Posner, for example, insists that the philosophical debate has no significance for legal cases because he is skeptical that there is any essential content to the concept <law> for the legal philosopher to discover, rendering the analytic enterprise futile; and, relatedly, Posner is skeptical that philosophical argument could persuade judges to adopt some particular approach to legal decision-making.[57] As it turns out, one can grant legal philosophy its assumptions—e.g., that competing theories of the concept do not always converge in their implications; that there is a fact of the matter as to which is correct; that there are compelling philosophical arguments for both positivism and anti-positivism that a judge might reasonably be persuaded by—and yet still maintain that the analytic project's relevance to a judge's professional obligations is vastly overblown.

---

[54] Dworkin (n15) 90: 'Any practical legal argument, no matter how detailed and limited, assumes the kind of abstract foundation jurisprudence offers.... Jurisprudence is the general part of adjudication, silent prologue to any decision at law.'

[55] Farrell (n52) 221-223; R Posner, *Law and Legal Theory in England and America* (Clarendon Law Lectures 1996) ch1.

[56] Posner (n55) 2-5; Murphy (n8) 4-5: 'We can say that the issue of the nature of law, its boundary with morality, need have no impact on the outcome of legal cases…. If the issue of the nature of law did affect the outcome of legal cases, more people, especially more lawyers, would be interested in the topic and continuing disagreement about it would be considered a problem.' Murphy does not explain *why* the issue should have no impact on the outcome of legal cases. The cited lack of interest in the problem might reveal only a general failure to properly appreciate the significance of the philosophical debate.

[57] Posner (n55) 2-5.

I believe the conception of the judge's professional role presupposed within legal philosophy is mistaken. It is true that judges incur duties constraining their conduct by virtue of their legal office, oaths, and obligations to the public, as well as, simply, as a matter of justice. But the relevant duties do not include an obligation to always treat those rules as law that *are* law, and certainly not in hard cases where it is not obvious whether the concept applies to a candidate rule. Call the principle obliging judges to track the concept perfectly 'the principle of legality':

> LEGALITY: If a rule in a case is, *ex ante* and independently of the judge's judgment, the law, then the judge has an obligation to follow it and declare it "law."[58]

The aim is to show that hard cases of the sort discussed in the previous section serve as counter-examples to the principle: a judge does not have reasons, internal to the law or otherwise, to comply with LEGALITY.[59] In the light of hard cases, one can show that LEGALITY is neither an S-rule nor an M-rule— in other words, the principle is neither embraced in contemporary jurisdictions as a matter of social convention nor is it a morally good rule for a legal system to embrace.[60]

---

[58] There is an epistemic version of the principle which the discussion also shows to be false.

> $\text{LEGALITY}_{\text{EPISTEMIC}}$: If a judge *believes* that a rule in a case is, *ex ante* and independently of the judge's judgment, the law, then the judge has an obligation to follow it and declare it "law."

I have left open whether the nature of the obligation referred to in LEGALITY and $\text{LEGALITY}_{\text{EPISTEMIC}}$ is distinctively *legal* or moral. The conclusion I argue for is that LEGALITY is false on either reading of the deontic modal.

[59] Another way to put it is that LEGALITY is neither a legal nor a moral rule. There is no legal obligation to follow legal rules in every case and there is no moral obligation either.

[60] I have assumed that for there to be a distinctively *legal* obligation to perform some action (say, e.g., follow the law in every case), there must be a rule *with the property of being law* that prescribes that action as obligatory. In other words, to be a legal obligation is in part to be an action deemed obligatory by a legal rule. This view leaves open whether legal obligations are robustly normative like moral obligations, a species of moral obligation, or purely 'formal' like the rules of chess. It may be that what it is to be a legal obligation *just is* to be an action deemed

First, LEGALITY is not an S-rule, or at the very least we lack grounds for supposing that it is conventionally embraced. One might be misled into thinking otherwise by the oft-repeated platitude that judges have a duty to strictly follow the law. Not just legal officials but the general public as well tend to endorse this common sentiment and it is tempting to infer from it both a general expectation that judges will conform to the principle *and* actual conformity by judges.[61] But the inference would be mistaken. The fact that it is widely accepted in modern jurisdictions that judges should obey the law does not entail that strict conformity with LEGALITY is what is generally expected or desired of judges. The assertion "judges should always follow the law" is ambiguous. On one interpretation, the speaker has particular rules in mind that happen to fall under the legal concept in her jurisdiction: for instance, rules stated in the American Constitution forbidding cruel and unusual punishment and unjust takings, canons of statutory

obligatory by a legal rule, and what it is to be a legal rule *just is* to be a rule conventionally followed. The nature of legal obligation is a topic of considerable controversy, but the argument of this paper is intended to be ecumenical. The argument is compatible with an even stronger requirement on legal obligation, where in order for agent A to be legally obliged to φ in C, there needs to be a first-order legal rule, R_1, obliging A to φ in C *and* second-order 'backing' rules (R_2, $R_{3,...}$) with the property of being law obliging A to follow the lower order rule (R_2, e.g., obliges the judge to follow R_1…).. In other words, in order for a first-order legal rule to create a robust and distinctively legal obligation, it may need appropriate 'backing' from higher-order legal rules that oblige the relevant actor to follow the first-order rule precisely on account of its legality.

I have suggested that rules be construed as functions from circumstances to actions, and one might worry about how to make sense of a rule that invokes a normative notion like requirement or obligation. We can conceive of rules specifying obligations as function mapping circumstances to action-types that are *ruled out*, which is just to say that in following such rules, we never perform the relevant type of action in the appropriate circumstance. An agent following such a prescriptive rule makes sure that she only performs an action in the relevant circumstance that isn't of the forbidden type. In other words, obligatoriness can be cashed out in terms of the nature of the function or its role in our practical deliberation and action.

[61] For data on public beliefs about and evaluation of judging, see JM Scheb and W Lyons, 'The Myth of Legality and Public Evaluation of the Supreme Court' (2000) 81 Social Science Quarterly 928, 938: 'Americans who believe the Court bases its decisions on [legality] are more likely to render positive assessments of the Court.'

interpretation, prior judicial pronouncements, and so on.[62] She might use the concept <law> to pick out these specific rules and believe that judges should strictly follow them quite apart from whether (and for what reason) the rules fall under the legal concept. For instance, her belief that American judges should follow rules stated in the Constitution may be based on her reverence for the values enshrined in it. But a quite different (and more complex) thought one might express in making the assertion "judges should always follow the law" is that judges should follow all and only those rules that fall under the concept <law>, without having specific rules in mind.

The familiar *de re* / *de dicto* distinction marks these two different senses of an assertion and applies also to beliefs and desires. I desire *de re* to read a novel by Virginia Woolf if I have a particular novel in mind—say, 'To the Lighthouse'—one that I believe satisfies the description: *a novel written by Virginia Woolf*. I desire it *de dicto* if I desire to read any novel that satisfies the relevant description. Knowing whether a novel was in fact written by Woolf is quite important if your goal is to satisfy my *de dicto* desire. On the other hand, if you know precisely which novel I have in mind when I express the *de re* desire, you need not know whether it satisfies the description I use to refer to it in order to help me achieve my ends.

The generally embraced platitude that judges should follow the law supports LEGALITY only if individuals *de dicto* desire or expect that judges will follow the law. But the *de dicto* interpretation is implausible. A person's reasons for requiring judges to follow rules that happen to be legal, like constitutional prohibitions against restricting free speech or religion, will likely implicate a wide range of moral and prudential

[62] U.S. Const. am. 5.

concerns. Since how judges rule on the bench is a matter of some political importance to individuals, it seems unlikely that persons would want case outcomes to be determined by a pre-established legal concept whose precise contours may or may not line up with their preferred policy preferences.

A less cynical reason for favoring the *de re* over the *de dicto* interpretation of public expectations concerns the opacity of our legal concept. It is simply not clear, even to those who have considered the issue carefully, what is entailed by the *de dicto* claim that judges should follow the law in cases where M-rules and S-rules come apart. In the case of the Fugitive Slave 'laws,' for instance, it is uncertain whether the rules that happened to be declared "law" by judges really were. In attributing a *de dicto* desire to the public, we would be committed to thinking that individuals want judicial behavior to be constrained in ways whose consequences are uncertain, and by a concept whose precise contours most of us perceive rather dimly. In general, we should be wary of attributing to persons *de dicto* desires with less than transparent aims because it renders persons and their interests hard to rationalize.[63]

While the above considerations are far from decisive, they shift the burden of justification on the proponent of the *de dicto* interpretation of the widely-held platitude in support of LEGALITY. No legal philosopher, as far I one can tell, has ever addressed the issue. The *de re* interpretation (along with several others)[64] of the platitude that judges

[63] The argument for why persons would be hard to rationalize if they had such desires comes later. It suffices for the present view to suppose that even if persons think judges should obey the law *de dicto*, their commitment to this principle is probably very weak.

[64] Even if the *de dicto* interpretation were correct, this wouldn't necessarily support LEGALITY. It might be that what individuals have in mind when they assert 'judges should follow the law' is something like a *prima facie* or *pro tanto* obligation to follow the law *de dicto* as opposed to a decisive obligation. There are other readings of the platitude that similarly undercut support for LEGALITY. It isn't clear whether the assertion 'judges should follow the law' is a *generic* claim

should follow the law strongly undercut support for LEGALITY as a conventionally embraced and supported rule.

My suspicion is that upon considering hard cases, where it is unclear whether the legal concept applies to a rule, most of us would be disinclined to think judges should follow a rule simply because of its legality (in part because this is clearly the rational response to hard cases, as we shall see in a moment). It seems reasonable to assume, at least tentatively, that what the public expects is for judges to always follow *paradigmatic* legal rules—those familiar jurisdiction-specific rules that both anti-positivists and positivists can and should agree fall under the concept.[65] There is no room to infer from this *de re* expectation a further general expectation or practice of judicial conformity to the principle of LEGALITY.

Perhaps grounds for thinking LEGALITY is conventionally embraced (and, accordingly, an S-rule) might instead be found in judicial promises and constitutional rules. But the actual content of judicial oaths and constitutions in modern legal systems suggests otherwise. The American oath of office for judges includes the words: "I … swear (or affirm) that I will administer justice without respect to persons, and do equal right to the poor and to the rich, and that I will faithfully and impartially discharge and perform all the duties incumbent upon me …. under the Constitution and laws of the United States." The oath speaks of duties incumbent under the constitutions and the laws, but leaves the specific content and character of these duties unspecified. Whether these

(like 'Mosquitoes carry the West Nile virus') as opposed to a universal generalization. A claim that the generic judge or judges generally should follow the law *de dicto* would not imply that all judges *in every case* have an obligation to follow law *de dicto*.

[65] For example: well established constitutional rules, detailed and clearly stated statutory requirements, and so on.

duties include conformity to the principle of LEGALITY is entirely underdetermined. Moreover, while Article VI of the U.S. Constitution specifies that judicial officers "shall be bound by oath or affirmation, to support this Constitution," it is hardly obvious that support for constitutional legal rules entails upholding all rules that fall under our legal concept. There is simply no clear statement of a rule, in the Constitution or elsewhere in American law, that requires judges to fetishize the concept <law> by following only those rules that fall under the concept simply for falling under the concept.[66] The American oath of office and legal background are not at all unusual in this respect compared to the laws of other jurisdictions.[67]

It is one thing to repudiate the notion that LEGALITY is conventionally embraced, and quite another to show that no ideal legal system or public would adopt the principle given the possibility of hard cases. It is the normative question we turn to now: whether there is *reason* to want judicial behavior to be constrained by the legal concept in all cases. This relates to the question: could LEGALITY be an M-rule? For even it is not *in fact* followed as part of an established convention, the principle's moral characteristics—like

[66] Rule 2.2. of the American Bar Association's Model Code of Judicial Conduct notes that judges shall uphold and apply the law fairly and impartially. But this is far from a clear statement of LEGALITY. The comments acknowledge the appropriateness of each judge bringing her on personal philosophy of interpretation to bear on the law. The only form of judging the code explicitly prohibits is deciding what the law is based purely on personal taste.

[67] No *de dicto* promise to obey all and only legal rules is to be found in the British, Australian, and Swedish versions of the judicial oath. While the Indian Constitution requires judges to 'uphold the Constitution and the laws,' the use of the plural form suggests reference to a plurality of legal rules rather than to any rule with the property of being law. The German oath requires judges to 'fulfill the duties of an honorary judge… faithfully to the law' and adds a requirement 'to serve only truth and justice.' It would be question begging to assume that fulfilling the duties of the judicial office and fidelity to law involve respecting law *de dicto*. As I go on to argue, it is far more reasonable to interpret such requirements *de re*, in terms of paradigmatically legal/constitutional rules. Could there be a jurisdiction that explicitly incorporated LEGALITY *qua* legal rule? Of course. But it suffices for present purposes that most jurisdictions are not like this and that LEGALITY is not an essential feature of legal systems.

its being a good principle for judges to follow—might render it legal. On Dworkin's theory, for example, the fact (if it is one) that LEGALITY is a morally good principle for judges to follow would give us reason to charitably interpret our practices—the platitude that judges should follow the law, judicial oaths, constitutional rules, and so on—in a way that supports LEGALITY, even if our practices do not in fact involve commitment to such a principle. And so, we have not yet ruled out the possibility that LEGALITY may be binding on judges *qua* legal rule on anti-positivistic grounds.

There are practical consequences to a rule's being called "law" by judicial officials. Rules declared "law" are enforced by the state and obeyed by others. The moral question is: how should judges decide which rules to call "law" in hard cases given the significant consequences of their speech acts? One way to address this question is to entertain the perspective of an institutional designer—someone with power and ability to decide which general principles will govern judicial decision-making.[68] The principle of LEGALITY is an example of a decision criterion the institutional designer might impose on judges in hard cases, one that tasks judges with the responsibility, in deciding which rule to follow, of tracking the legal concept whatever its precise constraints turn out to be.

The inferiority of LEGALITY as a decision criterion—or higher-order rule for deciding cases where first-order legal rules are hard to figure out—can be brought out by considering how we might weigh the following alternatives: (a) a rule that requires judges to follow and declare as "law" all and only M-rules, and (b) one that requires judges to follow and declare as "law" all and only S-rules. Note that, depending on whether the positivist or anti-positivist is right about <law>, one of these will have the

[68] The perspective of the institutional designed is intended merely as a heuristic: a means of identifying a moral ideal for the design of a judicial system.

same implications as LEGALITY for which rules judges should regard as law. Between the two principles, we would naturally prefer the criterion that when employed by judges would make things go morally best, all things considered.[69] If, for instance, judges are bad at identifying morally good rules (rules that it would be morally good for us all to follow), the positivistic principle of law-declaration would be best from the institutional point of view, for it does not require of judges that they appraise the moral features of rules in deciding which rules to follow.

But note that in deciding between the two principles, extensional equivalence with LEGALITY is irrelevant. The institutional designer should not care whether M-rules or S-rules are the ones our concept <law> has been tracking all along. Knowledge of the conceptual fact simply does not bear on what the best criterion is that judges should be using. Suppose we find it would make things go best if judges obeyed and pronounced as "law" rules that have the Hartian S-property. This is possible despite M-rules being the ones that meet a minimal degree of moral acceptability when obeyed. If it is best for judges to decide cases using S-rules from an institutional design stand-point, this fact is in no way undermined by a discovery that our concept <law> has tracked M-rules all along. Indeed, it would be implausible to think that, as users of the concept, we fixed on the M-property as opposed to the S-property precisely because it was the property it would be best for judges to track in deciding cases. Far more plausible is the assumption

[69] If one dislikes this way of putting it which assumes consequentialism in ethics, we can instead say: the institutional designer should pick whichever decision criterion has all the moral right-making features.

that judges and other users of the legal concept latched on to a property of rules that serves as an *adequate*—rather than best—basis for judicial decision making.[70]

The point can be motivated using a non-legal example. Suppose one becomes interested in the ordinary concept <marriage> because relationships that fall under the concept seem worth valuing. An examination of what marital relationships have in common might reveal a range of properties that make them valuable. Should it matter, ultimately, which of these features is essential to the ordinary concept? It should not if the aim is to identify relationships worth valuing. There may be features that are not essential to marriage but nevertheless typically found in marital relationships that make them valuable—like romantic devotion. In other words, one might discover upon reflection on the nature of marriage, features that are more relevant to the normative question than whatever it is that is essential to all and only those relationships that fall under the ordinary concept <marriage>. Moreover, in cases where it is non-obvious whether a relationship counts as a marriage because the ordinary concept's precise application conditions are obscure,[71] the question of whether the relationship is worth valuing should be decided independently of whether it counts as a marriage. For deciding the conceptual question involves deciding an issue that is irrelevant to the evaluative question—namely, which features of relationships have *we* been sensitive to in classifying certain ones as

[70] We shall return to this point in section IV when addressing objections to the view—in particular, the objection that the legal concept might already be optimally tailored to our ends.
[71] Consider a case where a couple has lived together for several years under the false belief that they were officially married by a priest who turns out to be an unlicensed charlatan. The ordinary concept's application, I submit, is obscure.

"marriage."[72] A relationship's value is determined by its intrinsic features; *not* by our sensitivity to those features within our linguistic practices.

The concept <law>, like the concept <marriage>, is a practically significant category. In well-functioning legal systems, rules that fall under the concept are at least typically ones we morally ought to obey. Accordingly, an institutional designer trying to find a decision criterion for judges to use in hard cases that would make things go morally best would do well to study the legal concept. She might discover features of rules that it would be beneficial for judges to track and use to decide cases. But, ultimately, the legal concept will be useful only in the limited way the concept <marriage> is useful to someone interested in the defining features of valuable relationships. In cases where the legal concept's application to candidate rules is uncertain because the precise application conditions are obscure, the institutional designer should determine what rules judges should follow based on their *non-legal* features rather than their legality. Their legality is not just obscure; it is of no *intrinsic* moral significance, being a function of our entirely contingent sensitivity to certain non-legal features of rules over others.[73]

[72] Could our sensitivity to certain features reflected in the precise conceptual constraints be an indirect yet ultimately useful guide to value? Perhaps. But, in all likelihood, the conceptual constraints are arbitrary. I address this concern more fully in the next section.

[73] If the boundaries of the concept were obvious, the analysis would be different. If the law were clear, determinate, and antecedently recognizable, there would be reason to follow it at least in part because those who are subject to the law's demands act based on their reasonable expectations concerning the law, and there are good reasons not to frustrate reasonable expectations. Moreover, there may be reasons for judges to *treat* legality as though it were intrinsically significant when the law is clear that do not apply in cases where the law is non-obvious. The obscure edge of legality is thus critical to the overall argument for why judges have authority to ignore law. The issue is discussed in further detail in section IV.

Fully specifying the ideal decision criterion for judges to use in hard cases is beyond the scope of this paper. But we must resist the temptation to think that just because we have not settled on an answer to this difficult question of institutional design, our default rule must be one that requires judges to fetishize the legal concept. LEGALITY cannot be a default rule that defines the judge's role precisely because the concept used to state the principle is far from transparent: it is not clear what the principle commands judges to do in hard cases.

To sum up, the conception of the judge's role I have defended invokes, first and foremost, a negative thesis: the *legal* rules that govern judicial decision-making do not include one that mandates treating all and only those rules as law that fall under the legal concept just because they are law. This negative thesis is compatible with the view that judges should resist the temptation to treat rules as law based on whether doing so would maximize morally good outcomes. In other words, the view is in principle consistent with a traditionally conservative take on how judges should ultimately decide hard cases.

Can we say, more positively, what a judge's professional duties are when it comes to deciding hard cases, if the duty isn't simply to follow the law as such, or, for that matter, to follow rules the judge believes to be law? One might worry that my argument generalizes and undermines any distinctively *legal* theory of judicial professional responsibility for hard cases, and not just one according to which the principle of LEGALITY best characterizes the judge's legal obligations. Consider the second-order question: what legal rule governs judicial decision-making when a judge confronts a hard first-order question of law? As a reminder, the judge confronts a hard first-order question of law when there are two or more incompatible first-order rules—say, an M-rule and an

S-rule—that would decide the question, but it isn't perspicuous which has the property of being law. First-order legal rules are rules prescribing actions or outcomes in circumstances characterized in non-legal terms (in terms that do not involve reference to other legal rules): e.g. *if a defendant is convicted of murder, he shall be imprisoned for life*. Second-order legal rules make essential reference to legality, either by referring to other legal rules or some legal status: e.g. *when confronted with an illegal statute, the judge is obliged to strike it down*. Suppose that the second-order question—how, from the law's point of view, should the judge decide a hard case?—is itself hard: as between various second-order decision rules the judge might follow (LEGALITY; a principle that prescribes following the M-rule in hard cases quite apart from the M-rule's legality; one that recommends following the S-rule in hard cases; and so on…), the legality of any particular decision rule will be subject to reasonable disagreement.[74] By my own lights, the *legality* of a rule is not sufficient to generate a legal obligation to follow it. When a candidate rule's legality is not perspicuous and subject to reasonable disagreement, judges are not obliged to follow the rule even if they correctly believe it to be legal. A rule's being legal but not perspicuously so undermines its normative significance *qua* law, because there is no general rule with the property of being law that directs judges to always follow the law when the law is not perspicuous. So, even if there is some higher-order decision rule other than the principle of LEGALITY with the property of being law, this will not entail that judges have a *legal* obligation to follow it, given that its legality will be at best non-perspicuous.[75]

[74] Thanks to an anonymous referee for raising this worry.

[75] There is a general view of distinctively *legal* obligation at work through-out this paper. Legal obligations arise when a rule specifying a type of action as obligatory has the property of being law. And there can be higher-order legal rules that oblige actors to follow the law in various

There are several things to say about this line of reasoning. First, suppose it is true that there are no second-order legal rules for deciding hard cases that judges are obliged to follow simply on account of their legality because their legality is non-perspicuous. It remains an open question whether judges might be *morally* obliged to decide hard cases in some particular way. The moral question would be a natural one to ask if the legality of decision rules is irrelevant from the point of view of a judge deciding a hard case. We shall consider an answer to this moral question in a moment. But quite apart from how that question is settled, it would be an important result in its own right that there are no distinctively legal obligations—that is, no rules with the property of being law that the judge is legally obliged to follow on account of their legality—in hard cases. This result would turn on (a) the absence of a general legal rule that directs judges to always follow the law, even when the law is not perspicuous; and (b) the legality of all second-order rules for deciding hard-cases being non-perspicuous. If (a) and (b) hold, it would be pointless for a theory of judging to be concerned with *what the law is* in hard cases, either at the first-order or second-order level, at least if the theory aims to be practical and action-guiding.

But I doubt that the general application of my argument to the legality of second-order decision rules is sound. The claim was never that mere disagreement about the legality of rules results in a hard case in the relevant sense. The argument that cases

---

contexts. E.g. the principle of LEGALITY, if it were conventionally embraced, might be such an obligation-generating rule. For our purposes, there is no need to settle the hard question of whether legal obligations are *genuine* (in the sense of having the same kind of normative force as moral obligations) or purely 'formal,' like the rules of chess. On formal vs robust normativity, see e.g. Stephen Finlay, "Defining Normativity," *in* (D. Plunkett, S. Shapiro, K. Toh, *Dimensions of Normativity: New Essays in Metaethics and Jurisprudence*, OUP, 2018. See also discussion in fn. 60.

where S-rules and M-rules come apart are hard appealed not just to reasonable disagreement about whether S-rules or M-rule are law, but on the nature of the evidence that would settle the issue: the complex and hard-to-access facts that determine the application conditions of the juridical concept of law.[76] Given the difficulty in discerning precise conceptual application conditions, reasonable disagreement on whether M-rules or S-rules are law is likely to persist. The argument against LEGALITY focused on such extreme cases of legal 'anti-luminosity.' There is no general rule that is conventionally embraced or one that is morally compelling that obliges judges to follow the law *as such* even in such extreme cases. That is, there are no clear conventions of law-following in such cases, and neither is there moral reason to fetishize law in such cases. So, it does not follow from the argument developed here that mere disagreement about which higher-order decision-rules are *legal* rules is sufficient to undermine the possibility of such rules obliging judges simply on the account of their legality.

Secondly, it is not at all obvious that there is or can be reasonable disagreement concerning the legality of *all* second-order decision rules for hard cases. So long as there are reasonably clear conventions in a jurisdiction for how to decide hard cases that also happen to involve decision-rules that are morally quite defensible, then lack of clarity concerning the precise contours of the legal concept certainly won't render a candidate decision-rule's legality non-perspicuous. The argument of this paper shows only that the legality of rules is subject to reasonable disagreement (and juridically insignificant) when the rules conventionally followed come apart in dramatic ways from the rules morally

[76] See discussion in section II.

favored. But that is consistent with there being morally good and conventionally embraced rules within our legal system for deciding hard cases.

Indeed, there are plenty of examples of morally good conventionally embraced decision-rules for hard cases that seem to be perspicuously law. It is fairly uncontroversial that judges cannot decide what to do in hard cases on the basis of mere whim or in deliberately immoral ways. In other words, it seems a plausible *legal* rule governing hard cases—one conventionally embraced within developed jurisdictions and morally sound—that judges must be *principled* and not have as one of their ultimate (or 'final') reasons for favoring a particular decision-criterion that it results in immoral outcomes (which is, of course, consistent with following a rule because one deems it immoral but not ultimately for the sake of its immorality). These legal constraints on judging in hard cases leave a lot underdetermined regarding the precise outcome that the judge is (legally) obliged to pursue. But the main point is just that the argument of this paper is compatible with there being *some* legal rules that *do* bind judges in hard cases simply on account of their legality, such that falling to follow such rules amounts to a failure of professional duty. American judges, for example, consent in their professional undertaking to abide by rules of principled decision-making even in hard cases.

So, it is a positive result of our discussion that although the principle of LEGALITY is not itself a legal rule constraining judges, there plausibly are distinctively legal norms governing juridical decision-making in hard cases. The edge of legality is not a legal free-for-all. Even if the legality of first-order rules does not matter in such contexts on account of first-order uncertainty, uncertainty at the first-order level does not translate into

sweeping higher-order legal uncertainty. There are higher-order rules that oblige judges and are perspicuously legal:

> PRINCIPLE PRINCIPLE: In hard cases, judges must declare rules as "law" in a principled way

PRINCIPLE PRINCIPLE is quite plausibly *the law* and perspicuously so.[77] A careful study of the conventions within jurisdictions and reflection on morally ideal institutional design might result in the discovery of even more fine-grained distinctively legal norms governing hard cases. But that study is a project for a different paper.

Does the positive claim, that judges must follow some legal rules simply on account of their legality even in hard cases, vindicate LEGALITY? The view imputes to legal systems a higher-order rule that permits judges to ignore the legality of first-order rules in hard cases and requires only that judges decide cases in a principled way. For example, when a judge decides whether to declare a morally abhorrent but conventionally embraced rule as the law, she can ignore whether the rule is *in fact* law in deciding what to do. Instead of the rule's legality, her decision might be guided by the principle of doing what is morally best out of a sense that the law requires her to make *some* principled decision. So, it seems, contrary to the paper's negative claim, judges *do* end up following the law as such in hard cases and indirectly comply with LEGALITY by complying with the PRINCIPLE PRINCIPLE as a legal rule.

[77] I have taken for granted that judges are legally obliged to remain faithful to well-recognized and paradigmatic *laws* (in the *de re* sense)—that is, there is a legal rule requiring judges to follow clear cases of law. That seems to be to be a truism, and the only plausible reading the platitude that judges should follow the law. So, there is a distinctively *legal* obligation to follow PRINCIPLE PRINCIPLE just because it is the law.

The fact that PRINCIPLE PRINCIPLE is a higher-order rule that judges are legally obliged to follow on account of its legality does not entail that judges must follow the law as such in hard cases. At best, judges should follow a higher-order legal rule – one that leaves the precise outcome to pursue in a hard case open – while ignoring the legality of first-order rules. There is simply no obligation to follow the law *tout court* in hard cases.

As mentioned earlier, even if it turns out that *legality* is simply not what matters at any stage of juridical deliberation about what to do in hard cases (because even a rule like the PRINCIPLE PRINCIPLE is at best only non-perspicuously legal), a theory of judging can fruitfully engage with the question *what are judges morally obliged to do quite apart from the law*? Again, answering this question is not the point of this paper. But if the correct moral theory turns out to be some form of rule-consequentialism, the answer to the moral question will have the following form:

> RULE-CONSEQUENTIALIST JUDGING: In hard cases, judges (morally) must declare rules as "law" based on principles and decision criteria that if widely followed by judges would make things go morally best, all things considered.

RULE-CONSEQUENTIALIST JUDGING is consistent with the claim that judges should, both legally and morally, remain faithful to well-recognized and paradigmatic *laws* (in the *de re* sense) and make a good faith attempt to figure out what the law is. It entails, only, that in hard cases where the line between legal and non-legal rules is unclear, judges have a moral obligation (and legal permission, so long as the argument of this paper is sound) to follow whichever principle for law-declaration would be morally best from the general institutional point of view.

IV. Objections and Replies

Some might consider it a *reductio* of my view that it entails judges can satisfy their professional obligations despite deliberately ignoring the law. In difficult cases where S-rules and M-rules come apart, the claim is it is simply irrelevant to what judges should be doing that historically "law" has been associated, say, with S-rules. If it is institutionally best for judges to be tracking and pronouncing M-rules "law" in hard cases, then the judge is legally permitted (and maybe morally obliged) to ignore what the law is and declare the M-rule "law." Is this not tantamount to a fraud on the public? Relatedly, one might think that by declaring rules to be law while believing them not to be law, the judge is morally blameworthy for being deceptive.[78]

As I have argued, cases where ignoring preexisting law is legally and morally appropriate are ones where it is not obvious what the law is. A judge who pronounces an M-rule "law" because of her sincere belief that her pronouncement conforms with what institutional designers would expect of judges need not be certain that all and only M-rules are law. The judge can in good conscience decide the case as though she were an anti-positivist. Even if the judge is for some reason certain that positivism is true, she should not be criticized for deviating from what she believes to be law insofar as she employs the morally best decision criterion for judges. The entire point of the previous discussion was that a judge can have respect for the law (or, better yet, respect for *laws*) in a way that fully satisfies her *legal* duties despite ignoring the law in hard cases. Even if she is morally blameworthy for calling a rule law while not believing it to be law (though

[78] Thanks to an anonymous referee for raising this worry.

it should be noted that it is highly controversial whether moral blame attaches to all instances of deceptive conduct or pretense),[79] I have primarily been interested in the question whether the judge is criticizable from *the law's point of view.* She would be legally criticizable if she were violating an explicit or implicit legal rule that governs her conduct. But, as argued, no legal rule requires her to conform her law-declaration to her beliefs about the law in hard cases, because there is no legal rule that obliges the judge to follow the law *as such* and in every case.

There is a different objection to the overall view that appeals to the interests of those held accountable under legal rules, one that might seem especially compelling when illustrated using the example of criminal law (and its *nullum crimen sine lege, nulla poena sine lege* principle).[80] Individuals act based on their expectations of what the law is and there are good reasons to avoid frustrating reasonable expectations. The ideal that laws be antecedently recognizable has been deemed central to law by a broad range of legal theorists.[81] In the case of criminal law, it seems especially important from the point of view of justice not to hold individuals accountable under rules that could not have been antecedently recognized by them as legal, given the severe consequences of being found to have violated the criminal law. But the proposed view of judicial obligations neglects this fairness constraint. If it is a conceptual fact that only S-rules are law, then calling an M-rule "law" and punishing a person under it offends against principles of justice having to do with adequate notice and fair warning.

---

[79] On the question of whether moral blame attaches, the juridical context might be one where pretense is morally appropriate, because the practice or conventions within the legal system authorize it. I have argued elsewhere that our practices can change our default moral obligations.
[80] 'No crime without law, no punishment without law.'
[81] See, e.g., L Fuller, *The Morality of Law* (Yale University Press. 1964).

The objection from citizens' reliance on law only goes through if the kind of linguistic infidelity the view permits frustrates the reasonable expectations of persons. But since it is very unclear whether the ordinary concept tracks M-rules or S-rules, persons are unlikely to have (reasonable) expectations about whether judges will follow one sort of rule or the other in borderline cases where the two come apart. The point can be made more generally and not just with respect to cases that separate positivism from anti-positivism. Hard cases of law where it is difficult to discern what the preexisting law is arise not just because the legal concept's application conditions are non-obvious, but also due to more familiar factors, such as semantic ambiguity in the statement of rules, uncertainty about judicial practice, evaluative complexity, and so on.[82] Are judges, in adjudicating such cases, obliged by considerations of fairness to take into account the reliance interests of those in the business of trying to do as much as could conceivably be permitted by law (Holmesian "bad men", in other words)?[83] It seems not. A much more plausible view is that citizens should not stray into the 'gray zone' or edge of legality; or, at the very least, that if they do, they give up reasonable grounds for complaint when their conduct is judged to be illegal. To the extent that people do have reasonable expectations about law, we serve that interest as I suggested before in discussing what the public *in fact* expects of judges: by examining the *paradigmatically* legal rules of a jurisdiction—that is, the clearest cases of law.[84]

---

[82] Dworkin (n9).

[83] OW Holmes, 'The Path of the Law' (1987) 10 Harvard Law Review 457, 459: 'If you want to know the law and nothing else, you must look at it as a bad man, who cares only for the material consequences which such knowledge enables him to predict, not as a good one, who finds his reasons for conduct, whether inside the law or outside of it, in the vaguer sanctions of conscience.'

[84] An issue I have not discussed here but analyze elsewhere is the precise relationship between the degree of *ex ante* certainty about the law on some issue and the law's intrinsic normative

The question of whether and to what extent laws need to be antecedently recognizable was front and center during the Nuremberg trials after World War II. There was considerable anxiety over holding Nazi war criminals liable under rules whose basis in prior international law was dubious. The criminalization of aggressive war in particular was seen as something of a legal novelty.[85] The agents of Nazi Germany could hardly have expected that the rules pronounced as international law at Nuremberg would be and yet, as David Luban has argued, an ideal of the rule of law that would prohibit such pronouncements, one that would "require lifelong protection for the elite of a genocidal regime," would not be worthy of respect.[86] In the case of Nazi Germany, the wickedness of the Nazi state deprived its agents of a reasonable complaint against being punished for their complicity after a fair trial. Relatedly, if considerations of justice favor our calling M-rules as opposed to S-rules "law" going forward, these are not defeated by the fact that some citizens (say, die-hard positivists) expect law at the margins to be constrained by entirely non-obvious conceptual facts.

V. A Reply to Dworkin

While not all theoretical inquiries need to be justified in terms of their practical upshot, our interest in the category of law is chiefly practical. Legal officials, among others, are tasked with the responsibility of figuring out what the law is, and it matters

significance for judges. The claim defended is that as the degree of *ex ante* certainty about law diminishes so does legality's normative significance.

[85] D Luban, 'The Legacies of Nuremberg,' (1987) 54 Social Research 779.

[86] ibid 800.

that they do this job correctly. The argument from judicial role highlights a way in which legal philosophical investigation can become untethered from its practical moorings. If judges need not be concerned with legality in hard cases, where the concept's application is non-obvious, then nailing down the precise application conditions does not serve judicial ends. The project of conceptual analysis in legal theorizing needs a different justification than the one traditionally given: namely, that answering the conceptual question will help judges comply with their legal obligations better.

The more positive lesson to draw from the argument is that there remains room for several juridically significant philosophical projects. Theorists offering candidate analyses of the concept of law can afford to be less concerned with getting the legal concept's extension exactly right and instead focus on ensuring that if judges were to rely on the suggested analysis in cases where there is disagreement over the concept's extension, they would be acting in ways that are morally best. In other words, the debate between positivists and anti-positivists might be re-oriented around the moral and political question: what criterion of law-declaration for hard cases results in morally sound judicial decision-making generally? There is, relatedly, an important project that is part sociological and part moral. It involves figuring out the precise conventions amongst judges for deciding hard cases within jurisdictions that involve morally good (even if not ideal) decision-criteria. These conventionally embraced morally good rules might be perspicuously legal, like the PRINCIPLE PRINCIPLE, and, if so, would provide additional content to the distinctively legal obligations that constrain (or purport to constrain) judges within jurisdictions simply on account of their legality. What that set of conventionally embraced and morally good rules won't include, we can be reasonably sure, is LEGALITY.

The suggestion that legal philosophy would benefit from being re-oriented around a normative enquiry is by no means novel.[87] It has been suggested that positivists and anti-positivists may have been implicitly debating a normative question—something like: which conception of law is best for us to use?—despite seeming to debate the actual contours of a fixed concept.[88] Nevertheless, we have discovered reasons for favoring a reorientation that appear to have been neglected in the broader debate and are unique to the legal context: reasons having to do with the professional role and responsibilities of judges. Whether or not the predicate "is law" expresses a concept that refers to all and only M-rules is not what ultimately justifies, from the legal point of view or otherwise, a judge's use of the word to refer to an M-rule in a hard case. If anything justifies such usage, it is a fact of ideal institutional design: that judges *should* pronounce M-rules "law" given the practical consequences of their generally doing so.

---

[87] See, e.g., Schauer (n28) 493; Murphy (n8) 9: 'So the methodology I favour for thinking about the boundary of law is what could be called a practical political one: the best place to locate the boundary of law is where it will have the best effect on our self-understanding as a society, on our political culture.' The present argument can be viewed as highlighting additional reasons for preferring this prescriptive methodology.

[88] D Plunkett and T Sundell, 'Antipositivist Arguments from Legal Thought and Talk: The Metalinguistic Response' in *Pragmatism, Law, and Language* (Routledge 2014). Plunkett and Sundell refer to the phenomenon of parties implicitly debating the question of what concept we *should* be tracking with our words as 'meta-linguistic negotiation.' More generally, see S Haslanger, 'What are we Talking About? The Semantics and Politics of Social Kinds' (2005) 20 *Hypatia* 10. I am skeptical that philosophers—or judges, for that matter—have been disputing how we *should* use "law" independently of its meaning. Some of my reasons for being skeptical are Dworkin's, who derided positivist attempts to re-interpret what judges are doing when they disagree about law. Dworkin (n15) 42-46. One would think that parties engaged in a 'meta-linguistic negotiation' would directly assert the fact whose acceptance is up for negotiation. That is, one would expect parties engaged in a dispute over how to use a word—say, "law"—to clearly assert, as part of the dispute, that the word ought to be used one way or another quite apart from its public or ordinary meaning. But, in fact, judges, and philosophers debating the nature of law, do not do that. Moreover, the argument developed here suggests that *there is no need* for meta-linguistic negotiation on the part of judges (or philosophers) in hard cases. When the precise contours of a concept are obscure and a predicate may very well express the ideal concept, one can use the predicate in the hopes that it expresses the ideal.

Finally, it would be a mistake to close an argument in which hard cases play such a major role without commenting on Ronald Dworkin's classic discussion of such cases.[89] Dworkin famously used hard cases to argue in favor of anti-positivism, and the present analysis sheds new light on Dworkin's insights as well as on the argument's shortcomings.

Dworkin observed that in hard cases judges both *do* and *ought to* decide which rules to declare "law" based on their moral features (based on their consistency with what Dworkin calls 'arguments of principle' or 'right').[90] Moreover, he pointed out that social facts seem to underdetermine what the law is in such cases. For instance, in *Everson v Board of Educ.*, the US Supreme Court had to decide whether a law purporting to grant free busing to parochial schools violated the First Amendment prohibition against religious establishment.[91] The social facts having to do with prior judicial decisions, constitutional text and history, interpretive practice, etc., did not determine whether assistance to parochial schools constitutes religious promotion of the sort the Constitution prohibits.[92] The Court seems to have decided the case by appeal to its conception of principles that ought to govern a democratic society that aspires to religious neutrality.[93] Dworkin's argument that in relying on general moral principles to decide such cases, judges end up behaving as they *should* is complex. But his principle claim was that it

[89] Dworkin (n9).
[90] ibid 1059.
[91] 330 U.S. 1 (I947). See discussion in Dworkin (n9) 1083-4.
[92] See the dueling opinions of the majority lead by Justice Black, 330 U.S. 1-17 (I947), and the dissent led by Justice Rutledge, ibid 28-63.
[93] Although, the majority claimed that blocking this form of assistance which *enables* parochial schools to operate was 'obviously not the purpose of the First Amendment,' it did not defend its claim. 330 U.S. 1, 17-18 (I947). The majority was likely motivated by the thought that it would be *morally wrong* to interpret the First Amendment to prevent a State from 'extending its general state law benefits to all its citizens without regard to their religious belief.' ibid 16-17.

would be wrong for unelected officials to decide hard cases by exercising discretion unconstrained by moral principles of right.[94]

Dworkin infers based on these observations that positivism must be false. What the precise argument against positivism is supposed to be is not entirely clear, but there are at least two compelling possibilities. Firstly, the positivist owes us an explanation of why, if the law has run out, judges continue to rule as though it has not based on a moral analysis of rules.[95] Secondly, a judge committed to positivism has no way of deciding the relevant cases as judges manifestly do (and should do) without violating her legal obligations, for she must treat the moral features of rules as ultimately irrelevant to their legality.[96] If she were to base her decision-making on moral considerations, then, by her own lights, she would be involved in the pretense of treating extra-legal considerations as though they were legal.

Positivists tend to respond to the explanatory challenge with the observation that just because judges act *as if* they are discovering law in hard cases, we need not take their behavior at face value. But the response remains under-motivated without an account of *why* a deeper, less surface-level interpretation of judicial behavior is called for. The present discussion provides principled reasons for being skeptical of appearances. We would have reason to think judges are single-mindedly attempting to discover pre-

[94] Dworkin offers two arguments against unconstrained judicial originality: 'The first argues that a community should be governed by men and women who are elected by and responsible to the majority…. The second argues that if a judge makes new law and applies it retroactively in the case before him, then the losing party will be punished, not because he violated some duty he had, but rather a new duty created after the event.' Dworkin (n9) 1061.
[95] As Dworkin puts it: 'The rights thesis… provides a more satisfactory explanation of how judges use precedent in hard cases.' ibid 1064.
[96] Social conventions *might* render moral analysis relevant—see Endicott (n3) on inclusive positivism—but we are imagining a case where the social conventions do not necessarily entail that moral analysis is appropriate in the case at bar.

existing law in hard cases if we had grounds to assume the truth of LEGALITY. But, as argued, the rule that judges in hard cases should make law-declarations based on actual law is unlikely to be conventionally embraced nor is it a good rule for a legal system to adopt. The entirely contingent conceptual fact as to whether rules supported by considerations of moral principle fall under the legal concept has no intrinsic normative significance for what judges should do in cases where the conceptual constraints are obscure. If judges should declare morally good rules as "law" in hard cases it is not because by doing so they would be discovering law, although they very well *might be*. If judging in hard cases *as though* anti-positivism were true is warranted; it is so for the moral and political reasons that Dworkin gives *and* because it is obscure what the law is in such cases; and not because anti-positivism is actually true. Accordingly, positivism is entirely compatible with judicial behavior in hard cases.

The second version of the argument seems also to rely on LEGALITY, at least implicitly. It presupposes that if there is a fact of the matter about what the law is, as it seems there might be in hard cases, positivist judges would be violating their legal obligations if they based their decision-making on what they deem to be non-legal considerations. But, as we have seen, judges would *not* be violating any legal obligations by basing their decision-making on what may or may not turn out to be non-legal considerations (e.g. the moral characteristics of rules), so long as they decide in a principled way—legality is not ultimately what matters in hard cases.

Since these ways of responding to Dworkin's arguments rely on the falsity of LEGALITY, they also undercut the stakes in the debate between Dworkin and the positivist. Neither side can claim the truth concerning the precise boundaries of law and morality

would make any difference to how judges should rule. Positivists may well prefer a response to the argument that does not rob the debate of practical significance in this way. It is not positivism that we have been concerned with defending, however. Our aim was simply to clarify the true significance of hard cases.

Conclusion

Those who take the central question animating analytic jurisprudence—what are the precise boundaries of the concept of law?—to be practically significant assume that we needed to settle the question to specify the professional duties of judges. That presumption is mistaken because it takes as unquestioned a long-standing shibboleth of legal philosophy— namely, that if a rule is, *ex ante*, the law, then judges have at least a *legal* obligation (and maybe even a moral one) to declare it as such and follow it. In fact, a rule's being law is neither necessary nor even sufficient for judges' being obliged to follow it, a fact especially apparent in hard cases of the sort where positivism and anti-positivism come apart, rendering the truth of either theory and the precise limits of our legal concept irrelevant from the juridical point of view. The practical upshot of legality's irrelevance in hard cases for judicial decision-making is significant. It entails, for example, that judges can, consistent with their legal professional duties, ignore rules they believe to be law in hard cases.

Chapter 6. Legal vs. Factual Normative Questions under the Common Law

Abstract. *When is a normative question a question of law rather than a question of fact? The short answer, based on common law and constitutional rulings, is: it depends. For example, if the question concerns the fairness of contractual terms, it is a question of law. If it concerns the reasonableness of dangerous risk-taking in a negligence suit, it is a question of fact. If it concerns the obscenity of speech, it was a question of fact prior to the Supreme Court's seminal cases on free speech during the 1970s, but is now treated as law-like. This variance in the case law cannot be explained by traditional accounts of the law/fact distinction and has fueled recent skepticism about the possibility of gleaning a coherent principle from judicial rulings.*

This Article clarifies a principle implicit in the settled classifications. I suggest that judicial practice is consistent: it can be explained by the distinction between normative questions that are convention-dependent and those that are convention-independent. Convention-dependent normative questions, or those that turn essentially on facts about our social practices (roughly, what we do around here) are reasonably classified as questions of law. By contrast, convention-independent normative questions, which turn instead on fundamental moral norms concerning what persons are owed simply on account of being persons, are properly classified as questions of fact. This principle, echoed in recent holdings, clarifies law/fact classifications in such diverse areas as torts, contracts, First Amendment law, and criminal procedure.

The principle also promises to resolve a looming constitutional controversy. In Ring v. Arizona, the Supreme Court held that all factual findings that increase a capital defendant's sentence must be decided by the jury under the Sixth Amendment. Two recent denials of cert. suggest that members of the Court wish to revisit, in light of Ring, the constitutionality of judges deciding whether a criminal defendant deserves the death penalty. Applying the principle to Ring, I argue that the question of death-deservingness is a convention-independent normative question, and for that reason should be deemed a factual question for the jury.

Introduction

When is a normative or evaluative question that arises at trial a question of *law* as opposed to a question of *fact*?[1] The short but not very helpful answer based on judicial rulings is: it depends.[2] If the question concerns the reasonableness of an "implied" term in a contract or the unconscionability of the contract as a whole, it is a question of law.[3] If it concerns the unreasonableness of the defendant's conduct in a negligent suit, it is a

[1] Described as the "controlling distinction between the power of the court and that of the jury," Dimick v. Schiedt, 293 U.S. 474, 486 (1935), the law/fact distinction determines whether the judge or jury will decide an issue, the standard of review on appeal, whether burdens of proof and discovery rules apply, as well as the decision's precedential value. *See* discussion *infra* Part I. The maxim that "judges do not answer a question of fact, and juries do not answer a question of law" has been traced to the 16th century. *See* Stephen A. Weiner, *The Civil Jury Trial and the Law-Fact Distinction*, 54 CAL. L. REV. 1867 (1966) (citing COKE, COMMENTARY ON LITTLETON 460 (1818)).
The distinction determines factors that powerfully predict a party's likelihood of prevailing on the issue. *See e.g.*, Kevin M. Clermont & Theodore Eisenberg, *Trial by Jury or Judge: Transcending Empiricism*, 77 CORNELL L. REV. 1124 (1992) (finding that plaintiffs in product liability and medical malpractice cases prevail at a much higher rate before judges than they do before juries). *See also* Robert Anderson IV, *Law, Fact, and Discretion in the Federal Courts: An Empirical Study*, 2012 UTAH L. REV. 1, 5 (2012); Kimberly A. Moore, *Judges, Juries, and Patent Cases--An Empirical Peek Inside the Black Box*, 99 MICH. L. REV. 365 (2000); Timothy J. Storm, *The Standard of Review Does Matter: Evidence of Judicial Self-Restraint in the Illinois Appellate Court*, 34 S. ILL. U. L.J. 73 (2009) (finding empirical evidence of a difference in affirmation and reversal rates based on standard of review); Patricia M. Wald, *The Rhetoric of Results and the Results of Rhetoric: Judicial Writings*, 62 U. CHI. L. REV. 1371, 1391 (1995) ("Appellate courts have to decide what the 'standard of review' is, and that standard more often than not determines the outcome.").

[2] I am specifically interested in the use of the distinction by *judges* to classify issues that plain statutory law does not specify as judge or jury issues. How legislatures treat questions is not the focus of present discussion.

[3] *See* RESTATEMENT (SECOND) OF CONTRACTS § 204, cmt. a, c (AM. LAW INST. 1981); U.C.C. § 2-302(1) (AM. LAW INST. & UNIF. LAW COMM'N 2011). Although the U.C.C.'s enactment makes contractual unconscionability a question of law by legislative fiat, the U.C.C.'s drafters drew on the historic practices of common law judges in deciding the issue. *See* Donald R. Price, *The Conscience of Judge and Jury: Statutory Unconscionability as a Mixed Question of Law and Fact*, 54 TEMPLE L. Q. 743, 745–48 (1981); discussion *infra* Part I.B.

question of fact.[4] In criminal law, whether the defendant's conduct was especially "cruel" or "heinous" to warrant a higher penalty is a question of fact for the jury.[5] Additionally, there have been shifts in classification across time. In the context of defamation actions, whether a false statement was made with "actual malice" was traditionally a question of fact reviewed deferentially.[6] But ever since *Bose Corp. v. Consumers Union*, it is reviewed *de novo*, the standard for questions of law.[7] A similar shift occurred in the case of obscenity: whether a publication was "obscene" or "prurient" used to be a paradigmatic question of fact for the jury until the Court's seminal cases on free speech in the 1970s, when it began to be treated as law-like.[8]

The case law presents a puzzle with deceptively high stakes. Judicial practice flatly contradicts standard theories of the common law's 'law/fact' distinction. The dominant view amongst legal theorists is that the law/fact distinction tracks or maps on to the distinction between normative and empirical questions. On this view, normative questions—questions concerning what *ought* to happen or how persons ought to behave—are necessarily legal; while all and only empirical questions—those concerning (roughly) what happened in the world—are factual. [9] The first part of my project involves

---

[4] *See* Weiner, *supra* note 1, at 1877, n. 43 ("The courts of all American jurisdictions, with possibly one exception, adhere to this principle . . . . There are literally hundreds of cases in which this concept has been asserted."); *id.* at n. 43 (citing cases invoking the rule); discussion *infra* Part 1.B.

[5] *See* Apprendi v. New Jersey, 530 U.S. 466, 490 (2000); Blakely v. Washington, 542 U.S. 296, (2004).

[6] James L. Oakes, *Proof of Actual Malice in Defamation Actions: An Unsolved Dilemma*, 7 HOFSTRA L. REV. 655, 688–98 (1979) (observing a long history of jury findings on the question of malice in civil and criminal defamation cases). *See* discussion *infra* Part 1.C.

[7] Bose Corp. v. Consumers Union, 466 U.S. 485, 501–02, n. 27 (1984).

[8] *See* Jenkins v. Georgia, 418 U.S. 153, 159–61 (1974); Miller v. California, 413 U.S. 15, 30 (1973); discussion *infra* Part 1.C. Specifically, the question of whether the depiction of sexual conduct was "patently offensive" came to be treated as law-like.

[9] I discuss the empirical-normative distinction more carefully in Part I. Other theories similarly struggle to explain the practice of judges—for instance, the view that legal questions are ones of

showing that this dominant view is mistaken: judges in the common law have long treated some normative questions as legal and others as factual, which suggests that the law/fact distinction, at least as it has been interpreted by judges, cuts across the normative domain.[10] Courts recognizing the difficulty in deriving a coherent principle from the settled classifications have described the jurisprudence as "elusive," "slippery," and as having a "vexing nature."[11] The hard question of interpreting the case law comes up frequently and with constitutional ramifications: the law/fact distinction is a trigger for Sixth and Seventh Amendment jury trial rights. The Supreme Court interprets the scope of the jury trial rights based on the common law practice of distinguishing essentially legal from factual questions.[12] As a result, the issue of how to classify normative questions under the law/fact distinction has resulted in some controversial decisions.

---

general applicability whereas factual questions are case-specific or particular. *See* discussion *infra* Part I. The challenge that this part of the case law poses for traditional theories of the law/fact distinction has been widely discussed in the literature. *See* Ronald J. Allen & Michael S. Pardo, *The Myth of the Law-Fact Distinction*, 97 NW. U. L. REV. 1769, 1771 (2003) ("The ubiquitous distinction, despite playing many key doctrinal roles, is muddled to the point of being conceptually meaningless."); Randolph E. Paul, *Dobson v. Commissioner: The Strange Ways of Law and Fact*, 57 HARV. L. REV. 753, 812 (1944) ("What is law to one Justice is fact to another, and perhaps vice versa when the next case comes along."); Robert L. Stern, *Review of Findings of Administrators, Judges and Juries: A Comparative Analysis*, 58 HARV. L. REV. 70 (1944); Weiner, *supra* note 1, at 1868 (observing that courts have shown no inclination to fashion definitions of law and fact which can serve as useful guidelines). Skepticism about the distinction goes back many decades. *See e.g.*, LEON GREEN, JUDGE AND JURY 270 (1930); Francis H. Bohlen, *Mixed Questions of Law and Fact*, 72 U. PA. L. REV. 111 (1924); Nathan Isaacs, *The Law and the Facts*, 22 COLUM. L. REV. 1, 2 (1922).

[10] I suspect that the story of how the distinction maps on to the empirical domain is complicated as well, but we will be concerned solely with the normative domain in what follows.

[11] *See* Williams v. Taylor, 529 U.S. 362, 385 (2000); Thompson v. Keohane, 516 U.S. 99, 111 (1995); Pullman-Standard v. Swint, 456 U.S. 273, 288 (1982). *See also* Miller v. Fenton, 474 U.S. 104, 113 (1985) (observing the difficulty of stating a rule that would "unerringly distinguish a factual finding from a legal conclusion").

[12] *See* discussion *infra* Parts I and IV.

Consider *Cooper Industries v. Leatherman Tool*.[13] In that case, the majority held that the question of whether punitive damages in an unfair competition action were proportional or not could be reviewed *de novo* as intermediate between law and fact, despite the proportionality question having been historically regarded as a factual question meant for the jury and reviewed deferentially.[14] The majority emphasized that punitive damages involve "moral condemnation,"[15] echoing the traditional scholastic view that whereas questions of law are normative questions involving "the establishment, disestablishment, modification, or interpretation of legal rules,"[16] factual questions are those concerning "who did what, where, when, how, why, with what motive or intent."[17] But as Justice Ginsburg reasonably emphasized in her dissent in *Cooper Industries*, normative findings have a long history in the common law of being characterized "as *factfindings—e.g.*, the extent of harm or potential harm caused by the defendant's misconduct, whether the defendant acted in good faith, . . . whether the defendant behaved negligently, recklessly, or maliciously."[18] Whatever one thinks of Justice Ginsburg's ultimate conclusion in the case, she is surely right that normative questions are often classified as factual in the common law.

---

[13] Cooper Indus. v. Leatherman Tool Group, Inc., 532 U.S. 424 (2001).
[14] *Id*. at 437.
[15] *Id*. at 432.
[16] Randall H. Warner, *All Mixed Up About Mixed Questions*, 7 J. APP. PRACT. & PROC. 101, 112 (2005).
[17] RICHARD J. PIERCE, JR., 2 ADMIN. LAW TREATISE § 10.5, at 732 (4th ed. 2002). *See also* Ronald J. Allen, Alexia Brunet & Susan Spies Roth, *An External Perspective on the Nature of NonEconomic Compensatory Damages and their Regulation*, 56 DEPAUL L. REV. 1249, 1262 (2007) (reciting evidence that "fact" has historically been understood to concern events in the external world capable of identification through empirical inquiry).

[18] *Cooper Indus., Inc.*, 532 U.S. at 446 (Ginsburg, J., dissenting).

More recently, the issue of how normative questions should be classified came up again in two recent denials of cert., from 2013 and 2016, concerning capital sentencing procedure.[19] The plaintiffs sought review of Alabama's practice of letting judges independently decide the question whether the defendant deserves the death penalty based on the cumulative weight of the aggravating and mitigating factors in the defendant's case. Under *Apprendi v. New Jersey* and *Ring v. Arizona*, all "findings of fact" that increase the severity of a defendant's sentence must be found by the jury in light of the Sixth Amendment.[20] The issue is whether a finding on the death-deservingness question is a "finding of fact." Although the Court denied cert., Justices Sotomayor and Breyer wrote a strongly worded dissent from the 2013 cert. denial, observing that "[t]he statutorily required finding that the aggravating factors of a defendant's crime outweigh the mitigating factors is therefore necessary to impose the death penalty. It is clear, then, that this factual finding exposes the defendant to a greater punishment than he would otherwise receive: death, as opposed to life without parole. Under *Apprendi* and *Ring*, a finding that has such an effect must be made by a jury."[21]

I suspect that the likely sticking point that separates the majority from Justices Sotomayor and Breyer is that the question of death-deservingness (and the overall weight of the aggravating and mitigating evidence) is a *normative* question. The Supreme Court has routinely emphasized the normative character of this final determination, suggesting that "in the final analysis, capital punishment rests on not a legal but an ethical

[19] *See* Woodward v. Alabama, 134 S. Ct. 405, 407 (2013) (Sotomayor, J., dissenting). *See also* Brooks v. Alabama, 136 S. Ct. 708 (2016).
[20] Ring v. Arizona, 536 U.S. 584 (2002); Apprendi v. New Jersey, 530 U.S. 466 (2000).
[21] *Woodward*, 134 S. Ct. at 410–11.

judgment—an assessment of . . . the 'moral guilt' of the defendant."[22] As discussed, such questions are not always treated as factual. Justices Sotomayor and Breyer did not articulate a reason for treating the normative question on which the death penalty turns as a question of fact.[23] If there is a principle immanent in the settled classifications that would confirm the Justices' view, it has yet to be articulated.

This Article argues that there is indeed a useful principle implicit in the established case law. While this principle may not afford a complete explanation for how judges have treated normative questions, it is at least part of the explanation for judicial practice and, moreover, the principle should inform future classifications under the distinction.[24] Judges have been tracking a distinction between two kinds of normative questions: essentially convention-dependent and convention-*independent* normative questions. Conventions can be understood as social practices—roughly "what we do around here" or what norm we actually follow. There are merchant conventions, conventions of legislators and judges, and conventions of various other sorts. Normative

[22] Spaziano v. Florida, 468 U.S. 447, 481 (1984) (Stevens, J., concurring in part and dissenting in part). *See also* McCleskey v. Kemp, 481 U.S. 279, 310 (1987) (outlining the capital jury's task of expressing "the conscience of the community on the ultimate question of life or death"); Woodson v. North Carolina, 428 U.S. 280, 297–98 (1976) (reflecting on the importance on the moral views of society in the administration of death penalty); discussion *infra* Part IV.

[23] *See* discussion *infra* Part IV.

[24] My preferred methodology for legal theorizing is rational reconstruction or charitable interpretation. The approach is partly descriptive and partly normative. The aim is to interpret judicial behavior in a way that casts judges and their decision-making in the best possible light. By unifying under relatively simple general principles a disparate body of case law, we preserve continuity with past practice while also discovering decision criteria that can be useful going forward. Finding reason in judicial practice, even if it is not entirely faithful to the actual intentions of all parties involved, is a worthwhile exercise: it serves the important functions of rendering the law more integrated and fostering respect for law. In other work, I defend the importance of charitable interpretation for legal systems. A significant virtue of this way of proceeding is that judges seem to rely on such a method in figuring out what the law is in hard cases, as Ronald Dworkin has argued.

questions are essentially convention-dependent when they only admit of a determinate answer by appeal to convention facts. When the relevant conventions are inconclusive or ambiguous, such normative questions do not admit of determinate answers.[25] By contrast, convention-*independent* normative questions do not turn primarily on facts about what we do around here. They turn instead on, and are determinately answered by, fundamental moral norms—e.g., those concerning what persons deserve simply on account of being persons.[26]

This distinction helps explain and rationalize judicial classification of normative questions as legal or factual. Judges classify normative questions as legal or factual at some stipulated level of generality. They classify *types* of normative questions as legal or factual—e.g., the question of reasonableness in negligence actions, or the question of unconscionability in a contracts dispute.[27] The relevant question is how *likely* is a type of normative question classified under the law/fact distinction to be essentially convention-dependent or -independent in particular cases. And the central claim of this Article is that

[25] Convention-dependent norms, I argue, tend to be ones that (1) concern the distribution of benefits and burdens that do not implicate matters of fundamental right; (2) solve moral problems that require large-scale collective action; (3) arise in contexts where a paramount concern is respecting the expectations of participants in a convention.

[26] These admit of determinate answers, either by appeal to fundamental moral facts alone, or on the basis of the fundamental moral facts *plus* facts about our conventions.

[27] The law, as Lee Fennell, notes is lumpy: it operates through general rules of thumb. Lee A. Fennell, *Lumpy Property*, 160 U. PA. L. REV. 1955 (2012). My view predicts how judges will classify a type of normative question, once the type (or level of generality) has already been chosen by judges. For example, the category 'unconscionability in a contracts dispute' is a more general category than 'unconscionability in mortgage contract disputes.' The classification of questions of unconscionability as legal has occurred at the more general level of contracts disputes. The distinction between types of question and particular tokens or instances is worth bearing in mind in what follows. Just because there are instances of a type that are convention-dependent, this does not settle whether the type is more likely to be convention-dependent or -independent. In follow-up work, I explore instances where it might be better for courts to ask the convention-dependent/-independent (law/fact) question at a different level of generality—in particular, at a more fine-grained level.

if a type of normative question is more likely to be convention-independent—that is, if it is more likely to implicate fundamental moral norms, then it is reasonably classified as a question of fact. Juries are well suited to deciding such questions. By contrast, there are sound conceptual and pragmatic reasons for classifying as *legal* those types of normative questions that are likely to be essentially convention-dependent.[28] There are echoes of this principle in court opinions.[29] But its primary virtue is that it promises to explain the practice of common law judges.

The principle explains, *inter alia*, the contrasting treatment of key reasonableness questions in torts and contracts. The question whether a factory owner behaved reasonably in failing to implement safety protocols that could have prevented severe injuries suffered by her employees is a different kind of question from whether a price term in a contract is reasonable. Reasonableness norms governing dangerous risk-taking often turn not simply on convention facts, but, as Gregory Keating writes, on "rights and

[28] *See* discussion *infra* Part II. The conceptual argument appeals to what distinguishes essentially *legal* from pure *moral* normativity—the former's at least partial dependence on social conventions—drawing on a point of relative consensus in legal philosophy. Both in ordinary language and within the law, we recognize different varieties of normativity (legal, moral, semantic). The pragmatic argument emphasizes, among other things, that judges have a special competence to decide normative questions that turn on conventions (law-related or otherwise), but it is unlikely and at the very least highly controversial that they have any special expertise over the jury in deciding fundamental questions of moral fact.
Although, I have put the principle in terms of a purely statistical notion of likelihood, the issue may involve more than mere likelihood. The relevant question might be whether a question is *sufficiently* likely to implicate matters of basic right and wrong.

[29] *See, e.g.*, Bose Corp. v. Consumers Union of United States, Inc., 466 U.S. 485, 501 n. 16 (justifying de novo review in cases where a normative finding "cannot escape broadly social judgments"). Though it might seem like my theory gets things backwards—convention-independent norms appear more robustly normative than convention-dependent norms, after all—it does not. When pure moral questions are implicated under the law, answers to such questions cannot and should not be deemed legal, or so I argue. Moreover, the social or convention-dependent character of legal norms does not make them any less normative. Finally, my analysis rejects the values vs. facts view of the law/fact distinction. *See* discussion *infra* Parts I and II.

obligations that attach to persons *simply as persons*."[30] One reason why basic moral rights are often implicated in negligence cases is that such cases routinely involve harms to interests that have a special moral priority, including "the interests in one's own physical health and vigor, the integrity and normal functioning of one's body, the absence of absorbing pain and suffering or grotesque disfigurement."[31] Given the harms at stake, it is reasonable to assume that basic moral rights bear on what constitutes reasonable risk-taking. Of course, conventions may also bear on the issue, but they are not independently determinative of the normative question. Plausibly, the relevant conventions are themselves the result of agents trying to do what they morally ought to do, and are thus indicative of background pre-conventional moral norms.[32] Basic moral rights and obligations are sufficiently implicated in negligence actions to rationalize the classification of the reasonableness of dangerous risk-taking as a question of fact.

By contrast, the reasonableness of an implied price term in a contract is not the sort of question that can be settled independently of conventions or by appeal to basic moral principles. If there are no set conventions in place regarding how to price widgets, there will simply be no determinate fact of the matter as to what constitutes a "reasonable price" to be implied by the court within a broad range of possible prices, when parties forget to settle on a price term. This is because the distribution of benefits and burdens stemming from voluntary trade does not implicate interests of foundational moral

---

[30] Gregory C. Keating, *Is the Role of Tort to Repair Wrongful Losses?*, *in* RIGHTS AND PRIVATE LAW 367, 383 (Donal Nolan & Andrew Robertson eds., 2012) (emphasis added).
[31] JOEL FEINBERG, THE MORAL LIMITS OF THE CRIMINAL LAW, VOLUME 1: HARM TO OTHERS 37 (1987).
[32] There is an important difference between the way conventions bear on what is reasonable in torts—they play an *evidential* role, indicating background moral norms—and the way non-moral conventions independently determine contractual norms. See discussion *infra* Part III.A and fn. 195.

importance. More generally, in the economic context, judges reasonably regard questions concerning the reasonableness of implied contractual terms and overall contractual unconscionability as essentially convention-dependent—as determined by merchant and regulatory conventions. The modern marketplace is quite plausibly an arena of relaxed interpersonal expectations, with conventions of self-interested and even predatory behavior having displaced pre-conventional moral norms of good behavior.[33] Thus, judges will frequently refer to the unique "morals of the market place"[34] and refuse to "impose . . . duties higher than the morals of the marketplace."[35]

I use the framework to explain other aspects of the case law on legal and factual normative questions. But its prescriptive upshot for the impending capital sentencing controversy discussed earlier warrants special emphasis. The analysis bolsters the case for the unconstitutionality of judges issuing death sentences independently of juries. The question whether a criminal defendant deserves to be executed is paradigmatically not the kind of question that can be *affirmatively* answered by appeal to social convention facts. An affirmative answer to the question of death-deservingness implicates fundamental questions of fairness and basic dignity.[36] Moreover, the Court has regularly emphasized the foundational ethical character of this final determination, and that it must be a "reasoned *moral* response to the defendant's background, character, and crime."[37] Members of the Court recognize that the law demands moral deliberation from the sentencer, that the life-or-death decision is a question concerning the defendant's "moral

[33] *See* discussion *infra* Part II.
[34] Meinhard v. Salmon, 164 N.E. 545 (N.Y. 1928).
[35] Lee v. LPP Mortg. Ltd., 74 P.3d 152, 162 (Wyo. 2003).
[36] *See* discussion *infra* Parts II and IV.
[37] Penry v. Lynaugh, 492 U.S. 302, 319 (1989).

entitlement to live,"[38] and that it must be based on a "moral inquiry into [his] culpability."[39] These holdings can be interpreted as standing for the proposition that the normative question on which the death penalty ultimately rests is a convention-*independent* moral question. Accordingly, given the framework I outline, it is a question of fact. To the extent that the common law rule controls in this case (and I argue that it does or at least weighs heavily) there is a strong argument to be made that only juries can affirmatively answer the question of death-deservingness.[40]

Part I explains the basic role of the law/fact distinction and its significance within the common law and constitutional law. It also describes traditional theories of the distinction and the challenges facing these theories using settled law/fact classifications in torts, contracts, criminal procedure, and First Amendment law. Part II introduces and defends an alternative theoretical frame for understanding the law/fact distinction as it has been interpreted by courts: the difference between convention-dependent and convention-independent norms.[41] Part III demonstrates the framework's potential for explaining how judges have handled normative questions. Part IV is primarily prescriptive, focusing on the framework's material implications for the scope of *Ring*, the Sixth Amendment jury trial right, and the unconstitutionality of Alabama judges independently deciding whether a criminal defendant deserves the death penalty.[42]

---

[38] Spaziano v. Florida, 468 U.S. 447, 469 (1984).
[39] California v. Brown, 479 U.S. 538, 545 (1987) (O'Connor, J., concurring).
[40] A negative answer—that is, a determination that life not death is warranted—can be made by judge and jury, I argue. An affirmative answer needs to overcome basic moral constraints. A negative answer does not and can be based on a social practice of mercy and forgiveness.
[41] I am primarily interested in the principles immanent in the behavior of common law judges—that is, the wisdom of the common law judge. My account does not address how or why legislatures have assigned questions to judge or jury.
[42] Recent cases suggest that members of the Court believe the practice of judges independently sentencing defendants to death is headed for a constitutional challenge. See discussion *infra* Part

I. How Courts Deal with Normativity Under the Law/Fact Distinction

There are two strands of case law on the law/fact issue that should be distinguished. There is, on the one hand, a line of Supreme Court cases interpreting the scope of the Sixth and Seventh Amendment rights to have juries decide questions of fact in civil and criminal cases.[43] The constitutional case law is concerned not just with distinguishing factual questions from legal ones but with determining whether a question of fact *must* be decided by the jury for constitutional purposes.[44] By contrast, the common law practice of distinguishing questions of law from questions of fact is not necessarily directed at satisfying constitutional requirements. The primary goal of the common law practice is to allocate decision-making responsibilities between judge and jury in a *principled* way and set standard of review when plain statutory law and the

---

IV. On April 11th, 2017, Alabama governor Kay Ivey signed into law a bill banning judges from independently of juries sentencing to death defendants convicted after April 11th, but the law does not apply retroactively to defendants convicted before April 11th. S.B. 16, Reg. Sess. 2017 (Ala. 2017). Judges remain free to exercise the 'life-to-death override' for pre-April 11th murder convictions. The legislation will not affect the 183 inmates currently on Alabama's death row. Alabama judges have overridden jury life sentences 112 times since 1976. *See Alabama Ends Death Penalty by Judicial Override,* ASSOCIATED PRESS (Apr. 11, 2017), https://www.usnews.com/news/best-states/alabama/articles/2017-04-11/alabama-ends-death-penalty-by-judicial-override.

[43] U.S. CONST. amend. VII ("In suits at common law, where the value in controversy shall exceed twenty dollars, the right of trial by jury shall be preserved, and no fact tried by a jury, shall be otherwise re-examined in any Court of the United States, than according to the rules of the common law."). The Sixth Amendment in relevant part states "[i]n all criminal prosecutions, the accused shall enjoy the right to a speedy and public trial, by an impartial jury of the State." U.S. CONST. amend. VI. *See also* Allen & Pardo, *supra* note 9, at 1779; discussion *infra* Parts I.B, I.C.

[44] Welsh White makes a similar point regarding the importance of distinguishing the two lines of jurisprudence. Welsh S. White, *Fact-finding and the Death Penalty: The Scope of a Capital Defendant's Right to Jury Trial*, 65 NOTRE DAME L. REV. 1, 4–5 (1989).

Constitution are *silent* as to whether a judge or jury should decide a question that arises at trial.[45]

This is an important difference to bear in mind in what follows. While the common law distinction guides judges on the allocative question, it is not always considered decisive. Courts will, for example, sometimes reserve for judges the responsibility of deciding a factual question for pragmatic reasons.[46] Alternatively, paradigmatic questions of fact may be assigned to judges by legislative fiat. By contrast, the constitutional question is a question of *right*: when *must* a plaintiff or defendant be afforded the right to have juries decide a question of fact raised at trial? The constitutional and common law rulings on law and fact are related, however. The Supreme Court regards the common law practice of treating an issue as factual or legal as a factor in deciding the constitutional question.[47] Nevertheless, it is important to keep the two lines of jurisprudence separate, given that the constitutional analysis turns on more than just the law/fact issue. It is also worth bearing in mind that our concern is solely with *judicial* practice—that is, how judges have interpreted the law/fact distinction; and not necessarily with the behavior of, say, legislatures in assigning questions to judges or juries.

---

[45] *See* discussion *infra* Part I.B; Weiner *supra* note 1, at 1867–68 (describing state statutes that leave it to judges to define what 'law' and 'fact' mean). *See also* OLIVER W. HOLMES JR., THE COMMON LAW 122–27 (1881).

[46] *See* White, *supra* note 41 (noting that judges make factual findings in the evaluation of evidence). Pragmatic considerations even give rise to a "complexity exception" to the Seventh Amendment right to have juries decide issues of fact. *See, e.g.*, In re Boise Cascade Sec. Litig., 420 F. Supp. 99 (W.D. Wash. 1976).

[47] *See* Ring v. Arizona, 536 U.S. 584, 599 (2002) (citing approvingly the work of Welsh White); Markman v. Westview Instruments, Inc., 517 U.S. 370, 378 (1996). *See also* White, *supra* note 44, at 4–5 (observing that English legal scholars and the common law understanding of the distinction between law and fact "undoubtedly influenced the framers").

The procedure for deciding the Seventh Amendment question involves the Court determining, first, whether a right to a jury trial exists for the overall cause of action.[48] If the right exists, the Court engages in a historical inquiry to see whether a specific question arising in the case would have been assigned to the jury in 1791, when the Amendment was ratified.[49] If the historical inquiry is inconclusive, the Court examines a wide variety of factors, including prudential and "functional" considerations, in deciding whether to mandate jury involvement.[50] The functional analysis takes into account judicial practice as well as conceptual differences between questions of law and fact.[51] The key point, which will be demonstrated at length in what follows, is that both constitutional and common law rulings have been guided by the assumption that the conceptual distinction between legal and factual questions is both objective and related to the pragmatic question of who—judge or jury—is better placed to decide the issue.[52]

A. Traditional Conceptions of Law and Fact: Normative vs. Empirical, General vs. Particular

Before getting to the case law concerning normative questions, it will be helpful to have theories of the law/fact distinction on hand. On a widely-embraced view,

---

[48] *See, e.g.*, Tull v. United States, 481 U.S. 412, 417–18 (1987).
[49] *See* Chauffeurs Local 391 v. Terry, 494 U.S. 558, 564 (1990) for the two-prong test. *See also* Allen & Pardo, *supra* note 9, at 1779.
[50] *See* Allen & Pardo, *supra* note 9, at 1779.
[51] *See, e.g.*, Cooper Indus., Inc. v. Leatherman Tool Group, Inc., 532 U.S. 424, 437 (2001) (describing historical facts as paradigmatically factual and legal questions as essentially normative).

[52] It would be surprising if the conceptual distinction that judges have used did not serve the *practical* end of allocating questions between judge and jury in a useful way. Any account of the conceptual distinction should explain why the distinction is practically useful. But on this analytic approach, the order of explanation runs *from* the conceptual distinction to the pragmatic justification, not the other way.

questions of fact are *empirical* or *historical* questions concerning "who did what, where, when, how, why, with what intent or motive,"[53] while questions of law are *normative* questions involving "the establishment, disestablishment, modification, or interpretation of legal rules."[54] Judge Richard Posner provides the example of questions about what happened during the reign of Richard III. These have a different "ontological status" from questions about rules or norms derived from statutes, judicial opinions, and other such sources, which are legal.[55] Similarly, Richard Friedman suggests that ordinary "fact-finding" involves determining or describing some part of reality, whereas "law-discovery" always involves interpreting normative standards, such as: "cruel and unusual."[56] The normative vs. empirical conception of the law/fact divide has at times been cited approvingly by the Supreme Court.[57]

Most modern theorists acknowledge that legal questions remain *factual* at least in one sense—namely, there can be a determinate fact of the matter regarding what the law

---

[53] PIERCE, *supra* note 17, at 732. *See also* Allen, Brunet & Roth, *supra* note 17, at 1262 (2007) (reciting evidence that 'fact' has historically been understood to concern events in the external world capable of identification through empirical inquiry).

[54] Warner, *supra* note 16, at 112.

[55] RICHARD A. POSNER, THE PROBLEMS OF JURISPRUDENCE 198 (1990).

[56] Richard D. Friedman, *Standards of Persuasion and the Distinction Between Fact and Law*, 86 NW. U. L. REV. 916, 918. *See also* Bohlen, *supra* note 9, at 112 ("[Law is defined as a] body of principles and rules which are capable of being predicated in advance . . . awaiting proof of the facts necessary for their application."); Arthur W. Phelps, *What is a "Question of Law"?*, 18 U. CIN. L. REV. 259, 259 (1949) ("([A] legal system which postulates norms (roughly, rules and principles of law) must make some differentiation between a norm and the question of the existence of the facts which call for its application.").

[57] *See, e.g.*, Cooper Indus., Inc. v. Leatherman Tool Group, Inc., 532 U.S. 424, 437 (2001) ("Unlike the measure of actual damages suffered, which presents a question of historical . . . fact, the level of punitive damages is not really a 'fact' 'tried' by the jury." (citations omitted) (quoting Gasperini v. Center for Humanities Inc., 518 U.S. 415, 459 (1996) (Scalia J., dissenting)). *But see id*. at 446 (Ginsburg, J., dissenting).

is on some issue.[58] The law/fact distinction should be understood in terms of the difference between questions concerning legal facts, on the one hand, and non-legal facts on the other. The relevant question for theorists is what distinguishes legal from non-legal facts. It could, for example, have something to do with the difference between empirical and normative facts.

A related account distinguishes non-legal from legal facts based on their degree of specificity or particularity. Such an account can allow normative findings to count as "factual" so long as they are highly particular, for instance, a finding that a defendant was negligent based on a rich and complex combination of factors that were true in the individual case. By contrast, more general normative truths are law-like. They concern what is true in a wide range of cases, for instance, that a failure to comply with a statutory requirement designed to protect persons from harm is negligence *per se*.

Proponents of both the normative/empirical and general/specific accounts of the distinction tend to concede that "law" and "fact" may not be binaries. There may be types of questions that are not easily classified as either legal or factual because they resemble both.[59] What legislatures intend to do appears to be a historical/empirical fact.

---

[58] *See e.g.*, Richard D. Friedman, *Standards of Persuasion and the Distinction Between Fact and Law*, 86 NW. U. L. REV. 916, 917 (1992) (noting that the relevant difference is between legal and non-legal facts); Henry P. Monaghan, *Constitutional Fact Review*, 85 COLUM. L. REV. 229, 235 (1985). *See also* Allen & Pardo *supra* note 9, at 1792–94 ("[T]he answers to legal questions are propositional statements with truth value and are therefore, like other propositions with truth value, factual . . . . One can be objectively right or wrong about the rules of basketball or chess even though these are also human-made, linguistic constructs."). The law/fact distinction tracks questions concerning legal facts, on the one hand, and *non-legal* facts, on the other (the current use of "fact" in discussing the distinction will follow this usage, with "non-legal" occasionally added for emphasis). Thanks to Nomi Stolzenberg for pressing me to make this clear.

[59] *See* e.g., Walter Wheeler Cook, *'Facts' and 'Statements of Fact,'* 4 U. CHI. L. REV. 233, 244 (1937) ("[T]he time-honored distinction between 'statements of fact' and 'conclusions of law' is merely one of degree. . . ."); Monaghan, *supra* note 58, at 233 n.24.

Yet legislative intent is paradigmatically a question of law decided by judges in the course of interpreting statutory language. Thus, Henry Monaghan writes that the "distinction posited between 'law' and 'fact' does not imply the existence of static, polar opposites. Rather, law and fact have a nodal quality; they are points of rest and relative stability on a continuum of experience."[60]

The "mixed" status of certain questions needn't be a mark against a theory of the distinction. Many of our ordinary concepts track genuine differences in the world, even while failing to determinately apply in borderline cases. There may not be a fact of the matter regarding whether persons in an intermediate stage of hair loss count as bald or not, but that does not mean that there isn't a genuine difference between those who are bald and those who aren't. The crucial question for theorists is whether the features that determine whether a type of question is more legal than factual can be specified *ex ante*. The theoretical challenge is to say precisely which features characterize the 'nodes' and, accordingly, determine where on the continuum between legality and factuality a question lies. The relevant features may be generality or normativity, but the account should ideally square with the actual practice of judges tasked with interpreting the distinction.

B. The Domain Relative Treatment of Normative Questions

The trouble with existing accounts of the distinction is that they struggle to explain the actual case law.[61] The theory that normative questions are necessarily legal is

[60] Monaghan, *supra* note 58, at 233. *See also* Warner, *supra* note 16.

[61] *See* Allen & Pardo, *supra* note 9, at 1800–06.

hard to reconcile with the numerous cases in which normative issues are classified as questions of fact. In tort law, for example, it is a firmly entrenched rule that whether the defendant acted unreasonably and thereby breached a duty of care owed to the plaintiff is a question of fact for the jury.[62] Jury instructions routinely emphasize that "[t]he law does not say what a reasonably careful person using ordinary care would or would not do . . . ."[63] Instead, juries are responsible both for articulating the norm governing reasonable and unreasonable behavior in the context of subjecting others to risk *and* for applying the norm to the facts of the case. While the question of fact status of the negligence issue attracts a fair bit of academic criticism, including from those who think normative questions are necessarily questions of law, the Supreme Court has endorsed the classification, hinting that the important values at stake and the prevalence of reasonable disagreement on what constitutes negligence is part of the justification.[64] Other concepts in the tort context, like "proximate causation," whose application and interpretation are assigned to the jury, often turn out to be "cryptonormative"—that is, concepts whose application non-obviously involves determining normative questions.[65]

---

[62] *See* Weiner, *supra* note 1, at 1877, 1877 n.43 ("The courts of all American jurisdictions, with possibly one exception, adhere to this principle. . . . There are literally hundreds of cases in which this concept has been asserted."); *id.* at 1877 n.43 (citing cases invoking the rule).

[63] Patrick J. Kelley & Laurel A. Wendt, *What Judges Tell Juries about Negligence: A Review of Pattern Jury Instructions*, 77 CHI. KENT. L. REV. 587, 608 (2002).

[64] For critics of the common-law treatment of negligence, *see* Allen & Pardo, *supra* note 9, at 1781 n.76. *See also* Sioux City & Pac. R.R. Co. v. Stout, 84 U.S. 657 (1873) (endorsing the question of fact status of negligence).

[65] Patrick J. Kelley, *Proximate Cause in Negligence Law: History, Theory, and the Present Darkness*, 69 WASH. U. L. Q. 49, 51 (1991) (discussing the view that "proximate cause has nothing to do with causation and little to do with proximity" and instead refers to the "scope of duty in negligence cases"). On cryptonormative judgments generally, see Alex Worsnip, *Cryptonormative Judgments*, 25 EUR. J. OF PHIL. 3 (2017).

Whether the defendant's conduct proximately caused the plaintiff's injury can turn, among other things, on whether the injury was *reasonably* foreseeable.[66]

The treatment of normative questions as questions of fact is not confined to tort law. In criminal law, juries decide, often by constitutional mandate, such questions as whether the defendant's conduct was especially "cruel" or "heinous," which involves determining the normative significance of empirical facts (like the use of a type of weapon or injuries to bystanders).[67] Indeed, virtually all "sentencing factors" or factors used to increase sentences beyond statutory maximums have question-of-fact status under constitutional law. In an important line of cases beginning with *Apprendi v. New Jersey*, the Supreme Court has emphasized that the Sixth Amendment right to a jury trial includes the right to have the jury decide all questions of fact, including those concerning the presence of aggravating and mitigating factors that increase the maximum sentence the defendant can receive from that allowed by a finding of guilt alone.[68] More recently, in *Alleyne v. United States*, the Court held explicitly that the finding of virtually any fact

---

[66] *See* Kelley, *supra* note 65.

[67] *See* Blakely v. Washington, 542 U.S. 296, 303 (2004) ("[T]he 'statutory maximum' . . . is the maximum sentence a judge may impose solely on the basis of the facts reflected in the jury verdict or admitted by the defendant.") (emphasis omitted). The Court found the judge's finding of "deliberate cruelty" at sentencing to be an unconstitutional invasion of the jury's fact-finding role. *Id.* at 313–14. For a discussion of the trend of greater jury involvement in sentencing, see W. David Ball, *Heinous, Atrocious, and Cruel: Apprendi, Indeterminate Sentencing, and the Meaning of Punishment*, 109 COLUM. L. REV. 893, 902 (2009); Paul F. Kirgis, *The Right to a Jury Decision on Sentencing Facts After Booker: What the Seventh Amendment Can Teach the Sixth*, 39 GA. L. REV. 895 (2005).

[68] *See* United States v. Booker, 543 U.S. 220, 243–44 (2005) (validating *Apprendi* and extending it to the Federal Sentencing Guidelines); Ring v. Arizona, 536 U.S. 584, 589 (2002) ("Capital defendants, no less than noncapital defendants . . . are entitled to a jury determination of any fact on which the legislature conditions an increase in their maximum punishment."); Apprendi v. New Jersey, 530 U.S. 466, 490 (2000).

that increases a punishment in any way, including the statutory minimum, is a fact that must be found by the jury.[69]

In the capital sentencing context, jury-findings of "aggravating" and "mitigating" factors determine whether the defendant receives the death penalty.[70] In nearly all states that allow the death penalty, juries play a decisive (and often final) role in determining whether a capital defendant should be sentenced to death, based on a weighing of mitigating against aggravating factors.[71] The Court has emphasized the normative character of this final determination, suggesting that "in the final analysis, capital punishment rests on not a legal but an ethical judgment—an assessment of . . . the moral guilt of the defendant."[72]

Cases such as these strongly suggest that question-of-law status is unlikely to be a function merely of a question's normative character. Upon reflection, it is indeed hard to see why, given that in ordinary language we find it helpful to distinguish different types of normative questions (legal, moral, epistemic), *every* normative question raised under

---

[69] 133 S. Ct. 2151, 2158 (2013) (holding that any fact that increases the mandatory minimum must be submitted to the jury including normative findings of fact).

[70] *See Ring*, 536 U.S. at 609 (holding that aggravating factors present a question of fact for the jury). *See also* Stephen P. Garvey, "*As the Gentle Rain from Heaven": Mercy in Capital Sentencing*, 81 CORNELL L. REV. 989, 1003 n.56 (1996) ("Capital sentencing juries are said to represent the 'conscience of the community.' However, they 'represent' the community only because they are members of the community, not because they discern and then apply community standards.").

[71] In 27 of the current 32 death penalty states, the jury's decision to sentence a defendant to life imprisonment is final and cannot be overridden by a trial judge. *See* Woodward v. Alabama, 134 S. Ct. 405, 407 (2013) (Sotomayor, J., dissenting from denial of certiorari).

[72] Spaziano v. Florida, 468 U.S. 447, 481 (1984) (Stevens, J., concurring in part and dissenting in part). *See also* McCleskey v. Kemp, 481 U.S. 279, 310 (1987) (outlining the capital jury's task of expressing "the conscience of the community on the ultimate question of life or death"); Woodson v. North Carolina, 428 U.S. 280, 297–98 (1976) (reflecting on the importance on the moral views of society in the administration of death penalty); Witherspoon v. Illinois, 391 U.S. 510, 519 (1968) ("[A] jury that must choose between life imprisonment and capital punishment can do little more—and must do nothing less—than express the conscience of the community on the ultimate question of life or death.").

the law would be, simply for being normative, a *legal* question. For instance, we distinguish moral questions from questions of etiquette or prudence.

It is also not the case that those who support the normative/empirical theory have the distinction backwards: the case law offers many examples of normative questions classified as legal.[73] Contract law, and more generally laws governing conduct by economic actors (such as unfair competition law), provide several illustrations.[74] Judges are often asked to fill the gaps in contracts with "implied terms," and in doing so draw upon considerations of fairness/reasonableness.[75] Thus, § 204 of the Restatement explicitly directs courts to imply terms that comport with "community standards of fairness and policy,"[76] and § 2-204(3) of the U.C.C. says courts should supply terms if "there is a reasonably certain basis for giving an appropriate remedy."[77] Courts acknowledge that the reasonableness of implied terms is a question of law just as universally as they do that it is a normative question.[78] In Pennsylvania, for instance, the

---

[73] For a comparison of the treatment of normative issues in torts and contracts, see Mark P. Gergen, *The Jury's Role in Deciding Normative Issues in the American Common Law*, 68 FORDHAM L. REV. 407 (1999). The disparate treatment of normative issues under the law/fact distinction has been widely commented on. *See, e.g.,* Allen & Pardo, *supra* note 9, at 1781–83; Gergen, *supra*; Weiner, *supra* note 1, at 1893–95.

[74] *See generally* William C. Whitford, *The Role of the Jury (and the Fact/Law Distinction) in the Interpretation of Written Contracts*, 2001 WIS. L. REV. 931 (2001); *id.*

[75] *See* Allen & Pardo, *supra* note 9, at 1782; Gergen, *supra* note 73, at 443; Charles T. McCormick, *The Parole Evidence Rule as a Procedural Device for Control of the Jury,* 41 YALE L.J. 365 (1932).

[76] RESTATEMENT (SECOND) OF CONTRACTS § 204, cmt. d. (AM. LAW INST. 1981).

[77] U.C.C. § 2-204(3) (AM. LAW INST. & UNIF. LAW COMM'N 2011).

[78] Judge Posner observes that "whether we say that a contract shall be deemed to contain such implied conditions as are necessary to make sense of the contract, or that a contract obligates the parties to cooperate in its performance in 'good faith' to the extent necessary to carry out the purposes of the contract, comes to much the same thing." Market St. Associates v. Frey, 941 F.2d 588, 596 (7th Cir. 1991). *See also* Tymshare, Inc. v. Covell, 727 F.2d 1145, 1154 (D.C. Cir. 1984); CHARLES FRIED, CONTRACT AS PROMISE 75 (1981) ("[I]t seems as if contractual relations depend not on the will of the parties but on externally imposed substantive moral judgments of what the relations between the parties should be."); Randy E. Barnett, . . . *And Contractual Consent,* 3 S. CAL. INTERDISC. L.J. 421, 427 (1993); Larry A. DiMatteo, *The Counterpoise of*

doctrine of "necessary implication" states that "[i]n the absence of an express term, [this doctrine] . . . may act to imply a requirement *necessitated by reason and justice* without which the intent of the parties is frustrated."[79]

Similarly, the "unconscionability" of contractual terms is a question of law.[80] Judges can refuse to enforce a contract deemed unconscionable. According to the U.C.C., "[t]he basic test is whether, in the light of the general commercial background and the commercial needs of the particular trade or case, the clauses involved are so one-sided as to be unconscionable under the circumstances existing at the time of the making of the contract . . . The principle is one of the prevention of oppression and unfair surprise . . . ."[81] This manifestly normative enquiry involves appraisal of both the procedural fairness of a contract, having to do with symmetry of information and bargaining strength of the contracting parties, as well as substantive fairness, having to do with the relative gains

---

Contracts: The Reasonable Person Standard and the Subjectivity of Judgment, 48 S.C. L. REV. 293, 298 (1997) ("The reasonable person, under the banner of sanctity of contract or that of fairness, is used to fill gaps in otherwise inchoate contracts."); *id.* at 299 ("The implication of society's rules of fairness and reasonableness is generally accomplished through the courts' fabrication of the reasonable person."); Harold Dubroff, *The Implied Covenant of Good Faith in Contract Interpretation and Gap-Filling: Reviling a Revered Relic*, 80 ST. JOHN'S L. REV. 559, 616 (2012) (observing three categories of implied terms including "terms imposed by the court for reasons of policy or fairness, or in obedience to rules of law.").

[79] Emily M. S. Houh, *The Doctrine of Good Faith in Contract Law: A (Nearly) Empty Vessel?*, 2005 Utah L. Rev. 1, 15 (discussing doctrine of "necessary implication" as referenced in Somers v. Somers, 613 A.2d 1211, 1214 (Pa. Super. Ct. 1992)).

[80] U.C.C. § 2-302(1) (AM. LAW INST. & UNIF. LAW COMM'N 2011) ("If the court as a matter of law finds the contract or any clause of the contract to have been unconscionable at the time it was made the court may refuse to enforce the contract . . . ."); U.C.C. § 2-302(2) (AM. LAW INST. & UNIF. LAW COMM'N 2011). *See also* Landsman Packing Co. v. Cont'l. Can Co., 864 F.2d 721, 729 (11th Cir. 1989) (reversing lower court's decision to allow the jury to decide unconscionability).

[81] UCC § 2-302 cmt. 1 (AM. LAW INST. & UNIF. LAW COMM'N 2011). It is also important to note that an unconscionability finding by a judge is not equivalent to a finding that no reasonable juror could find the contract conscionable. It is a determination made entirely at the judge's discretion and not controlled by whether reasonable persons might disagree.

and losses allocated to either side.[82] Although the U.C.C.'s enactment makes contractual unconscionability a question of law by legislative fiat, the U.C.C.'s drafters drew on the historic practices of common law judges in deciding the issue. As Donald Price notes, juries were excluded from unconscionability determinations well before the issue was legislatively assigned to judges.[83]

Another helpful example is the treatment of punitive damages in economic contexts. In *Cooper Industries, Inc. v. Leatherman Tool Group, Inc.*, which involved unfair competition and false advertising claims, the jury awarded $4.5 million in punitive damages, and the issue on appeal was the appropriate standard of review in assessing the damages awarded for proportionality.[84] The Supreme Court held that *de novo* review—the standard for questions of law—was appropriate for the proportionality of punitive damages. It deemed the finding of punitive damages to have intermediate status between law and fact, noting in particular the element of moral condemnation involved.[85] As we have noted, however, moral findings, including those involving condemnation of a defendant's conduct, are routinely treated as factual and reviewed deferentially, a fact that Justice Ginsburg emphasized in her dissent in *Cooper*.[86] For reasons not entirely

---

[82] *See* E. ALLAN FARNSWORTH, FARNSWORTH ON CONTRACTS (1990) (defining procedural unconscionability in terms of the "absence of meaning choice on the part of one of the parties"); MARGARET J. RADIN, BOILERPLATE: THE FINE PRINT, VANISHING RIGHTS, AND THE RULE OF LAW 125 (2013) ("[s]ubstantive unconscionability refers to defects in the bargain itself: the notion that some contracts may look so one-sided or unequal or oppressive that the court in good conscience simply should not tolerate enforcing them."); Arthur Allen Leff, *Unconscionability and the Code—The Emperor's New Clause*, 115 U. PA. L. REV. 485 (1967).
[83] Donald R. Price, *Conscience of Judge and Jury: Statutory Unconscionability as a Mixed Question of Law and Fact*, 54 TEMPLE L. Q. 743, 745–48 (1981).
[84] Cooper Indus., Inc. v. Leatherman Tool Grp., Inc., 532 U.S. 424, 426 (2001).
[85] *Id.* at 437.
[86] *Id.* at 446 (Ginsburg, J., dissenting).

clear, the moral condemnation in this instance and the normative question raised at trial was deemed to be law-like.

If the normative/empirical account of the distinction fails to explain the case law, the general/specific account does not fare much better. Proponents of the latter view think that normative issues are questions of law when they implicate general as opposed to specific normative considerations. However, it is far from obvious that contractual norms interpreted and applied by judges have a greater degree of generality than those arising in the tort and criminal context.[87] There have been few attempts to unify under general normative rules' findings of contractual unconscionability or reasonableness.[88] In fact, scholarship on unconscionability tends to emphasize the highly case-specific nature of the enquiry.[89] Courts seem to agree: "the precise number of days . . . which will constitute a 'reasonable time,' . . . depend[s] upon circumstances as variable and uncertain as are the transactions and characters of men; and finally to be determined by the discretion, not to say, caprice of the Court."[90] Moreover, normative questions

---

[87] *See* sources cited *supra* note 9.

[88] *See, e.g.*, M. Neil Browne & Lauren Biksacky, *Unconscionability and the Contingent Assumptions of Contract Theory*, 2013 MICH. ST. L. REV. 211, 222 ("Procedural unconscionability can result from any of the following elements: (1) absence of meaningful choice; (2) superiority of bargaining power; (3) the fact that the contract is an adhesion contract; (4) unfair surprise; or (5) sharp practices and deception."); Robert A. Hillman, *Debunking Some Myths About Unconscionability: A New Framework for U.C.C. Section 2-302*, 67 CORNELL L. REV. 1 (1981).

[89] Browne & Biksacky, *supra* note 88, at 215–16 (arguing that the policing of unconscionability by the courts has been "inherently contradictory . . . since the inception of the doctrine").

[90] Goodwin v. Davenport, 47 Me. 112, 117–18 (1860) (emphasis omitted). *See also* A.N. Deringer, Inc. v. Strough, 103 F.3d 243, 246 (2d Cir. 1996) (invalidating a geographical limitation on competition for unreasonableness and supplying a new limitation); Seaver v. Lincoln, 38 Mass. 267, 268–69 (1838) (noting that "one decision goes but little way in establishing a precedent for another"); Commercial Nat'l Bank v. Zimmerman, 185 N.E. 210, 217 (N.Y. 1906) ("[W]hat constitutes reasonableness of time . . . cannot be determined by any fixed rules."); U.C.C. § 2-204(3) (2011); RESTATEMENT (SECOND) OF CONTRACTS § 204 (AM. LAW INST. 1981).

classified as factual do seem to be decided under general moral principles. For instance, jurors in criminal sentencing seem to decide issues on the basis of general maxims, as evidenced by patterns in the treatment of emotional disturbance and severe environmental deprivation (or SED) evidence.[91] It is, to say the least, a controversial notion that the varying classification of normative questions as legal or factual is to be explained in terms of differences in the generality or specificity of the normative truths in question.

There has been a tendency amongst theorists to treat the case law that is hard to square with theory as exceptional in one of two ways. The more common approach has been to acknowledge inconsistencies and explain them as cases of judges ignoring the law/fact distinction for case-specific pragmatic reasons but failing to do so explicitly.[92] For instance, a judge may refer to what is really a question of fact as a legal question if she feels that judges are uniquely qualified to decide the issue. Randall Warner suggests

---

[91] *See* Michelle E. Barnett, Stanley L. Brodsky, & Cali Manning Davis, *When Mitigation Evidence Makes a Difference: Effects of Psychological Mitigating Evidence on Sentencing Decisions in Capital Trials*, 22 BEHAV. SCI. & L. 751 (2004) (finding that mock jurors are less likely to assign a death sentence in light of SED evidence); Stephen P. Garvey, *Aggravation and Mitigation in Capital Cases: What Do Jurors Think?*, 98 COLUM. L. REV. 1538, 1564–66 (1998) (finding based on empirical study that mental or emotional disturbance is treated as highly mitigating and poverty and deprivation are assigned mitigating weight by half of jurors).

[92] *See, e.g.*, Bohlen, *supra* note 9, at 115 (arguing that assignment of the negligence issue to juries is not based on strict application of the law/fact rule); Friedman, *supra* note 56, at 922 ("We should not be fooled [by negligence]. The jury in such a case does more than determine an aspect of reality. It also determines the norms that will be applied in that case."); Oliver W. Holmes, *Law in Science and Science in Law*, 12 HARV. L. REV. 443, 457 (1899) ("I venture to think . . . that every time that a judge declines to rule whether certain conduct is negligent or not he avows his inability to state the law . . . if a question of law is pretty clear we can decide it, as it is our duty to do, if it is difficult it can be decided better by twelve men taken at random from the street."); Monaghan, *supra* note 58, at 232 n.22 ("[T]he allocation of negligence questions to the jury rests on grounds of policy, not on abstract conceptions of the intrinsic nature of the question itself."); *id.* at 234–35 ("The difficulty comes when the judges seek to force such [pragmatic] allocation decisions into the conventional categories of law and fact. Distortions in the analytic content of the categories occur."). *See also* Antilles S.S. Co. v. Members of Am. Hull Ins. Syndicate, 733 F.2d 195, 205–06 (2d Cir. 1984) (Newman, J., concurring) (arguing that negligence is left to the jury for practical reasons).

that "evaluative determinations" are "special" and classification by judges of an evaluative issue as law or fact does not turn on the essence of the law/fact distinction but instead on "a policy choice concerning the judicial actor better positioned to decide a particular issue."[93] What lends support to this view is the frequency with which judges justify their classificatory choices by appeal to prudential considerations. In deeming a question to be one of law, judges appeal to factors like the complexity of an issue or the need for uniformity and predictability.[94] They might also appeal to the importance of bringing community sentiment to bear on an issue in assigning a question to the jury.[95] It is far from obvious whether such pragmatic concerns have anything to with the essential difference between legal and factual questions.[96]

There are several reasons why theorists should resist the temptation to explain away difficult case law as judges, letting prudential considerations that are both case-specific and unhinged from the conceptual differences between legal and factual questions, guide classification. The case law keeps legal theorizing from devolving into data-free speculation. It is therefore important to try and explain as much of it as one can by appealing to general theory. If judges routinely use case-specific pragmatic analysis rather than the analytic distinction to decide the law/fact issue, then this raises the possibility that the analytic distinction itself does no real work. Instead, 'law' and 'fact' may just be labels judges use to indicate their independently reached pragmatic conclusions regarding who should decide an issue. As Allen and Pardo write, "[t]he

[93] Warner, *supra* note 16, at 130.
[94] *See, e.g.*, Markman v. Westview Instruments, Inc., 517 U.S. 370, 390–91 (1996).
[95] *See, e.g.*, Sioux City & Pac. R.R. Co. v. Stout, 84 U.S. 657, 664 (1873).
[96] *See* Allen & Pardo, *supra* note 9, at 1782.

extent to which pragmatic considerations determine the allocative question is plain in these areas. . . . [B]ut this does not make 'legal' issues out of factual issues . . . ."[97] The pragmatic considerations may have an important role to play in reinforcing classification. But our theory of legality and factuality should be consistent with as much of the difficult case law as possible.

Another reason that the pragmatic account of the controversial cases is less than satisfying is that the prudential reasons cited for assigning an issue to the judge or jury tend be less than persuasive.[98] Take, for instance, the view that judges' expertise at interpreting written documents makes them especially good at discerning the reasonableness of implied contractual terms, given the central role of the explicit contractual language in determining such reasonableness. In cases of genuine gap-filling, there is often no clear language in the contract that speaks to the issue, and judges must rely on considerations unrelated to exegesis, including the going rates in the industry and their own sense of fairness. Similarly, it is often claimed that judges tend to be uniform in their decisions, which results in greater predictability. Yet there are reasons to doubt that this holds true. The research suggests that juries are often as predictable as judges.[99]

A different approach that is sometimes taken by theorists in response to the differential treatment of normative questions is to treat them as "mixed questions of law

---

[97] *Id.* at 1783.

[98] *Id.* at 1782 ("The second rationale for this rule–administrative concerns–involves the need for uniformity and predictability with frequently reoccurring fact patterns. . . . No justification has been given, however, as to why this only applies to contracts and not to other areas such as negligence, or any other area for that matter.").

[99] *See* Kimberly A. Moore, Markman *Eight Years Later: Is Claim Construction More Predictable?*, 9 LEWIS & CLARK L. REV. 231 (arguing, based on empirical studies, that the transfer of patent claim construction from a question of fact for the jury to one of law for the judge has not made the law more predictable)

and fact," or questions that share features of both and can therefore be classified as either.[100] As discussed earlier, an analytic theory of the law/fact distinction can accommodate "mixed questions" as well as the idea that questions can have both law-like as well as factual characteristics. Crucially, however, the theorist needs an account of the features in virtue of which a question counts as mixed. In other words, the theory needs to explain what determines the more law-like character of normative fact-finding in the contractual or economic domain, for example, than that of normative fact-finding in capital sentencing proceedings. To say that the question of reasonableness in contracts is more law-like because judges decide it would be circular. On the analytic approach, the order of explanation must run the other way: one must appeal to essential differences between legal and factual questions to explain *why* judges should decide the issue. Unfortunately, there is no generally agreed upon theory of mixed questions in the academic literature or elsewhere. Courts have reasonably responded by describing mixed questions as "elusive abominations"[101] and the overall jurisprudence in this area as "lack[ing] clarity and coherence."[102]

C. Changes in Classification Over Time

Another challenge for theorists is to explain why normative questions of fact can become legal over time and vice versa. The Supreme Court's First Amendment jurisprudence provides several examples of such transitions. In certain cases, the Court

---

[100] *See, e.g.*, Warner, *supra* note 16.
[101] S & E Contractors, Inc. v. United States., 433 F.2d 1373, 1378 (Ct. Cl. 1970), *rev'd,* 406 U.S. 1 (1972).
[102] United States v. McConney, 728 F.2d 1195, 1200 (9th Cir. 1984), *abrogated by* Pierce v. Underwood, 487 U.S. 552 (1988).

will legitimize *de novo* review (the standard for questions of law) of issues historically treated as factual and reviewed deferentially. For instance, in the context of defamation actions, whether a false statement was made with "actual malice" or "reckless disregard of truth or falsity" was traditionally a question of fact reviewed deferentially.[103] In *New York Times v. Sullivan*, the Court made a finding of actual malice a condition on punitive damages in defamation actions brought by public officials.[104] Shortly thereafter, the Court in *Bose Corp. v. Consumers Union* determined that, in light of its previous holding and the important constitutional right of free speech that turns on it, the malice question was intermediate between law and fact—a "constitutional fact," and, accordingly, should be reviewed *de novo* by appellate courts.[105]

Similar examples of *de novo* review of what used to be paradigmatic factual findings abound in other areas of First Amendment law. The normative question of whether speech appeals to "prurient interest" or is "patently offensive," a finding that precludes protection under the First Amendment, is reviewed *de novo*.[106] Prior to the

[103] Oakes, *supra* note 6, at 688–98 (observing a long history of jury findings on the question of malice in civil and criminal defamation cases).
[104] New York Times v. Sullivan, 376 U.S. 254, 267 (1964).
[105] *See* Bose Corp. v. Consumers Union, Inc., 466 U.S. 485, 508–10 n.27 (1984). In recent years, the Court has emphasized that actual malice requires "material" falsity. Air Wis. Airlines Corp. v. Hoeper, 134 S. Ct. 852, 861 (2014).
[106] *See* Jenkins v. Georgia, 418 U.S. 153, 159–61 (1974) (reviewing *de novo* and reversing a unanimous jury determination that the movie *Carnal Knowledge* was patently offensive while recognizing that the determination was factual); Miller v. California, 413 U.S. 15, 25 (1973) ("[T]he First Amendment values . . . are adequately protected by the ultimate power of appellate courts to conduct an independent review of constitutional claims when necessary"); Roth v. United States, 354 U.S. 476, 497 (1957) (suggesting that whether an attacked expression is suppressible is a problem requiring "particularized judgments which appellate courts must make for themselves") (Harlan, J., concurring in the result, dissenting in part). There is some controversy as to which part of the 'prurience' determination calls for *de novo* review. Under *Miller*, prurience turns on what is ordinarily found to be prurient, whether the work 'offensively' depicts sexual conduct specifically defined in statutory law, and whether the work lacks serious value. *See* United States v. Various Articles of Merch., 230 F.3d 649, 653 (3d Cir. 2000).

Supreme Court's seminal cases on obscenity and free speech in the 1970s, jury findings of obscenity were deferentially reviewed as paradigmatic findings of fact for decades by appellate courts at the state and federal level, and the Court had explicitly declined to review findings *de novo*.[107] Similarly, whether "fighting words" are "inherently inflammatory," a finding that determines whether or not the speech is entitled to constitutional protection, is reviewed *de novo* as a "constitutional fact" intermediate between law and fact.[108] Outside the free speech context, the Court has been far more reluctant to invoke the constitutional fact rationale, even where normative findings, classified as factual, trigger important constitutional rights. The Court declined to require that questions of discriminatory intent in racial discrimination cases be reviewed *de novo* as questions of law.[109]

---

[107] Alexander v. United States, 271 F.2d 140, 146 (8th Cir. 1959) ("The primary responsibility for determining the obscenity issue is upon the jury. . . . The jurors are entitled to make their own evaluation of the books upon the basis of all the evidence before them . . . ."); United States v. Roth, 237 F.2d 796, 801 (2d Cir. 1956), *aff'd*, 354 U.S. 476 (1957) (finding that a reasonable jury could find materials obscene); Commonwealth v. Isenstadt, 62 N.E.2d 840, 847 (Mass. 1945) ("The principal question in the case is whether . . . we can say as matter of law that an honest jury . . . would not be acting as reasonable men in concluding beyond a reasonable doubt that this book, taken as a whole, possesses the qualities of obscenity, indecency, or impurity. The test is not what we ourselves think of the book, but what in our best judgment a trier of the facts might think of it without going beyond the bounds of honesty and reason."); *id*. (finding that a reasonable jury could find the publication obscene). Justice Harlan, in his partial dissent in *Roth*, objected to the Court's failure to review the obscenity question *de novo*: "I do not understand how the Court can resolve the constitutional problems now before it without making its own independent judgment upon the character of the material upon which these convictions were based." *Roth*, 354 U.S. at 497–98 (Harlan, J., concurring in the result, and dissenting in part). *See also* Whitney Strub, *Slouching Towards Roth: Obscenity and the Supreme Court, 1945-1957*, 38 J. SUP. CT. HIST. 121 (2013) (noting that prior to *Roth* the Court had last substantively weighed on the obscenity issue in 1896).

[108] Street v. New York, 394 U.S. 576, 592 (1969) (exercising independent judgment on the question).

[109] Pullman-Standard v. Swint, 456 U.S. 273, 286–88 (1982) ("[W]hether the defendants had intentionally maintained a racially segregated school system at a specified time in the past. . . . [is] essentially factual, subject to the clearly-erroneous rule."); *id.* at 287–88 ("That question, as we see it, is a pure question of fact."); Rogers v. Lodge, 458 U.S. 613, 622–23 (1982) (applying "clearly erroneous standard" to review issue of discrimination in election rules). *But see* Norris v.

To put the issue in terms of the law/fact distinction, what is needed is some account of why certain normative findings have come to be treated as mixed questions of fact and law whereas others haven't. As several commentators have pointed out, mere appeal to the importance of implicated constitutional values cannot explain disparities in the way the doctrine has been applied.[110] Moreover, what is needed is an account of why so-called constitutional facts were previously regarded as wholly factual, receiving deferential review, if they were "constitutional" all along. To quote Allen and Pardo, "[w]e suppose it is possible to confuse lions with zebras, even when staring at them for a couple of centuries, but it is unlikely."[111] The theorist can try to explain these changes in terms of mistake or error, but crucially, an account is needed for why judges might have been mistaken and why confusing certain mixed or law-like factual questions for pure questions of fact is not like confusing lions for zebras.

Additionally, there are cases of judge-decided legal issues that come over time to be treated as questions of fact for the jury. To take a relatively recent example, consider the increased role of the jury in finding facts that bear on criminal sentencing outside of statutory requirements/guidelines. Judges previously had discretion to raise or lower a defendant's sentence based on factors that militate in favor of a harsher sentence. In *Apprendi*, the Court prohibited judges from enhancing criminal sentences beyond statutory maxima based on facts other than those decided by the jury under the Sixth

---

Alabama, 294 U.S. 587 (1935) (reviewing *de novo* racial discrimination in the exclusion of jurors from state juries). On discrimination being a moral category, see BENJAMIN EIDELSON, DISCRIMINATION AND DISRESPECT (2016).

[110] *See, e.g.*, Allen & Pardo, *supra* note 9, at 1786; Monaghan, *supra* note 58, at 266–67.

[111] Allen & Pardo, *supra* note 9, at 1784.

Amendment jury trial right.[112] The rule was further extended to the capital sentencing context in *Ring*, where the Court held that judicial determination of a capital defendant's "death eligibility" based on "aggravating factors" was inconsistent with *Apprendi*.[113] The majority emphasized that the Sixth Amendment is not a limitation on judicial power but a reservation of jury power. It limits judicial power only to the extent that it infringes on the fact-finding responsibility of the jury. Determining sentencing factors often involves normative assessment. Whether a murder was "Cold, Calculated, and Premeditated" or "Heinous, Atrocious, and Cruel" turns not just on empirical facts, like the use of a weapon or injuries to bystanders, but on an assessment of the degree to which such facts bear on the heinousness of the crime and, ultimately, militate in favor of a harsher sentence.[114]

Prior to *Ring* and *Apprendi*, there was far greater judicial involvement in determining the existence of factors that determine appropriate punishment.[115] In *Ring*, Justice Ginsburg identified five states in which capital sentencing, including evaluation of aggravating and mitigating factors, was entirely the responsibility of judges.[116] Since the Court's decisions constitutionally mandating jury determination of sentencing factors, the role of the jury has grown. In almost all states the jury's decision to sentence a defendant

[112] Apprendi v. New Jersey, 530 U.S. 466 (2000).
[113] Ring v. Arizona, 536 U.S. 584, 609 (2002) ("The right to trial by jury guaranteed by the Sixth Amendment would be senselessly diminished if it encompassed the factfinding necessary to increase a defendant's sentence by two years, but not the factfinding necessary to put him to death.")
[114] *See id.* at 592 n.1.
[115] *See* Morris B. Hoffman, *The Case for Jury Sentencing*, 52 DUKE L.J. 951, 963–68 (2003); Kirgis, *supra* note 67, at 897–88 (noting that judges routinely made determinations after the defendant's guilt for a crime had been established to determine an appropriate sentence).
[116] *Ring*, 536 U.S. at 608 n.6.

to life imprisonment is *final* and cannot be overridden by a trial judge.[117] Only Alabama affords judges the power to override the jury and independently decide the death-deservingness question.[118] The constitutionality of the practice is suspect, an issue we shall return to later. The crucial point for present purposes is that the law has trended in the direction of limiting the judge's role in determining aggravating factors that affect criminal sentencing. Increasingly, any normative finding that bears on a defendant's sentence has come to be viewed as factual rather than legal, even though courts used to assign such findings to the judge.

D. Skepticism About the Distinction's Overall Coherence

The aspects of law/fact jurisprudence we have considered mainly concern the challenge that *normative* questions pose for courts grappling with the distinction. But, as Randall Warner points out, courts have struggled not just with evaluative or normative questions.[119] In *res judicata* disputes, for example, the identity of a cause of action with a previously litigated one was historically a question of fact, but came be to be viewed in

---

[117] *See* discussion *infra* Part IV. *See generally* Jeffrey Abramson, *Death-Is-Different Jurisprudence and the Role of the Capital Jury*, 2 OHIO ST. J. CRIM. L. 117 (2004); Bryan A. Stevenson, *The Ultimate Authority on the Ultimate Punishment: The Requisite Role of the Jury in Capital Sentencing*, 54 ALA. L. REV. 1091 (2003).

[118] *See* discussion *infra* Part IV.

[119] *See* Warner, *supra* note 54. *See also* Rebecca Haw Allensworth, *Law and the Art of Modeling: Are Models Facts?*, 103 GEO. L.J. 825 (2015) (arguing that scientific predications based on empirical models should not be treated as findings of fact); Steven J. Madrid, Note, *Annexation of the Jury's Role in Res Judicata Disputes: The Silent Migration from Question of Fact to Question of Law*, 98 CORNELL L. REV. 463, 465 (2013) (arguing that the law/fact distinction has been misapplied in res judicata cases); John O. McGinnis & Charles W. Mulaney, *Judging Facts Like Law*, 25 CONST. COMMENT. 69 (2009) (arguing that courts should treat congressionally determined "social facts" as special and review the findings *de novo*); Rebecca Sharpless, *Fitting the Formula for Judicial Review: The Law-Fact Distinction in Immigration Law*, 5 INTERCULTURAL HUM. RTS. L. REV. 57, 60 (2010) (arguing that courts have misapplied the distinction).

the mid-18th century as a question of law, seemingly without acknowledgement by judges of the inconsistency.[120] This "chaotic legal landscape" leads critics of the distinction to conclude that the law/fact distinction is a "legal fiction," that "the quest to find 'the' essential difference between [law and fact] . . . that can control subsequent classifications of questions as legal or factual is doomed from the start, as there is no essential difference."[121]

The aim, in what follows, is to demonstrate that the case law is not as unprincipled as it seems. As far as the treatment of normative questions is concerned, there is a logic implicit in the holdings that is very much connected to the essential difference between legal and non-legal questions. There are several reasons for the present focus on normative questions, in particular—questions that invoke concepts like reasonableness, fairness, aggravation, and the like. First, the ultimate goal of this paper is not to fully characterize the boundaries of legality and factuality.[122] The goal is to provide a richer understanding of how the distinction has been and should be applied by judges in a range of hard cases that have been the focus of much recent discussion. Second, the proper treatment of normative questions under the distinction is an issue of unique importance to recent constitutional controversies, many of which turn on whether the jury or judge should be deciding a question of fairness or reasonableness. The analysis bears on the proper resolution of these controversies.[123] I am certain that, ultimately, more will need to be said to defend the theory put forward and its fit with

[120] *See* Madrid, *supra* note 119.
[121] Allen & Pardo, *supra* note 9, at 1770. *See also* sources cited *supra* note 9.
[122] In general, vindicating the usefulness of a concept and showing how it applies in hard cases does not require fully characterizing necessary and sufficient conditions for its application.
[123] *See* discussion *infra* Part IV.

decisions made under the rule. Moreover, an account will still need to be given of how non-normative or *empirical* questions fit under the rule, a question I have ignored altogether.[124] But in showing (i) that the skeptical case against the analytic distinction is considerably weaker than has often been assumed and (ii) that the distinction remains doctrinally relevant, the discussion to follow should be of significant interest to both critics and proponents of the judicial practice of distinguishing questions of law from questions of fact.

II. A Framework for the Classification of Normative Questions under the Law/Fact Distinction

This section's aim is chiefly theoretical: to distinguish essentially convention-dependent from convention-independent norms. Having defended the normative distinction, I offer the beginnings of an argument for construing the categories of law and fact in terms of it. There are both conceptual and pragmatic reasons for treating convention-dependent normative questions as questions of law and convention-

[124] It makes sense to theorize in this piece-meal fashion about the law/fact distinction because the distinction between normative and empirical facts is a robust one. I see no pre-theoretic reason for assuming that how the law/fact distinction is applied in the normative context must also explain how it is applied in the empirical domain. No doubt I am influenced in this judgment by my own meta-normative views. But I believe it to be a plausible assumption common to many contemporary views on the metaphysics of the normative domain, that the normative is distinct from the empirical.

independent ones as questions of fact. Later sections will explore whether and to what extent the theory helps us explain actual judicial practice.

A. Convention-dependent & Convention-independent Normative Questions

While all normative truths depend (or hold in virtue of) non-normative or empirical facts, normative truths can be distinguished based on the kinds of non-normative facts on which they depend.[125] For instance, there are normative truths that are hard to discern because they depend on a wide range of complex empirical facts.[126]

---

[125] The framework relies critically on a general truth about the normative domain: namely that normative facts *supervene* on (or are made true by) other, non-normative facts. *See* Jaegwon Kim, *Concepts of Supervenience*, 45 PHIL. & PHENOMENOLOGICAL RES. 153 (1984). Normative supervenience is taken as an uncontroversial starting point in meta-ethics. *See* Gideon Rosen, *What is Normative Necessity* 1 (2014) (manuscript) (describing it as the "least controversial thesis in metaethics"); Michael Smith, *Does the Evaluative Supervene on the Natural?, in* ETHICS AND THE A PRIORI 208 (2004) ("The supervenience of the evaluative on the natural thus purports to operate as a conceptual constraint on evaluative judgment. This too is accepted by nearly everyone writing about the nature of value.").

[126] On the nature of objectivity or truth in the normative domain, there is considerable disagreement within moral philosophy. *See* T.M. SCANLON, BEING REALISTIC ABOUT REASONS 1–15 (2014); Michael Smith, *Meta-ethics*, *in* THE OXFORD HANDBOOK OF CONTEMPORARY PHILOSOPHY (Michael Smith & Frank Jackson eds., 2005). A principal source of disagreement is whether moral or normative judgments represent facts in the way that ordinary, descriptive judgments do. According to non-cognitivists, moral judgments express desire-like mental states—e.g., the judgment that it is good to promote happiness in the world merely expresses the judging agent's desire to promote happiness in the world. The function of moral language is to express the relevant "non-cognitive" mental states. Non-cognitivists earn our right to talk about facts in the normative or evaluative domain by combining a desire-based view of moral judgment with a deflationary account of truth and factuality, where the judgment "it is a *fact* that promoting happiness is morally good" simply amounts to judging that it is good to promote happiness. In other words, by using the fact-based language we express the very desire-like attitude that constitutes moral judgment. *See, e.g.*, SIMON BLACKBURN, RULING PASSIONS: A THEORY OF PRACTICAL REASONING (1998). Cognitivists oppose this sort of view and very much take moral thought and talk to involve belief-like, representational states, such as the belief that Yale University is in New Haven, Connecticut. On one version of cognitivism ("moral non-naturalism"), moral thought and talk describes a part of reality consisting of properties and relations neglected by the natural sciences (irreducibly *moral* properties and relations).
As in the case of legal philosophy, we can mostly ignore such disagreements, for the only sense in which normative facts need to be objective, for our purposes, is in the sense of being fixed by factors other than judicial preference. All talk of normative/moral truth and fact in this article is meant to be neutral between cognitivist and non-cognitivist theories. Non-cognitivism provides

Whether it is good for a market economy to encourage highly self-interested and even exploitative economic behavior depends on a host of non-normative empirical facts pertaining to the consequences of such behavior for the overall economy, standards of living, disparities in income, the gravity of the harms to individuals who are exploited, etc. By contrast, there are normative truths that depend on relatively simpler or more accessible empirical facts. The wrongness of torturing someone for the sheer fun of it follows directly from the nature of pain or those features of a person in virtue of which they have moral status. For instance, under a Kantian conception of morality, the prohibition on torturing persons is made true by the dignity that persons have simply by virtue of their capacity for free and rational agency.[127] Our ordinary sense of what people are owed as a matter of right is presumably informed by our sensitivity to such basic features of persons and their mental states, features we recognize as morally relevant as a matter of course.

Some normative truths hold partly in virtue of law-related conventions like legislative enactments and judicial practice. To take a prosaic example, consider the reasons we have to drive on the right side of the road. The normative truth that one ought to drive on the right is at least partly determined (or made true) by the fact that we have settled on a convention of driving on the right. Of course, following the convention has various benefits including, first and foremost, motorist safety, and securing those benefits

---

all the resources to sustain the distinctions within the normative domain that will be discussed in what follows.

[127] *See, e.g.*, David Sussman, *What's Wrong with Torture?*, 33 PHIL. & PUB. AFF. 1 (2005); *see also* IMMANUEL KANT, GROUNDWORK OF THE METAPHYSICS OF MORALS 429–30 (1785); Onora O'Neill, *Between Consenting Adults*, 14 PHIL. & PUB. AFF. 252 (1985). For data on the sizable percentage of Americans who believe that torture is never or rarely ever justified, see Symposium, *U.S. Public Opinion on Torture, 2001–2009*, 43 POL. SCI. 437 (2010).

is one reason we follow the convention. But, nevertheless, the existence of the convention plays an essential role in *making it true* that one has reason to drive on the right. By contrast, the wrongness of torture does not seem to turn on conventions we have established. It follows directly from truths about the nature of persons from which flow basic moral rights.

Surprisingly, theoretical work on the general distinction between convention-dependent and convention-independent normativity is limited.[128] George Mavrodes' work in just war theory offers a helpful examination of convention-*dependent* norms, in particular.[129] Mavrodes argues that various ethical principles prescribing appropriate conduct in war that are assumed to be convention-independent are, in fact, more plausibly regarded as convention-dependent. His main example is the prohibition against harming enemy non-combatants, widely thought to be a norm whose reason-giving force stems directly from *a priori* facts about human dignity. Against the dignitarian view, Mavrodes argues that sometimes non-combatants are more responsible for the harms perpetrated by a state that justified going to war in the first place, making them apt targets of blame and punishment.[130] Moreover, if war is always unjust, then it is far from clear that immunizing non-combatants from harm is the best way of minimizing the injustices of

---

[128] What *is* often discussed is the way specific normative obligations presuppose practices or conventions. For example, there is an extensive literature on the nature of promissory obligation and the general practices of promisors & promisees. *See* T.M. SCANLON, WHAT WE OWE TO EACH OTHER, 295–327 (1998); Niko Kolodny & R. Jay Wallace, *Promises and Practices Revisited*, 31 PHIL. & PUB. AFF. 119 (2003); Seana Valentine Shiffrin, *Promising, Intimate Relationships, and Conventionalism*, 117 PHIL. REV. 481, 481–524 (2008). *See also* ARTHUR ISAK APPLBAUM, ETHICS FOR ADVERSARIES: THE MORALITY OF ROLES IN PUBLIC AND PROFESSIONAL LIFE (1999) (arguing that agents by entering professions like law can acquire unique moral license to violate familiar moral prohibitions). For helpful discussion on this point, thanks to Stephen Galoob.

[129] George I. Mavrodes, *Conventions and the Morality of War*, 4 PHIL. & PUB. AFF. 117 (1975).

[130] *Id*. at 120–23.

war. He suggests, ultimately, that the reasons to follow the principle—to the extent that it represents a genuine normative constraint on our conduct in war—may be wholly explained by the fact that there is a reasonably widespread convention in the international community of not harming non-combatants, a convention that we have reason to support because it has made war considerably less bad in various ways and its existence is better than having no such convention.[131] More generally, there can be reasons to comply with and support conventional practices that are, all things considered, less than morally ideal (in the sense that strictly following them does not always or even most often realize the morally best outcome) when having a widely-followed convention that does reasonably well at solving large scale moral problems is better than having none at all.[132]

The other important study of convention-based normativity is Mark Greenberg's recent work in legal philosophy.[133] Greenberg's work is less concerned with the contrast between convention-dependent and convention-independent normative truths, and more with identifying the circumstances that give rise to convention-*dependent* normative truths. Greenberg points out that legal institutions and practices can help constitute the moral obligations and duties we have. On his 'anti-positivist' theory, the law *just is* the subset of our moral obligations that depend on law-related social practices.[134] But we can separate Greenberg's theory of law from his helpful commentary on the relationship between law-related conventions and a certain class of normative truths.

---

[131] *Id.* at 124–30.
[132] *Id.* at 127.

[133] *See* Mark Greenberg, *The Moral Impact Theory of Law*, 123 Yale L.J. 1288 (2014).
[134] *Id.* On anti-positivism, see discussion *infra* note 149.

Greenberg provides several examples of cases where conventions bear on what one ought to do. Like Mavrodes, he points to circumstances that generate complex moral problems whose solution requires collective action.[135] In such circumstances, when some collective activity emerges that does reasonably well at achieving morally good outcomes, there can be reasons to support the practice. Greenberg's principal example of this phenomenon is "specific schemes for the public good," like tax laws:[136]

> Without a legal system, people will have general moral obligations to help others. But there will often be no moral obligation to give any particular amount of money to any particular scheme. For one thing, especially when it comes to problems of any complexity, many different possible schemes are likely to be beneficial, and the efforts of many people are needed for a scheme to make a difference.[137]

The tax system solves this problem, more or less effectively, by selecting one scheme directed at the public good. Once we have tax laws in place, our general moral obligation to help others becomes a more specific one so long as paying taxes represents an especially effective way of discharging the general duty to help others. Moreover, the reasons to support and participate in a scheme are especially strong when it is part of a broader social institution that we have reason to respect (e.g. a democratic system).

---

[135] Greenberg, *supra* note 129, at 1312.
[136] *Id.* at 1314.
[137] *Id.*

Greenberg describes such reasons for respecting social conventions as "reasons of democracy."[138]

Another way that conventions can end up influencing what one ought to do is through the logic of promising:

> By making promises and entering into agreements, people change their moral obligations. The fact of agreement has moral force. Even if what was agreed on is an arrangement that is seriously morally flawed—a different arrangement would have been much fairer, for example—the fact that the arrangement was agreed on may be sufficient to create a moral obligation.[139]

In this way, the existence of a convention, and one's explicit or implicit acceptance of it, can generate reasons for acting in accordance with the convention, reasons of a similar species as those that favor keeping one's promises having to do with respecting the expectations of others.[140] Market conventions plausibly have this kind of normative force for economic agents.

Now, consider convention-*independence*. It seems unlikely that the establishment of new conventions could explain or alter the significance of our most basic moral norms. Consider the question of whether a person morally deserves to be executed for his crime. It seems unlikely that an affirmative answer to this question turns in any way on how we have historically sentenced capital defendants. Even if there were a convention in place

[138] *Id.* at 1313.
[139] *Id.* at 1312–13.
[140] *See also* sources cited *supra* note 128.

of executing a particular class of defendants, this would not make it morally appropriate to visit such a grave harm on a person who did not independently deserve it. The appropriateness of ending someone's life, as a threshold matter, does not seem to turn on what we *do* or *have done* "around here." It turns, if anything does, on fundamental facts regarding what persons deserve in light of their actions and circumstances.[141]

The notion that there are norms that are relatively "inflexible" and derive their force not from what we say or do around here but from *a priori* facts about persons and the nature of certain harms is a familiar one.[142] Facts concerning fundamental rights of the sort emphasized by deontologists fall within this normative category. Here, for example, is Frances Kamm on the grounds of the right to free speech:

> The right to speak may simply be the only appropriate way to treat people *with minds of their own* and the *capacity to use means to express it* . . . . To say that any given person is not entitled to the strong right to free speech is . . . a way of saying that *certain crucial features of human nature are not sufficient to generate the right in anyone*.[143]

---

[141] The point is routinely emphasized in the case law. *See* sources cited *supra* note 72. Prominent theories of criminal law's justification declare proportionality to be an essential moral precondition on just punishment, and the principle finds approval in prevailing practice. *See, e.g.*, H. L. A. HART, PUNISHMENT AND RESPONSIBILITY 9 (1968); Richard S. Frase, *Excessive Prison Sentences, Punishment Goals, and the Eighth Amendment: "Proportionality" Relative to What?*, 89 MINN. L. REV. 571, 646 (2005) ("Limiting retributivism is a sound jurisprudential principle which enjoys widespread support, and the Supreme Court has used this principle to place constitutional limits on the imposition of capital punishment, fines and forfeitures, and punitive damages.").

[142] On the grounds of moral status and basic rights, see JEREMY BENTHAM, AN INTRODUCTION TO THE PRINCIPLES OF MORALS AND LEGISLATION (1789); ROBERT NOZICK, ANARCHY, STATE, AND UTOPIA (1974); WARREN QUINN, MORALITY AND ACTION (1993); Joel Feinberg, *The Nature and Value of Rights*, 4 J. VALUE INQUIRY 243 (1970).

[143] INTRICATE ETHICS 247 (2007).

As a rule of thumb, relatively fundamental moral truths concerning rights and basic obligations tend to be convention-independent. They tend to be convention-independent because they express truths about what persons are owed simply on account of being persons.[144] Even a consequentialist who thinks that all rights are subject to balancing and compromise can get behind the notion that some are especially weighty and basic in that their significance stems from fundamental facts about human experience.[145]

It might be helpful in what follows to refer to conventions *displacing* or *altering* pre-conventional moral norms, where this denotes the process by which the establishment of conventions generates reasons for following the convention as opposed to whatever pre-conventional moral rule that would have been authoritative absent the convention. Whereas conventions cannot displace or alter some basic moral principles and prohibitions, conventional establishment *can* shape, in combination with the fundamental moral truths, what ought to be done in circumstances where many competing values are at stake that are roughly on a par and trade-offs are inevitable. When we do not have simple deontological rights and prohibitions to resolve normative questions regarding what ought to be done, the role of conventions and reasons for following them tend to acquire greater significance.

---

[144] Crucially, it is not simply their *moral* status that determines their pre-conventionality. Recall Greenberg's example of the way tax schemes alter our general moral duty to help others. Instead, it is the subject matter and relative importance of the relevant moral truths that makes them pre-conventional.

[145] *See* James Dreier, *Structures of Normative Theories*, 76 THE MONIST 22 (1993); Philip Pettit, *The Consequentialist Can Recognise Rights*, 38 PHIL. Q. 42 (1988).

In sum, there is a genuine and structurally interesting difference between convention-dependent and convention-independent normative facts. The former, like the obligation to pay taxes, partly depend on conventional facts, including, sometimes the very social practices and conventions that are law-related (in that they are influenced by the actions of paradigmatic legal actors like judges and legislators). *Essentially* convention-dependent norms tend to be ones that wouldn't arise *but for* the establishment of conventions. These often (1) concern the distribution of those benefits and burdens that do not implicate matters of fundamental right; (2) solve moral problems that require large-scale collective action; (3) arise in contexts where a paramount concern is respecting the expectations of participants in a convention. Normative questions that do not turn primarily on conventions (law-related or otherwise) tend to implicate matters of fundamental right and wrong. Our answers to these reflect our pre-conventional sense of what we owe to each other.

Just as we can ask of a highly specific (or "token") normative question (e.g. did Susan, in light of all the empirical facts true in her case, behave unreasonably?), whether it is convention-dependent or -independent, we can ask of normative questions considered in general—or *types* of normative questions—how *likely* they are to be essentially convention-dependent or -independent in individual cases (e.g. questions concerning reasonable risk-taking). In other words, we can ask how likely a type of question is to be settled by basic moral principle of right and wrong. Bear in mind that judges apply the law/fact distinction to types of questions (like questions of reasonableness in torts) not to highly particular questions in an *ad hoc* case-specific way. So, the general question of likelihood is important for purposes of applying the law/fact

distinction. The likelihood that a given type of question will be convention-independent may in many cases be hard to determine.[146] But the potential for complexity should be no mark against the genuineness of the distinction between convention-dependent and convention-independent normativity.

B. The Conceptual and Pragmatic Reasons for Interpreting "Law" and "Fact" in Terms of the Normative Distinction

There are conceptual as well as pragmatic reasons for classifying normative questions that are more likely to be essentially convention-dependent as *legal* and those likely to be convention-*in*dependent as *factual*. Beginning with the conceptual reasons, despite considerable disagreement in legal philosophy, there is relative consensus that law is distinguished at least in part by its unique dependence on certain sorts of social practices and conventions.[147] Legal facts paradigmatically (if not always) depend on

---

[146] For instance, Mavrodes may be wrong about our reasons for obeying standard prohibitions against harming enemy non-combatants in war and it turns out that our reasons predominantly hinge on fundamental facts about human dignity, convention-based reasons may nevertheless form an important component of the totality of considerations in favor of complying with such a prohibition. *See, e.g.*, Robert K. Fullinwider, *War and Innocence*, 5 PHIL. & PUB. AFF. 90 (arguing that the prohibitions against harming non-combatants is not wholly convention-based). This issue will be taken up in a discussion of "mixed" questions of law and fact.

[147] Mark Greenberg, *How Facts Make Law*, 10 LEGAL THEORY 157, 157 (2004) ("Nearly all philosophers of law agree that . . . . ordinary empirical facts about the behavior and mental states of people such as legislators, judges, other government officials, and voters play a part in determining [law].").

The central task of analytic jurisprudence is to describe in the most fundamental terms what law is: the features in virtue of which a norm gets to be *legal* rather than something else. The hope is that work in this area of legal philosophy might help us distinguish "law" from "fact." Legal philosophers have had little to say about the law/fact distinction. This is somewhat surprising given that the common law rule provides a convenient test-case for philosophical theories of law. *See* JOHN AUSTIN, THE PROVINCE OF JURISPRUDENCE DETERMINED (1832); RONALD DWORKIN, LAW'S EMPIRE (1986); HART, *supra* note 141. For criticism of the notion that the legal concept has a distinctive essence, see Brian Leiter, *The Demarcation Problem in Jurisprudence: A New Case for Scepticism*, 31 OXFORD J. LEGAL STUD. 663 (2011).

what individuals such as legislatures, judges, elected officials, say, believe, do, or intend. I refer to the relevant law-determining activities of persons as law-related conventions.[148] To illustrate the point, if it is the law in a jurisdiction that first-degree murderers are imprisoned for life, this fact holds at least partly in virtue of such social practices and conventions as legislators having enacted, according to established procedures, a statute that prescribes life-imprisonment for murderers.

This essential connection between law and social conventions may not be the whole story regarding how legal facts are determined. Indeed, analytic jurisprudence has been embroiled in a famous disagreement over the remainder.[149] But it suffices for present purposes to find *some* characteristic feature of law that might be useful in contrasting legal from non-legal questions—namely, the dependence of legal questions on various social practices and conventions.

Convention-dependent normative questions are accordingly *law-like*. They essentially depend on conventions, law-related or otherwise: the practices of merchants,

---

[148] This is only slightly artificial usage, considering that "conventions" in the ordinary sense refers to things said and done as a matter of course (or customary human activity). If my usage of "convention" feels unnatural, the reader should feel free to substitute all instances of "convention" with "what people say and do." The distinction between normative questions that depend on "what people say or do" and those that are independent can do all the theoretical work that is required. *See* discussion *infra* Part II.B.

[149] The disagreement that has come to define the field is whether distinctly *moral* (or broadly normative) facts in addition to the social practice facts necessarily contribute to making the law what it is. Positivists believe that it is only social practice facts that are essential to law. Hart, for instance, famously thought that a rule's *being law* was wholly determined by the rule's being part of a broader hierarchy of rules habitually obeyed in the community, including "primary rules" that prescribe what individuals should do in various situations, and "secondary rules," which specify the circumstances under which a primary rule emerges. Hart, *supra* note 141, at 99. By contrast, anti-positivists think that moral facts, like the fact that it is *morally good* for a community to abide by the plain meaning of statutes, help determine the legal rules of a jurisdiction, along with the relevant social practices. *See* Dworkin, *supra* note 147, at 52, 87. The dispute between positivists and anti-positivists goes well beyond the Hart-Dworkin debate, and the battle lines continue to be drawn, in new and interesting ways, by contemporary positivists and anti-positivists. *See, e.g.*, SCOTT SHAPIRO, LEGALITY (2011); Greenberg, *supra* note 129.

regulators, parties to a case, judges, and so on. As in the case of paradigmatic legal questions, figuring out convention-dependent normative facts often requires looking to practices within legal institutions and what legal actors have done and intended to do (recall the example of market norms). This fact furnishes a sound conceptual reason to treat convention-dependent normative questions as questions of *law*.

The conceptual reason is buttressed by pragmatic considerations. It is often suggested that the law/fact distinction is primarily a device for allocating decision-making responsibilities between judge and jury based on their respective competencies. Judges are thought to be better suited than juries to decide questions of law but not questions of fact. The convention-dependence of law (in general) furnishes a straightforward general justification for the rule: it makes eminent sense for judges rather than the jury to decide questions of law if answering these questions requires figuring out such conventions as what judges and legislatures have decided and done, given that judges are trained and have expertise in interpreting judicial and legislative behavior. This justification extends to judges deciding essentially convention-dependent normative questions, even when the conventions aren't law-related, insofar as the skill that judges acquire at interpreting social practices is suitably general.

Secondly, conventions serve their coordinating function best when they are more widely known and well understood. Judges unlike jurors articulate *reasons* for their conclusions and this makes them especially well-situated to promulgate conventions in the course of deciding essentially convention-dependent normative questions. By contrast, even in cases of special verdicts, the jury gives little to no explanation for its ultimate rulings on issues. The jury's inability to give detailed accounts of its judgments

makes it less suited to articulate the facts concerning social conventions underlying its normative conclusions. Accordingly, judges are well-suited to decide convention-dependent normative questions.

Thirdly, as discussed earlier, there is often no determinate answer to convention-dependent normative questions when the conventions are ambiguous or inconclusive. If no social practices clearly indicate how widgets have been priced (perhaps because widgets are a recent innovation or because merchant practices are ambiguous), there is no determinately reasonable price within a broad range of prices. Judges (and indeed the law) have an important role to play in "settling" indeterminate convention-dependent normative questions, by simply *choosing* a convention—or, in other words, by convention-mongering in such cases. As quasi-legislative actors, judges have authority to make such choices.

By contrast, conceptual and pragmatic reasons militate *against* judges deciding convention-*in*dependent normative questions. It is much harder to justify the claim that judges might be better suited than juries to decide questions concerning matters of fundamental moral right—for instance, those concerning what persons are owed simply on account of being persons. Furthermore, in a pluralistic society characterized by significant moral disagreement on questions of fundamental right, it seems appropriate to have the jury (rather than a single judge) decide convention-independent moral questions if and when they come up at trial insofar as the jury consists of multiple persons drawn from a representative cross-section of the community. We no doubt tolerate deviations from this principle, as in the case of judicial interpretation of basic rights enshrined in the Constitution. But in such cases judges have *explicit* constitutional or legislative

authority to decide the basic moral question. When a question of basic morality arises at trial and the Constitution and plain statutory law are silent as to who decides the question—which is precisely when the common law rule applies—it seems intuitive to think that judges lack default authority to decide the question. At the very least, in such cases judicial authority to decide the question does not seem like it can be grounded in the question's nature.[150] So, in general, it seems a sound rule for a legal system to adopt that questions of basic moral right and wrong that come up at trial and that haven't been explicitly assigned to judge or jury by the legislature or the Constitution are presumptively questions of fact for the jury.

With the meta-normative distinction in place and reasons for thinking that judges should be tracking it in allocating decision-making responsibilities between judge and jury, the next step is to see whether judges really have been tracking this distinction given their treatment of normative questions under the law/fact rule. I take up this challenge in the next section.

III. Judicial Sensitivity to Conventional and Pre-Conventional Norms

The preceding section suggests a rationale for distinguishing normative questions under the law/fact distinction. Some normative questions are essentially convention-dependent in that addressing them requires figuring out conventions (often law-related ones, like what prior judges or legislators have done, for example) or settling on new

[150] This argument based on judicial professional authority and legitimacy is the crux of the normative rationale for understanding the distinction as I suggest. The common law's key insight reflected in judicial interpretation of the law/fact distinction is that as a *default* matter judges lack authority to decide basic questions of justice because these questions are not essentially questions of law.

conventions. This makes them importantly 'law-like' on standard accounts of the nature of law, and judges happen to be uniquely suited to resolving them. There are also normative questions that are convention-*in*dependent. These typically concern fundamental moral requirements and prohibitions (matters of justice or basic rights). Convention-independent normative questions can reasonably be regarded as "questions of fact"—they have little if anything to do with *law* as such.[151] The aim of this section is to show that this basic principle helps explain judicial practice concerning the classification of normative questions under the law/fact distinction.

A. Convention-independent Norms in Negligence Law and Sentencing

The question of reasonable conduct in negligence cases and other areas of tort law has historically been treated as factual under the common-law rule. Its status as a relatively convention-*independent* normative question finds support in corrective justice theories of tort law.[152] On the corrective justice approach, the concept of a moral wrong is central to tort law, especially in the case of intentional torts, like assault and battery, as

[151] There are echoes of this reasoning in the case law. *See, e.g.*, Bose Corp. v. Consumers Union, Inc., 466 U.S. 485, 501 n.16 (1984) (justifying *de novo* review in cases where a factual finding "clearly implies the application of standards of law" and where it "cannot escape broadly social judgments") (quoting Baumgartner v. United States, 322 U.S. 665, 670–71 (1944)). The present account fills in the details regarding when evaluative questions "imply the application of standards of law." They do when convention-dependent norms are implicated.

[152] *See, e.g.*, JULES COLEMAN, THE PRACTICE OF PRINCIPLE: IN DEFENSE OF A PRAGMATIST APPROACH TO LEGAL THEORY (2001); ERNEST J. WEINRIB, THE IDEA OF PRIVATE LAW (1995); Jules Coleman, *Corrective Justice and Wrongful Gain*, 11 J. LEGAL STUD., 420 (1982); John C.P. Goldberg & Benjamin C. Zipursky, *Torts as Wrongs*, 88 TEX. L. REV. 917 (2010); Keating, *supra* note 28, at 367; Ernest J. Weinrib, *The Special Morality of Tort Law*, 34 MCGILL L.J. 403 (1988). *See generally* PHILOSOPHICAL FOUNDATIONS OF TORT LAW (David G. Owen eds., 1995). Corrective justice theory is not the only approach to tort law that emphasizes the centrality of wrong or blameworthy action in torts. Keating's approach, for example, does away with the focus on remedial obligation. Imposing remedial obligations on tort-feasors is just a second-best way of ensuring that agents don't commit a certain class of wrongs in the first place.

well as negligence.[153] A standard negligence claim involves a plaintiff claiming redress for having been wronged by the defendant's failure to exercise the degree of care of a reasonable person. The wrongfulness of the defendant's conduct is explained not in morally neutral terms, such as the defendant's being the "least-cost avoider" of the harms caused, but in terms that imply moral culpability and blame.[154] The defendant exhibited *inadequate* regard for the interests of others.

The relevant aspect of this tradition in tort law scholarship is not just its emphasis on a distinctly moral concept of wrongfulness or unreasonableness at the heart of tort law, but precisely the conviction amongst proponents of the view that the question of reasonableness/unreasonableness often turns on fundamental facts regarding the rights of persons. As Gregory Keating writes, "tort norms articulate obligations to avoid harming people in various ways, and to respect their authority over their persons and their property in various ways. These wrongs are grounded . . . in *rights* people have *as persons*, such as the right to physical and psychological integrity."[155] Why does the wrongfulness of tortious conduct often turn on fundamental moral rights? Part of the explanation concerns the nature and significance of the interests that "reasonable care" in this context is meant to protect. The relevant moral norms of reasonable behavior

---

[153] *See, e.g.*, Keating, *supra* note 28, at 367; Coleman, *supra* note 152, at 9, 15, 36.

[154] *See, e.g.*, James A. Henderson Jr., *Why Negligence Dominates Tort*, 50 UCLA L. REV. 377 (2002) (noting that basing liability on wrongful conduct rather than on conduct that merely causes harm appeals to the shared intuitions of American judges.). *Cf.* GUIDO CALABRESI, THE COSTS OF ACCIDENTS 26 (1970) ("I take it as axiomatic that the principal function of accident law is to reduce the sum of the costs of accidents.").

[155] Keating, *supra* note 28, at 369. *See also id.* at 383 ("The facts that tort rights and obligations attach to persons simply *as* persons . . . and run from every person in the jurisdiction to every other person, need to be front and center in our thinking about the character and content of primary obligations."). Keating's view is wrong-based but does not emphasize remedial or corrective obligations.

"protect important boundaries against unauthorised [sic] crossings" and "our essential interests as persons."[156] These interests include one's sovereignty and power of discretion "over one's physical person or one's real property."[157] Joel Feinberg in his famous catalog of harms identifies certain interests of persons as critical including:

> [T]he interests in one's own physical health and vigor, the integrity and normal functioning of one's body, the absence of absorbing pain and suffering or grotesque disfigurement, minimal intellectual acuity, emotional stability, the absence of groundless anxieties and resentments, the capacity to engage normally in social intercourse and to enjoy and maintain friendships, at least minimal income.[158]

Negligence cases routinely involve the failure of agents to guard against setbacks to interests that fall in Feinberg's privileged set of critical interests—as when an employer's failure to implement a safety feature results in the amputation of an employee's limbs; and when they do not, as when the harms are purely economic and arise out of market exchange, courts tend to be reluctant to treat the case under tort law.[159]

---

[156] *Id.* at 390, 393.

[157] *Id.* at 390. *See also* W. PAGE KEETON, PROSSER AND KEETON ON THE LAW OF TORTS 8 (5th ed. 1984) ("The common thread woven into all torts is the idea of unreasonable interference with the interests of others."); Nicholas J. McBride, *Rights and the Basis of Tort Law, in* RIGHTS AND PRIVATE LAW 331 (Nolan & Andrew Robertson, eds., 2012).

[158] Feinberg, *supra* note 29, at 37.

[159] Under the "pure economic loss" rule, the law in most states is opposed to recovery under tort law in cases where the plaintiff's injuries are "purely economic" and there is no personal injury or damage to tangible property. *See* Herbert Bernstein, *Civil Liability for Pure Economic Loss Under American Tort Law*, 46 AM. J. COMP. L. 111, 112 (1998); Gergen, *supra* note 73, at 414

Another reason for thinking that the reasonableness enquiry in torts implicates matters of basic right, consistently with the corrective justice framework, is that a finding of negligence liability plausibly involves moral condemnation or *blame* of the defendant, and often results in punitive damages, as in cases of gross negligence. Basic principles of fairness militate against blaming or punishing agents unless they are truly morally blame and punishment-worthy: that is, only if they violate their moral obligations.

The corrective justice approach can be understood in terms of the framework developed earlier of convention-independent normativity. The tradition emphasizes that normative questions raised in tort law are often convention-independent; that is, they implicate 'pre-conventional' moral norms. The moral norms governing reasonable risk-taking in the sort of cases that frequently recur in the tort context—cases involving serious harms to agents—concern a form of regard we owe to others simply on account of their being persons whose interests matter. The relevant norms are not so easily displaced by contrary conventions or a general practice of doing less than what is morally required. Even if a community developed a habit of, say, driving drunk or recklessly on the road, for instance, this would not necessarily immunize a person from the charge of having behaved wrongfully or negligently when they cause substantial injury to another motorist on account of their recklessness. Given the gravity of the harms at stake, it seems sensible to assume that persons do not freely and intelligently consent to a relaxation of moral prohibitions and requirements in this domain, at least not without careful and near universal consideration of the merits of doing so. In other words, bad

("The narrow protection afforded economic interests in tort law signals the law's greater tolerance for selfishness when the interests affected are purely economic.").

driving habits that emerge organically, even when they generate expectations of reckless behavior on the road, cannot wholly undermine the moral obligation one has as a motorist simply to avoid being reckless. After all, people's lives are at stake. While sufficiently wide spread and deliberately chosen conventions of increased risk-taking—ones that have been legislatively enacted, for instance—could theoretically displace moral norms prohibiting negligence on the road, the fact that they would have to be widely shared and very deliberately chosen itself reflects the relative inflexibility of basic moral norms in the negligence context.

It is certainly true that the norms of the reasonable person are also meant to be ones that are generally obeyed in society (the norms, as it is often put, are those internalized by the "*ordinary*, prudent person").[160] However, it would be odd to think that it is the fact that most of us do not behave recklessly that grounds the *moral wrongfulness* of driving at high speeds or under the influence. Our practices have evidential value, indicating as they do the existence of basic norms of decency and reasonable behavior that most people follow as a matter of course. More importantly, our practices in the negligence context are not decisive. They must be tested against questions of basic rights and fairness.

Undoubtedly, there are cases in tort law where the harms in question seem less severe, falling short of violations of interests on Feinberg's list. Accordingly, it becomes more plausible that the norm violated evolved from and depends essentially on conventions. Arguably, part of what makes it morally unreasonable to fail to keep one's

[160] *See, e.g.*, N.Y. P.J.I. 2:10 (3d ed. 2000); Patrick J. Kelley & Laurel A. Wendt, *What Judges Tell Juries about Negligence: A Review of Pattern Jury Instructions*, 77 CHI-KENT L. REV. 587, 595 (2002) (describing the emphasis in jury instructions on "ordinary" care).

house in good repair when inviting people over is that people *expect* responsible community members to do the same, an expectation that may well be grounded in a general practice of repairing one's home rather than *a priori* moral facts about what invitees are owed.[161] If the general practice disappeared, perhaps some of the moral obligation to keep one's home in good repair would be undercut (and the onus would be on the invitee to take reasonable precautions or else assume the risk of harm). Conventions play an even bigger role, consistently with the corrective justice framework, in cases where the reasonable person standard is applied to professionals, like doctors and lawyers. In such cases, the standard is naturally articulated by reference to the practices and conventions amongst the relevant professional class regarding different sorts of risks.[162]

Even if conventions play some role in shaping our sense of what counts as morally reasonable risk-taking, this hardly settles whether the conventions *determine* the standard of care or instead exist precisely because members of the community have internalized fundamental norms requiring adequate concern for others.[163] Given the interests at stake—in the case of risk to invitees, their interest in physical safety—it is at least plausible that the conventions and general practice amongst home owners *reflect* or *operate in tandem* with the background pre-conventional norms governing interpersonal

---

[161] *See* RESTATEMENT (SECOND) OF TORTS § 343 (1977) (discussing premises liability); Patrick J. Kelley, *Who Decides? Community Safety Conventions at the Heart of Tort Liability*, 38 CLEV. ST. L. REV. 315, 324–25 (1990) (discussing the role of community safety norms in settling expectations).

[162] *See* Joseph H. King Jr., *In Search of a Standard of Care for the Medicial Profession: The "Accepted Practice" Formula*, 28 VAND. L. REV. 1213 (1975); 99 A.L.R.3d 1133 (1980).

[163] *See* discussion in Keating, *supra* note 152, at 369.

interactions rather than creating or co-opting those norms.[164] The view does not require being overly sanguine about the degree to which our community standards for risk-taking (or the standards internalized by professionals like doctors and lawyers) realize the best of all possible moral worlds; or that the ordinary person shows just the kind of care that we, independently of our conventions, morally owe to one another. The view depends only on its being plausible that our general conventions of protecting the critical interests of others, including their health and safety, get things at least roughly right from the pre-conventional moral point of view. Indeed, in many but not all cases our conventions are themselves the result of agents trying to do what they morally ought to. Our conventions need not be morally perfect for the point to stand that their evidential significance in the negligence context may go only so far as they line up with an independent and relatively basic moral structure of what we owe to each other.[165] Moreover, it is eminently plausible that our conventions of risk-taking need to be regularly tested against the basic moral question of what agents are owed as a matter of *right* given the harms at stake—in many cases, the fact that the defendant's actions were consistent what is conventionally done will not immunize her from liability.[166] To quote Learned Hand in the famous *TJ*

---

[164] *See, e.g.*, Kelley, *supra* note 161, at 324 (noting that community safety norms can develop through moral teaching).

[165] For a related view, see Benjamin Zipursky, *Sleight of Hand*, 48 WM. & L. REV. 1999 (2007). On Zipursky's view as I understand it, in interpreting negligence standards, jurors figure out obligations of reciprocity through their participation in (and commitment to) common conventions. The relevant obligations are not necessarily moral in nature. Conformity to them involves a kind of non-moral virtue. The point on which both Zipursky and I agree is that the question of what is reasonable in negligence is a *normative* question, and not identical to the question of what the conventions are. Our conventions are a means to understanding the relevant norms.

[166] The discussion to follow of various cases elaborates on this point.

Hooper case: "there are some precautions so imperative that even their universal disregard will not excuse their omission."[167]

Admittedly, the approach to negligence defended by corrective justice theorists and other theorists of tort law that focus on wrongful action is far from universally accepted. However, it suffices for present purposes that the view is plausible and widely-endorsed. For it is enough to *rationalize* (not fully vindicate) judicial behavior using our basic principle to motivate its explanatory potential. Corrective justice theory has been tremendously influential, historically, and continues to be widely embraced. Its proponents include Blackstone and influential nineteenth-century American jurists.[168]

So long as the question of reasonableness in negligence law is plausibly understood to implicate convention-independent norms, the case law appears driven by the claimed difference between legal and factual normative questions. Even if the norms applicable in the negligence context turn out to be far more convention-dependent than they appear, this would not make the historic practice of judges assigning the question to juries inexplicable. The structure of reasonableness norms in the context of dangerous risk-taking is far from obvious, and it is eminently reasonable to suppose that convention-independent moral principles concerning basic rights should have an important role to play in settling when injured persons should be entitled to demand redress and punishment when others inflict dangerous risks on them.

While a full-scale defense of corrective justice theory and related approaches to negligence law is not the point of this section, it is worth expanding on some of the ways

---

[167] 60 F.2d 737 (2nd Cir. N.Y. 1932). Thanks to Jens Ohlin for reminding me of Hand's discussion of the issue.

[168] *See* Goldberg & Zipursky, *supra* note 152, at 928 n.67 (citing jurists for the moral wrong based view of torts).

in which corrective justice theory can explain key aspects of negligence law that other theories struggle with. It will be helpful to compare the theory with its main rival amongst theorists: the law and economics approach.[169] The economic approach traditionally eschews moral categories in explaining tort law's normative concepts. What makes negligent conduct "unreasonable" is not that it violates some *a priori* or relatively basic moral norm. Instead, its unreasonableness is wholly explained by the fact that it inflicts a harm on the defendant that the negligent actor could have more cheaply avoided. Moreover, the imposition of liability is conceptualized not as punishment for immoral conduct but as akin to a licensing fee or tax imposed to incentivize efficient risk-taking: risk-taking that minimizes the monetary costs of accidents.[170] The resulting account characterizes tort law's ultimate aim as that of efficiently allocating costs in a way that leads to overall wealth maximization.[171] The economic approach may be very sensible in many domains of tort law—e.g., strict liability law and cases where the harms at stake fall outside of Feinberg's critical interests. But it struggles to explain the actual practice of juries empowered to decide the question of *negligence*.[172]

---

[169] *See, e.g.*, Calabresi, *supra* note 154; WILLIAM M. LANDES & RICHARD A. POSNER, THE ECONOMIC STRUCTURE OF TORT LAW 85-107 (1987); Henry T. Terry, *Negligence*, 29 HARV. L. REV. 40 (1915); *see also* United States v. Carroll Towing Co., 159 F.2d 169 (2d Cir. 1947) (articulating the famous Learned Hand formula for negligence). On the Hand formula, a person's conduct is unreasonable only if $PL > B$, where P is the probability of an injury occurring, L is the magnitude of the injury, and B is the expected benefit of engaging in the conduct. The economic approach appraises L and B in financial terms (often using willingness to pay as a proxy for value).

[170] *See, e.g.*, Calabresi, *supra* note 154.

[171] *See, e.g.*, Landes & Posner, *supra* note 169, at 154–57(defining values to be promoted in terms of willingness to pay and wealth maximization).

[172] Some theorists favor mixed theories of tort law for this reason. *See, e.g.*, Gary T. Schwartz, *Mixed Theories of Tort Law: Affirming Both Deterrence and Corrective Justice*, 75 TEX. L. REV. 1801 (1997).

In a systematic examination of state jury instructions, Patrick Kelley and Laurel Wendt find that jurors deciding the negligence issue are never told, through jury instructions or otherwise, to calculate and decide purely based on how costly it was to take relevant precautions.[173] Neither are they usually told to look at general conventions except in cases involving professional actors like doctors or lawyers. Kelley and Wendt write:

> To law professors, of course, unreasonable foreseeable risk conjures up Henry Taylor Terry's cost-benefit test for negligence, embodied in the first and second Restatements and summarized in Learned Hand's *Carroll Towing Company* test. But it seems to us that this is not the meaning that would be conveyed to the jury [of reasonable prudence]. . . . The instructions seem to call on the jury to determine whether the defendant's conduct, which resulted in harm to the plaintiff, was *a private injustice* to the plaintiff.[174]

The hypothesized reasonable persons, "though not paragons of virtue *simpliciter*, can be expected to act in reasonably careful or reasonably prudent ways. The only virtue this fully endows the hypothesized person with is the virtue of justice: the reasonably careful person exercising ordinary care under the circumstances gives the plaintiff what is her due."[175] While some states add various qualifications to the typical instructions (qualifications having to do with emergency situations, avoidable situations, and with the

[173] *Supra* note 59.
[174] *Id.* at 618–21 (emphasis added).
[175] *Id.* at 621.

inappropriateness of imposing an excessively demanding standard of an exceptionally reasonable person, for example), *none of them excuse the defendant if most people would have done the same or if the financial costs of taking precautions would have outweighed the benefits*.[176] State jury instructions routinely include the disclaimer: "the law does not say how the negligence standard applies, rather that it is for the jury to decide, based upon the facts in the case."[177] Moreover, recommendations made by scholars that normative language be excised from instructions on negligence so that they refer only to care taken by the "ordinary person"—rather than the ordinary, *reasonable* person—have been followed virtually nowhere.[178]

Jury instructions on negligence reveal more than just the difficulties confronting the economic theory of tort law (construed as a theory attempting to describe how the law *is* rather than how it ought to be). They reinforce the central claim made by corrective justice theorists—that jurors deciding the negligence issue are invited to tap into their basic moral sentiments, those "constitutive of the [ordinary] sense of justice itself."[179] While it is impossible to say with certainty how jurors *in fact* decide what counts as reasonable or unreasonable risk-taking, there is considerable evidence to suggest that jurors respond unfavorably to defendants who defend risk-taking by appeal to conventions or economic cost-benefit analysis.[180] The famous Ford Pinto case provides

[176] *Id.* at 603–07.
[177] *Id.* at 608 (citing Mich. 10.02 (2d ed. 1981 & Supp. 2001)).
[178] *Id.* at 613 (describing unsuccessful attempts to simplify instructions by excising 'reasonable').

[179] Keating, *supra* note 152, at 376.
[180] *See, e.g.*, Richard W. Wright, *The Standards of Care in Negligence Law*, *in* PHILOSOPHICAL FOUNDATIONS OF TORT LAW 249, 263 (David G. Owen ed., 1997) ("Defendants that are thought to have deliberately made such risk-utility decisions are often deemed by juries and judges not only to have been negligent, but also to have behaved so egregiously as to justify a hefty award of punitive damages . . . ."); Gary T. Schwartz, *Tort Law and the Economy in Nineteenth Century*

but one example of juror antipathy to economic and convention-based reasoning.[181] Ford was found to have sold its Pinto model car with a design element that made it especially prone to explosion upon impact from the rear, a feature that Ford knew would result in deaths and serious injuries amongst customers. Ford's engineers and management were especially explicit in their rationale for not implementing a safer design. They found that the cost of the safer design was outweighed by the amount Ford would have to pay in liability for deaths, pain, and suffering caused by exploding Pintos. Ford based its calculations on actuarial tables estimating the 'value' of a person's life.[182] Jurors found Ford guilty of negligence and imposed hefty punitive damages despite Ford's insistence that such cost-benefit calculations were routinely made in the industry. According to the jury, Ford displayed inadequate regard for people's lives and its decision to tolerate the risks inherent in selling the Pinto was not one that was Ford's to unilaterally make.[183] The Ford Pinto case is far from unusual in terms of juror behavior.[184] The actual practice of jurors lends further plausibility to the claim that the question of reasonableness in negligence law is widely seen to be likely to be a convention-independent moral question.

Moving on to the death penalty case, there is an even stronger argument to be made that normative questions that arise in capital sentencing (e.g., whether the defendant's conduct was sufficiently heinous to warrant the death penalty) are convention-*in*dependent (and hence factual). Whether and to what extent aggravating or

---

America: A Reinterpretation, 90 YALE L.J. 1717, 1757 (1981) (finding that New Hampshire and California courts are reluctant to find that economic costliness justified a defendant's risk-taking).
[181] *See* Grimshaw v. Ford Motor Co., 174 Cal. Rptr. 348 (1981); *see also* Gary T. Schwartz, *The Myth of the Ford Pinto Case*, 43 RUTGERS L. REV. 1013, 1034–38 (1991).
[182] Schwartz, *supra* note 175, at 1020.
[183] *Id.* at 1014.
[184] *See* sources cited, *supra* note 180.

mitigating factors are present in the defendant's case—that is, whether the facts militate for capital punishment—is almost universally decided by the jury.[185] The question is one of profound moral seriousness.[186] It amounts, ultimately, to the question whether the state is justified in taking a person's life, an affirmative answer to which depends on what we owe convicted criminals as a matter of right.[187] If criminal defendants who have suffered from severe mental handicaps have a moral right to the community's mercy, for example, conventions (law-related or otherwise) cannot displace their right to the community's mercy. The significance of the right depends, among other things, on the way mental handicaps impair a person's capacities for moral and rational action, and the gravity of the harm that is ending a person's life.[188] Even if we have previously executed defendants who have suffered from severe mental handicaps, our previous practices would not shake the fundamental moral prohibition—if there is one—on sentencing them to death.

---

[185] *See* discussion *supra* Part I.C.

[186] *See, e.g.*, Penry v. Lynaugh, 492 U.S. 302, 319 (1989) (noting that the death sentences must "reflect a reasoned *moral* response to the defendant's background, character, and crime"). Members of the Court have consistently recognized that the law demands moral deliberation from the sentencer, that the life-or-death decision is a question concerning the defendant's "moral entitlement to live," *Spaziano v. Florida*, 468 U.S. 447, 469 (1984), and that it must be based on a "moral inquiry into [his] culpability," *California v. Brown*, 479 U.S. 538, 545 (1987) (O'Connor, J., concurring); *see also* Kansas v. Marsh, 548 U.S. 163, 180 (2006) (noting that a jury is constitutionally tasked to engage in a "measured, normative process . . . when deciding the appropriate sentence for a capital defendant"); Blystone v. Pennsylvania, 494 U.S. 299, 317 (1990); Caldwell v. Mississippi, 472 U.S. 320, 329 (1985).

[187] *See* Gregg v. Georgia, 428 U.S. 153, 182 (1976) ("[T]he Eighth Amendment demands more than that a challenged punishment be acceptable to contemporary society. The Court also must ask whether it comports with the basic concept of human dignity at the core of the Amendment.").

[188] *See, e.g.*, Richard Lipke, *Social Deprivation as Tempting Fate*, 5 CRIM. L. PHIL. 277 (2011). The Supreme Court has essentially affirmed this style of moral reasoning, as having broad appeal, and not just among ethicists. *See* Penry, 492 U.S. at 319; *Williams v. Taylor*, 529 U.S. 362, 395 (2000).

The stakes are simply too high for normative evaluation of the aggravating or mitigating significance of the defendant's conduct and circumstance, at least one that *favors* the death penalty, to turn on convention-based reasoning.[189] The harm of undeserved execution is not outweighed by the value of following conventions in the death penalty context. It is no surprise, then, that the Supreme Court has explicitly prohibited the use of legal rules for discarding potentially mitigating evidence in juror instructions.[190] The frequently repeated mantra in such rulings of the importance of deciding issues on a "case-by-case" basis reflects the Court's recognition that legalistic reasoning cannot support a finding that the defendant deserves to die, not, at any rate, without allowing pre-conventional moral norms concerning what the defendant deserves in light of the facts of his case to constrain findings of death-eligibility.[191] Accordingly, the *factual* status of questions concerning the existence of aggravating factors in the death penalty context (and the question of death-eligibility more generally) is not just favored by courts in most states; it is constitutionally mandated.[192]

To sum up: questions concerning the reasonableness of the defendant's conduct in negligence cases and the question of death-eligibility in capital sentencing are

---

[189] Judges can, I argue, be involved in determining whether *mercy* is warranted. Mercy can be warranted, even for the most heinous offenders, by appeal to social practice.

[190] *See* Eddings v. Oklahoma, 455 U.S. 104 (1982). (holding that the sentencer cannot refuse to give the defendant's turbulent family history mitigating weight based on a legal test of criminal responsibility); *id.* at 113–14 ("Just as the State may not by statute preclude the [capital] sentencer from considering any mitigating factor, neither may the sentencer refuse to consider, *as a matter of law*, any relevant mitigating evidence."); *see also* McKinney v. Ryan, 813 F.3d 798 (9th Cir. 2015).

[191] *See, e.g.*, Blystone v. Pennsylvania, 494 U.S. 299, 317 (1990) ("[I]n the end it is the jury that must make the difficult, *individualized* judgment as to whether the defendant deserves the sentence of death." (quoting Turner v. Murray, 476 U.S. 28, 34 (1986))); Graham v. Collins, 506 U.S. 461, 515 (1993) (noting that a reasoned moral response requires individual assessment of personal culpability).

[192] Ring v. Arizona, 536 U.S. 584, 584 (2002).

convention-independent normative questions; answers to which do not turn primarily on conventions. Negligence law is involved in the punishment and remedy of wrongful risk-taking—the sort of risks that endanger the critical interests of others. The wrongfulness or unreasonableness of conduct that is familiarly the subject of a negligence action does not turn solely on what we do around here, but plausibly on the kind of respect that is fundamentally owed to others. The "question of fact" classification of such normative questions is precisely what our principle concerning the difference between legal and factual normative questions would predict. In capital sentencing, fundamental moral principles against undeserved execution *must* be taken into account when issuing death sentences, if a capital sentencing regime has any chance of being morally legitimate, a fact endorsed by the Supreme Court based on its interpretation of the Eighth Amendment's prohibition against "cruel and unusual" punishment.[193] Questions of mitigation and aggravation are therefore reasonably regarded as convention-independent normative questions or "questions of fact": in deciding them the jury invokes its pre-conventional sense of justice regarding what the defendant is owed.

B. The Convention-dependent Morality of Contract Law

Two independent lines of reasoning suggest that normative truths implicated in contracts disputes, having to do with the unconscionability of a contract or the reasonableness of implied terms, are in general determined essentially by conventional

[193] *See* sources cited *supra* note 191. *See also* Lockett v. Ohio, 438 U.S. 586, 604 (1978) (holding that the Eighth and Fourteenth Amendments require that "the sentencer . . . not be precluded from considering, as a *mitigating factor*, any aspect of a defendant's character or record and any of the circumstances of the offense").

facts. The first is connected to the essential role that non-legal conventions play in determining what counts as contractual reasonableness and/or unconscionability. Courts routinely look to the parties' prior course of dealings and the general practices of the merchant community when determining whether a price or other implied term would be reasonable.[194] Whereas in the case of dangerous risk-avoidance, our practices and the conduct of ordinary persons plausibly *reflect* or operate in tandem with (rather than constitute) pre-conventional norms of basic decency and regard, market conventions are not ultimately a function of the moral dispositions of merchants or reflections of basic moral rights.[195] "Conventional" prices are fixed based on parties pursuing their own self-interest under conditions of scarcity. Moreover, the going rates for goods sold in the marketplace or prices historically accepted by buyers and sellers seem to weigh significantly on the reasonableness of set prices, given that parties can expect to be

---

[194] *See, e.g.*, U.C.C. § 1-303 (2011); Western Indus., Inc. v. Newcor Canada Ltd., 739 F.2d 1198, 1204 (7th Cir. 1984) (trade custom can even override UCC gap-filler defaults and can give rise to binding contract provision); American Mach. & Tool Co. v. Strite-Anderson Mfg. Co., 353 N.W.2d 592, 596-98 (Minn. Ct. App. 1984) (evidence of trade usage and course of dealing admissible for determining delivery dates). *See generally* Robert A. Hillman, *Court Adjustment of Long-Term Contracts: An Analysis under Modern Contract Law*, 1987 DUKE L.J. 1 (1987); Randy E. Barnett, *The Sound of Silence: Default Rules and Contractual Consent*, 78 VA. L. REV. 821 (1992).

[195] Compare the norm-determining role of conventions in contracts from their evidential role in negligence. Plausibly, the moral norm prescribing concern for the safety of others when driving *gives rise to* certain conventions on the road. These conventions then bear on the relevant standard of care. By contrast, the conventions amongst merchants that determine what a reasonable price is for widgets are not similarly the result of agents conforming to moral norms. The conventions play a true norm-determining role only in the latter case.

Moral norm	Convention	Standard of care
Show concern for the safety of others when driving!	Motorists showing concern for the safety of others	The legally relevant standard of reasonableness
null	Merchants pricing widgets at $x	The legally relevant standard of reasonableness

charged such rates in the absence of explicit agreement.[196] So long as one of the central values to be promoted through contract law is the expectations of parties, as is often suggested, then the relevant market-based conventional facts can truly be said to determine in a non-derivative way the reasonableness of implied terms and contractual fairness.[197]

Given the dependence of contract law on non-legal conventions, it is not hard to see why what counts as contractual reasonableness and fairness might also depend on law-related conventions—such as judicial practice or regulation. In many instances, the appropriate price or some other missing term will not be settled by the prior contractual history of the parties or even industry-wide practice; therefore, judges must make *some* choice of default rule from among various acceptable options.[198] Judges are especially well-suited to make such choices and the relevant default rules known, as quasi-legislative officials who can be held accountable, given their responsibility for making the rationale behind their judgments explicit. Any party able and empowered to make the non-legal conventions and default rules known will inevitably end up *influencing* non-legal conventions as well, given that merchants often turn to case law for information about the merchant community's practices. Even a court's initially mistaken

---

[196] *See, e.g.*, C. A. Riley, *Designing Default Rules in Contract Law: Consent, Conventionalism, and Efficiency*, 20 OXFORD J. LEGAL STUD. 367 (2000) (arguing that gap-filling ought to be driven by the subjective consent of the parties and the customs and conventions immanent within the parties' community); *id.* at 374–82 (suggesting that the case for conventionalist defaults include considerations having to do with cost, fairness, tacit consent, fairness, reasonably reliance, and positive incentives); Richard Craswell, *Contract Law, Default Rules, and the Philosophy of Promising*, 88 MICH. L. REV. 489 (1989). It is implausible that there are *a priori* moral facts regarding reasonable pricing for most non-essential goods and services.

[197] *See, e.g.*, Lord Steyn, *Contract Law: Fulfilling the Reasonable Expectations of Honest Men*, 113 LAW Q. REV 433 (1997); Riley, *supra* note 196, at 375–80.

[198] *See generally*, Ian Ayres & Robert Gertner, *Strategic Contractual Inefficiency and the Optimal Choice of Legal Rules*, 101 YALE L.J. 729 (1992). The U.C.C. gives courts broad gap-filling powers. U.C.C. § 2-204(3) (2011).

interpretation of industry custom may become correct over time because the industry conforms to the interpretation. In other words, what judges *say* and *decide* has a role to play in determining what counts as reasonable in contractual exchange. It makes sense, therefore, to regard the question of reasonableness of contracts as an essentially convention-dependent normative matter.

It is worth emphasizing why this sort of analysis does not work in the case of negligence. Law-related conventions—legislative practice, for instance—obviously do influence what counts as negligent conduct. A defendant's failure to comply with statutory requirements, if it results in harms that the statute aims to prevent, is considered negligence *per se,* or negligence as a matter of law.[199] However, such rules drawing on legal conventions are treated as exceptions to the general practice of juries deciding what counts as negligent conduct. Moreover, the reason statutory requirements are often introduced in negligence cases is because they are plausibly *derivative* of more basic, pre-conventional moral norms that parties are supposed to respect. When a person injures another while driving drunk, what makes such conduct wrongful and unreasonable is the degree of disregard the person shows for the safety of others. Its wrongfulness is not principally explained by the fact that the law forbids drunk driving or the driver's failure to comply with the law. To reiterate the point made earlier, the law-related conventions in the tort context *reflect* important pre-conventional moral facts, grounded in the dignity and importance of other people and their fundamental interests. On this popular conception of tort law, the actor deciding the negligence issue typically does not need to know the law-related conventions to decide whether the conduct at issue is wrongful

[199] RESTATEMENT (THIRD) OF TORTS: PHYS. & EMOT. HARM §14 (2010).

enough to warrant compensation and punishment.[200] In short, the way in which contractual norms are convention-dependent is different in kind from the way legal and non-legal practices bear on the negligence question. In the contracts case, it does not make sense for the decision-maker to decide what terms are reasonable independently of conventions or by reference to their intuitive sense of what is owed to others as a matter of basic decency. This is because the interests under threat in contracts cases are not ones that we must, as a matter of basic decency, safeguard out of respect for our contractual partners.

The second line of reasoning that supports the unique convention-dependence of contractual norms is more involved, and it may be better to layout its basic structure before defending its key premises. The main idea is that the interests of individuals safeguarded by contractual norms (like reasonableness and unconscionability) are less morally vital than those protected by tort and criminal law norms. Because of these lower stakes in the economic context, pre-conventional moral norms governing contractual exchange are comparatively easily displaced by conventions requiring less than what pre-conventional morality requires (always keeping one's promises, for example), especially if such conventions can be justified by appeal to overall social good. Moreover, such conventional displacement of ordinary morality *has* plausibly occurred in the modern marketplace, through a combination of laws, economic policy, and cultural attitudes. By contrast, it is much less plausible that fundamental moral norms prohibiting the physical endangerment of others are flexible or have shifted in the face of contrary conventions.

---

[200] *See* discussion *infra* Part III.A, note 180.

Beginning with the lowered stakes, the prototypical contractual dispute over the reasonableness of implied terms or unconscionability implicates interests of parties that are primarily economic: losing a bargained-for benefit or losing more than one bargained-for.[201] As Mark Gergen notes, there is a long history in the common law of treating such economic losses arising out of voluntary transactions as having a lesser moral significance than harms to persons and property resulting from agents' tortious or criminal conduct.[202] There are several reasons for this disparity. For one, the loss of an economic benefit is at best instrumentally bad for an agent, whereas physical injury, emotional trauma, and interference in certain privileged domains (like a person's home) have an inherent or intrinsic badness.[203] Generally, an interest in a particular distribution of financial benefits and burdens stemming from voluntary trade between parties who are reasonably well off is not a fundamental or especially urgent interest of persons. It is no doubt true that morally important interests of persons are implicated in disputes over price terms or contractual unconscionability. But the mere fact that morally significant interest of persons are at stake does not entail that these interests generate rights-claims or implicate foundational moral values. The parties to such disputes do not have a claim as a matter of right to have their interests valued over all else. These local interests of parties to a dispute can be sacrificed more easily in the name of conventions and general social good, than say an interest in physical integrity.

---

[201] Not all contracts disputes implicate non-urgent economic interests of persons. Whereas a dispute over what counts as a reasonable price concerns the economic interests of the contracting parties, a dispute over whether there was fraud or coercion implicates a more fundamental interest: an interest in non-domination. Appropriately, questions of fraud and coercion are treated as questions of fact! Moreover, unconscionability is analyzed differently from coercion. For helpful discussion on this point, I thank Scott Altman.

[202] Gergen, *supra* note 73, at 412–13.

[203] *See* Feinberg, *supra* note 29, at 31.

Economic harms resulting from unconscionable contracts or contracts with unreasonable terms can certainly be substantial, as when the losses are sustained by impoverished agents like poor tenants. If a mortgagor is about to lose her home despite having paid off a good portion of her debt due to a uniquely burdensome mortgage, the potential harm to her, absent judicial intervention under unconscionability doctrine, is surely great.[204] Nevertheless, it is true that the losses sustained in the course of doing business with other self-interested economic actors *typically* lack the gravity of unconsented-to physical injury or damage to real property sustained at the hands of reckless, malicious, or negligent agents. This is a recurrent theme in the corrective justice literature on torts as well as in accounts of the differences between contracts and other areas of law, and the supposition is borne out by existing case law.[205]

Another reason for the differential importance of harms inflicted in paradigmatic tort and criminal cases compared to those inflicted in the relevant category of contractual suits is that the former are caused by morally blameworthy or wrongful agents. By contrast, it is far from clear that economic agents, even when they behave in exploitative and highly self-serving behavior, are to blame for the harms they cause (not including conduct by such actors deemed tortious or criminal). Harms resulting from morally blameworthy conduct warrant greater concern than harms caused by blameless agents, and thus generate stronger reasons for state-authorized corrective and *punitive* action.[206] Part of the explanation for these differences in moral responsibility concerns our unique

[204] *See, e.g.*, Skendzel v. Marshall, 301 N.E.2d 641 (Ind. 1973).
[205] *See* Gergen, *supra* note 73.

[206] *See generally* Seana V. Shiffrin, *Harm and its Moral Significance*, 18 LEGAL THEORY 357 (2012).

conventions and cultural expectations for economic actors—issues to be discussed later in this subsection. The imposition of duties and obligations in contracts disputes does not generally carry implications of blame or punishment.

Because the harms to individuals are less severe, it makes sense that any pre-conventional moral duties or prohibitions that we might have owed to one another in contractual exchange will be relatively flexible in the face of a convention or general practice of doing *less* than what pre-conventional moral norms require, especially when the existence of such conventions results in overall societal good. Certain foundational moral norms (the prohibition against torture, for instance) have a rigidity that, in the extreme, gives rise to inviolable rights—i.e., rights that generally cannot by justifiably compromised based on considerations of overall, societal good. The norms governing reasonableness and fairness in contractual terms are unlikely to be rigid in this way in light of the harms at stake.

Moreover, it is quite likely that the standards for appropriate and inappropriate behavior in the market context *have* shifted in light of our cultural practice of tolerating and even encouraging unbridled self-interest in the marketplace.[207] Capitalistic societies

[207] *See, e.g.*, JOHN K. GALBRAITH, A HISTORY OF ECONOMICS: THE PAST AS THE PRESENT 37 (1987) (noting that the emergence of "[m]ercantilism involved . . . a marked break with the ethical attitudes and instructions of Aristotle and of Saint Thomas Aquinas and the Middle Ages in general"); ROBERT L. HEILBRONER, THE NATURE AND LOGIC OF CAPITALISM (1985) (comparing norms of pre-capitalist and capitalist society); ALAN WOLFE, WHOSE KEEPER?: SOCIAL SCIENCE & MORAL OBLIGATION (1989) (arguing that modern economic systems have supplanted traditional moral norms); Harry J. Glasbeek, *Commercial Morality Through Capitalist Law: Limited Possibilities*, 27 LE REVUE JURIDIQUE THEMIS 263, 303 (1993) ("The ideological power, and legal acceptance, of the primacy of selfishness and the merits of a utilitarian approach to public policy-making ensures weak restraints, that is, weak regulation."); *id.* at 306 (noting the "advent of capitalist relations was truly revolutionary" with respect to the norms of market exchange); Lawrence E. Mitchell, *Groundwork of the Metaphysics of Corporate Law*, 50 WASH. & LEE L. REV. 1477, 1484–85 (1993); Christopher Calvo, *Responsibility, Ethics and American Economic Thought, 1776-1900.* History of Corporate Responsibility Project Working Paper No. #7 (2011) (noting the impact of the Jacksonian era market revolution on economic culture). *Cf.*

encourage the relentless pursuit of material gain out of a sense that it promotes the general good.[208] The dominant view amongst economic historians is that law and culture have strongly shaped participants' sense of what is and is not appropriate in the market domain, and so long as we can trust this collective sense of the unique "morals of the marketplace," it favors a convention-based account of the norms at play.[209]

Conventional-displacement or alteration of the moral landscape in the contractual arena sets up a kind of feedback loop, where the existence of fairly wide-spread conventions of doing less than what is pre-conventionally thought to be decent behavior affects the nature and gravity of the harms when parties fail to act decently. Once the marketplace becomes a domain in which it is well-known that ordinary moral norms of decency and kindness are relaxed, parties incur an obligation to recognize and guard against the risk of exploitation when they voluntarily participate in market exchange. Relatedly, participants can be said to impliedly consent to the risk of being harmed by the self-interested and even highly predatory behavior of others. So long as persons engaged in economic exchange should anticipate and guard against the predatory behavior of others, the harms suffered by parties due to unconscionable contracts have a diminished significance. Consider again the mortgagor who is tied to a possibly unconscionable contract that requires foreclosure despite considerable payments made towards her home.

St. Thomas Aquinas, *Summa Theologica*, *in* EARLY ECONOMIC THOUGHT 60 (Arthur Eli Monroe, ed., 1945) (arguing that it is impossible for buyers and sellers not to fall into sin); JAMES DAVIS, MEDIEVAL MARKET MORALITY: LIFE, LAW AND ETHICS IN THE ENGLISH MARKETPLACE, 1200-1500 (2012).

[208] ADAM SMITH, THE WEALTH OF NATIONS 508 (1937) ("The natural effort of every individual to better his own condition . . . is so powerful a principle, that it is alone, and without any assistance, not only capable of carrying on the society to wealth and prosperity. . . .")

[209] *See, e.g.*, Meinhard v. Salmon, 164 N.E. 545 (N.Y. 1928); Lee v. LPP Mortgage. Ltd., 74 P.3d 152, 162 (Wyo. 2003) ("We have said that the relationship between a lender and its customer is contractual in nature so we impose no duties higher than the morals of the marketplace.").

In one sense, a court's willingness to bail out the mortgagor by finding the contract procedurally or substantively unconscionable involves a willingness to help out potentially irresponsible risk-takers. One need not be a shill for the economically powerful to see that, as far as empathetic victims go, those who suffer in the market place at the hands of genuinely selfish and predatory actors are less sympathetic victims than those who have been subjected to another's unexpected negligence or malice. Outside the marketplace, it is implausible to assume that we have collectively consented to a system of social organization where parties can inflict dangerous risks to physical and mental health on one another.

This overarching theory of contractual norms delivers a prediction: if it is true that judges are in the business of enforcing convention-dependent norms rather than pre-conventional morality, we should expect to find that judicial interpretation of concepts like unconscionability might fail to jibe with people's ordinary sense of interpersonal morality. And that is indeed what we find. There is a large quantity of scholarship pointing out the wide gap between ordinary notions of what is morally unconscionable and what judges deem unconscionable.[210] Some critics take this tension to reveal a failing on the part of courts to apply the law of contractual unconscionability correctly.[211]

---

[210] *See* RADIN, *supra* note 82; Glasbeek, *supra* note 207, at 292; Leff, *supra* note 78; Daniel T. Ostas, *Predicting Unconscionability Decisions: An Economic Model and an Empirical Test,* 29 AM. BUS. L.J. 535, 583 (1992) ("Unconscionability decisions mix radical rhetoric with conservative actions."). Notably, when parties to a contract are merchants in a commercial setting, courts invoke a "presumption of conscionability." *See e.g.*, Xerox Corp. v. Graphic Mgmt. Servs. Inc., 959 F.Supp.2d 311, 320 (W.D.N.Y. 2013); Cayuga Harvester, Inc. v. Allis-Chalmers Corp., 465 N.Y.S.2d 606, 617 (N.Y.S.2d 1983) ("[C]ontract[s] in a commercial setting [] [have] a presumption of conscionability. . . .").

[211] *See* NANCY S. KIM, WRAP CONTRACTS: FOUNDATIONS & RAMIFICATIONS 88, 207–09 (2013) (critiquing courts' failure to rein in unfair contracts); Browne & Biksacky, *supra* note 84, at 250 ("[I]n cases where factors suggest unconscionability, judges still rule against unconscionability and implicitly evoke Adam Smith's laissez-faire statement: 'Every man, [so] long as he does not violate the laws of justice, is left perfectly free to pursue his own interest [in] his own way. . . .'")

But there is a way of making sense of what judges are doing that both acknowledges the authentic moral intuitions of these critics while also charitably interpreting the holdings. Judges are not in the business of enforcing pre-conventional moral standards of fairness and justice because, plausibly, these do not apply in a context where exploitative behavior is tolerated for the general good. Courts can be viewed as enforcing less familiar and less demanding standards of interpersonal fairness. The routine emphasis one finds in judicial opinions on the distinctive "morality of the marketplace" supports this view.[212] There are other, more general features of contract law that reinforce the conventional view of contractual norms. There is, for instance, the widely-discussed tolerance in contract law of efficient breach.[213] Parties to a contract can refuse to perform their contractual obligations, in which case they are only obliged to pay expectation damages, rather than being compelled to perform. This strikes many as evidence of the law's tolerance for promise-breaking.[214] As a number of writers have argued, however, ordinary norms of promise-keeping plausibly do not apply in the contractual context.[215]

(alteration in original); *id.* at 250–54 (discussing cases where the authors believe courts should have found contracts unconscionable); Melissa T. Lonegrass, *Finding Room for Fairness in Formalism--The Sliding Scale Approach to Unconscionability*, 44 LOY. U. CHI. L.J. 1, 4 (2012); Price, *supra* note 3, at 744 ("Examination of the case law decided under section 2-302 will demonstrate that the matter-of-law mandate has resulted in badly-reasoned decisions on the issue of unconscionability. . . .").

[212] *See, e.g.*, Meinhard v. Salmon, 164 N.E. 545 (N.Y. 1928); Lee, 74 P.3d at 162*; see also* Browne & Biksacky, *supra* note 84.

[213] *See* RESTATEMENT (SECOND) OF CONTRACTS ch. 16, intro. note (AM. LAW INST. 1981); Robert L. Birmingham, *Breach of Contract, Damage Measures, and Economic Efficiency*, 24 RUTGERS L. REV. 273 (1970).

[214] *See, e.g.*, Melvin A. Eisenberg, *Actual and Virtual Specific Performance, the Theory of Efficient Breach, and the Indifference Principle in Contract Law*, 93 CAL. L. REV. 975, 1012–13 (2005); Oliver Wendell Holmes, *The Path of the Law*, 10 HARV. L. REV. 457, 462 (1897).

[215] *See, e.g.*, Daniel Markovits & Alan Schwartz, *The Myth of Efficient Breach: New Defenses of the Expectation Interest*, 97 VA. L. REV. 1939, 1982 (2011) (rejecting the notion that efficient breach cannot be squared with promissory morality); Steven Shavell, *Is Breach of Contract Immoral?*, 56 EMORY L.J. 439 (2006).

Another telling feature of contract law is the absence of punitive damages and the reluctance of judges to express moral condemnation of actors, except in contexts where a party's conduct is tortious or criminal (as in the case of fraud).[216]

There is ample basis, then, for the common law to treat normative questions in contract law as convention-dependent and, hence, as questions of *law* to be decided by judges. The grounds include (i) the lesser prototypical harms at stake, (ii) the essential role of non-legal conventions and judicial practice in defining what counts as "reasonable" economic behavior, and (iii) the wide-spread, legal and cultural promotion of uniquely self-interested behavior in economic exchange. At the very least, these factors render plausible an explanation in terms of our overarching framework of why the common law treats normative questions arising in the torts and criminal context differently from those pertaining to contractual reasonableness and unconscionability. Moreover, the analysis suggests a general recipe for determining whether a type of normative question is one of law or of fact. Courts might begin by examining what typically turns on the normative question—in particular, the gravity of harms that the relevant norms aim to prevent. If the harms are less serious, courts should consider whether conventions, including law-related conventions, might have altered the normative landscape making our pre-conventional, ordinary sense of fairness and justice a poor guide to answering the normative question.

The framework provides a helpful lens from which to view other developments. Recall the decision in *Cooper Industries*, where the Supreme Court held that the

[216] *See* RESTATEMENT (SECOND) OF CONTRACTS § 355 (1981); Laurence P. Simpson, *Punitive Damages for Breach of Contract*, 20 OHIO ST. L.J. 284 (1959) (arguing that punishment has no place in commercial transactions).

appropriateness of punitive damages in a case involving unfair competition/false advertising could be reviewed *de novo* as a finding intermediate between law and fact. The Court emphasized the moral condemnation involved in punitive damages awards.[217] The dissent was keen to emphasize that moral findings are routinely treated as factual and reviewed deferentially as in negligence law.[218] What *Cooper Industries* left unsaid is that the moral question of appropriate punishment in the context of essentially economic conduct like unfair competition can only be answered by considering the unique morals of the marketplace, which, as discussed, are shaped by conventions. This makes the moral question of appropriate punishment law-like. In other words, squaring *Cooper Industries* with the wider case law on normative questions and the law/fact distinction involves deploying the framework of convention-dependent and convention-independent normativity

C. Explaining Intertemporal Shifts: The Cases of Obscenity, Malice, and Aggravation

An aspect of the puzzle concerning normative questions was the inconsistent treatment of such questions under the law/fact distinction over time. A theory of the distinction needs to explain such changes. If the shifts are to be explained in terms of judicial error, we need a plausible account of why judges might have gotten things

[217] Cooper Indus. Inc. v. Leatherman Tool Grp., Inc., 532 U.S. 424, 432 (2001).
[218] *Id.* at 446 ("But there can be no question that a jury's verdict on punitive damages is fundamentally dependent on determinations we characterize as factfindings—*e.g.*, . . . whether the defendant behaved negligently, recklessly, or maliciously.") (Ginsburg, J., dissenting).

wrong. Alternatively, what is needed is an account of the changes in the world on which the factual or legal status of these questions has turned.

The present framework sheds important light on these changes, beginning with the case of a factual question that comes to be regarded as legal, or in-between legal and factual. The question of whether a false statement was made with "malicious intent" or "reckless disregard for the truth" was historically a question of fact for the jury, in both criminal and civil defamation cases under English as well American common law.[219] After the decision in *New York Times v. Sullivan*—requiring a finding of actual malice before punitive damages can be awarded to public official claimants—appellate courts began reviewing the issue *de novo*. In *Bose Corp. v. Consumers Union of United States, Inc*., the Supreme Court explicitly affirmed the widespread use of *de novo* appellate review of the "actual malice" question, noting "'the vexing nature' of [the law/fact] distinction . . . ."[220] The Court ruled that the question of whether a false statement was made maliciously was intermediate between law and fact—"a constitutional fact"—given the important First Amendment value at stake (a finding of actual malice licenses punishment of the speech in question).[221] Multiple scholars have since pointed out that it is hard to find a principled basis for the Court's treatment of the issue, given its disinclination to treat as constitutional facts other traditionally factual questions on which important constitutional rights hang.[222]

---

[219] *See* Bose Corp. v. Consumers Union of United States, Inc., 466 U.S. 485, 517 ("In my view the problem results from the Court's attempt to treat what is here, and in other contexts always has been, a pure question of fact, as something more than a fact—a so-called 'constitutional fact.'") (Rehnquist, J., dissenting); discussion *supra* Part I.C.
[220] Bose, 466 U.S. at 501–02.

[221] *Id.* at 517–20 (Rehnquist J., dissenting).
[222] *See* sources cited *supra* note 110.

On the present reading, the change in classification of the normative question—was the intent to publish a false statement sufficiently malicious and/or reckless to license punishment?—can be understood in terms of a cultural shift in our attitudes towards reputational harm (the kind of injury that defamation law seeks to prevent). As Robert Post writes, "defamation law presupposes an image of how people are tied together, or should be tied together, in a social setting. As this image varies, *so will the nature of the reputation that the law of defamation seeks to protect*."[223] Post goes on to observe that American society has grown increasingly skeptical of the notion, popular in what he calls "deference societies," that harm to a person's reputation or honor is a matter of grave moral importance.[224] That is, reputational harm used to be conceived in terms of personal dignity and fundamental right. This dignitarian view of reputational harm—with its presuppositions regarding fundamental morality—was supplanted with an economic view, on which a person is harmed by attacks on their reputation only insofar as it prevents them from acquiring property. "Contemporary Americans are uneasy with the concept of honor" but "they are intimately comfortable with the concept of property."[225]

Why would a cultural shift from treating reputational harm as an affront to a person's dignity to treating it as a mere economic injury help explain the changed law/fact status of the "actual malice" question? When reputational harm was viewed as very serious, it made sense for the law to draw on convention-independent moral

---

[223] Robert C. Post, *The Social Foundations of Defamation Law: Reputation and the Constitution*, 74 CAL. L. REV. 691, 693 (1986) (emphasis added).
[224] *Id.* at 702.
[225] *Id.* at 726. In the early 20th century, the moral harm of defamation weighed more heavily than First Amendment values. *See, e.g.*, Chaplinsky v. New Hampshire, 315 U.S. 568, 572 (1942) (exempting defamatory speech from constitutional protection given the "social interest in order and morality").

concepts in specifying normative conditions on punishment. The question of malice was treated as *factual* (convention-independent) in light of the presumed relevance of pre-conventional moral concepts of malice and reckless disregard to the question of what ought to be punished and how much in the context of defamation. However, as the harms from malicious defamation (in the pre-conventional moral sense) come to be viewed as less serious, the relevance of pre-conventional moral norms to the question of what defamation law should punish becomes less obvious. The harm from malicious defamation begins to look increasingly susceptible to trade-off and balancing against benefits incurred from the law's tolerating such harms to promote free speech. Accordingly, courts reasonably responded by supplanting a convention-independent normative concept of "actual malice" with a convention-dependent one (defined in part by our law-related conventions including judicial interpretation of the concept).

What was once perceived as a convention-independent normative question comes to be seen as legal convention-dependent precisely *because* of our evolving understanding of the relevant moral facts.[226] A similar account can be given of other areas of First Amendment law where courts have employed the "constitutional fact" doctrine to treat a normative question historically regarded as factual more like a legal question. In the case of obscenity law, the questions of whether a publication is "obscene" or "patently offensive" have come to be treated as constitutional questions requiring *de novo* review despite having been historically treated as paradigmatic

[226] Unsurprisingly, we find the Court minimizing the significance of reputational harm in its rulings and continuing to interpret the "actual malice" condition on punishment in increasingly demanding ways. *See, e.g.*, Air Wisconsin Airlines Corp. v. Hoeper, 134 S. Ct. 852, 861 (2014) ("[W]e have required more than mere falsity to establish actual malice: The falsity must be 'material.'") (citations omitted); Post, *supra* note 223, at 736–38 (citing cases of judicial discomfort with the notion of reputational harm as harm to dignity).

questions of fact for the jury that were reviewed deferentially.[227] Similarly, the question of whether provocative speech is so "inherently inflammatory" as to count as "fighting words" which are likely to provoke the average person to retaliation, receives *de novo* review.

It is tempting to think that this entire area of law has been influenced by evolving societal attitudes towards the relevant sorts of harm.[228] In a society that takes harms caused by obscenity very seriously, the law will take its cue from pre-conventional morality in deciding what kind of speech to prohibit or protect. However, once the harms come to be seen as less serious, it no longer makes sense for the law to closely track pre-conventional moral concepts of obscenity. Accordingly, legal actors respond by supplanting convention-independent moral concepts in First Amendment law, those based on the ordinary sense of right and wrong or basic conceptions of human dignity, with concepts that are at least in part shaped by convention (what we *find* or judge to be obscene). Note that on such a view the law's treatment over time of normative questions under the law/fact distinction far from being unprincipled is responsive to changes in our understanding of the normative truths relevant to these questions. What drives the treatment of issues under the law/fact distinction, then, is our (evolving) understanding of what is or is not morally fundamental or open to compromise—as it should be. The resulting account of transitions in law/fact classification may be one of error and

---

[227] *See* discussion *supra* Part I.C.

[228] *See* Dennis Chong, *Tolerance and Social Adjustment to New Norms and Practices*, 16 POL. BEHAV. 21, 35 (1994) (noting that the Supreme Court's late 1950s decisions limiting the reach of obscenity statutes coincided with the sexual revolution); Harold T. Christensen & Christina F. Gregg, *Changing Sex Norms in America and Scandinavia*, 32 J. OF MARRIAGE & FAMILY 616 (1970) (noting that attitudes towards sex liberalized considerably during the 50s and 60s).

correction, but, importantly, the need for such corrections stems from the complexity of the moral terrain and not from incoherence in the classificatory scheme.

Compare the present analysis of the "constitutional fact" doctrine with one that treats the importance of the constitutional values at stake as the principle grounds for the classification. The latter view falters in part because of the implausibility of viewing the constitutional values at stake as less important in areas where the Court has refused to review normative questions *de novo*, such as in the case of racial discrimination, as discussed earlier.[229] There is a different angle from which to approach the doctrine, one that emphasizes the relative importance (or unimportance) of pre-conventional/fundamental morality in deciding a normative issue. That factor appears to be a key driver of the doctrine's invocation. Pre-conventional moral concepts and norms appear more relevant in cases where the Court has refused to invoke the doctrine. In discrimination cases for instance, the normative question of whether the defendant was wrongfully discriminated against is appropriately answered by appeal to pre-conventional moral concepts of discrimination on the basis of race and morally arbitrary features.

Let us turn to cases where what were previously questions of law come to be treated as factual questions, as in the case of normative questions in death penalty cases, like the existence of mitigating and aggravating factors.[230] Is it plausible that the convention-independence of the "death-eligibility" question simply escaped judges for so long? The notion that even the worst criminal offenders might have a basic dignity unsuited for a punishment as severe as the death penalty is a relatively modern one.[231]

---

[229] *See* discussion *supra* Part I.C.

[230] *See* discussion *supra* Part I.C.

[231] The Court's death penalty jurisprudence is based on the Eighth Amendment, which "draw[s] its meaning from the evolving standards of decency that mark the progress of a maturing

More generally, fundamental moral rights are frequently overlooked; one glimpse at the long-tolerated practice of slavery is sufficient to confirm this fact.[232] So long as convention-independent moral facts can be elusive, judges may incorrectly classify as questions of law some normative questions that are decisively determined by pre-conventional moral facts, like the question of whether sufficiently many aggravating factors are present in a defendant's case to warrant consideration of the death penalty.

Our charge was either to explain based on the overarching framework how historically changing law/fact classifications were *not* the result of judicial error, or else to explain how these errors could happen. I have offered both kinds of explanation. Whether a normative question raised at trial should be classified as a question of law or fact is vexing precisely because the relevance of conventions is not always obvious. Correct application of the law/fact distinction depends on the entirely non-trivial issue of whether and to what extent legal outcomes, in various domains of law, turn on fundamental moral norms; and the boundary between fundamental morality and law can be elusive. Once this complexity is appreciated, we can no longer move so easily from the fact that judges change their minds about the law/fact status of normative questions that arise at trial to skepticism about the distinction's analytic coherence *vis-a-vis*

---

society," and in the past twelve years alone, capital punishment has been outlawed for the mentally handicapped, for minors, and for crimes other than murder and treason—all to bring our sentencing practices into alignment with the evolving moral standards of the citizenry. Trop v. Dulles, 356 U.S. 86, 101 (1958)*; see also* Kennedy v. Louisiana, 554 U.S. 407 (2008) (abolishing the death penalty for the rape of a child where the death of the victim was neither the result nor the intent); Roper v. Simmons, 543 U.S. 551 (2005) (abolishing the death penalty for individuals under the age of eighteen at the time of their capital crimes); Atkins v. Virginia, 536 U.S. 304 (2002) (abolishing the death penalty for the mentally retarded).

[232] On our bad track record with ensuring fundamental rights generally, *see* Evan. G. Williams, *The Possibility of an Ongoing Moral Catastrophe,* 18 ETHICAL THEORY & MORAL PRAC. 971 (2015).

normative questions. It is not always easy to determine whether a normative question is convention-dependent and so legal or convention-independent and thus factual.

IV. A Framework for Sixth Amendment Jurisprudence: Does the "Life-to-death" Judicial Override Infringe on the Jury's Fact-finding Responsibility?

The aim of this Article so far has been descriptive—to bring to the fore a principle implicit in courts' classification of normative issues as legal and factual.[233] I conclude with an application of the principle to an ongoing controversy. The controversy concerns the constitutionality of the "judicial override"—a legal mechanism still used by Alabama judges to override jury life-sentences in capital trials.[234] The "life-to-death" judicial override presents a constitutional puzzle that the Supreme Court has been grappling with over the last few decades, and as recently as January of 2016, when it decided, in *Hurst v. Florida*, to strike down similar override provisions in Florida's capital sentencing

[233] In this paper, I haven't provided a detailed causal mechanism by which the meta-normative distinction I've described might have influenced judges. One possibility is that judges have assigned normative questions to juries when they feel that they lack legal tools to decide the question – that is, when they feel out of their element. Common law judges are more likely to feel out of their element when they confront questions of basic morality, as opposed to normative questions that they can decide based on conventions. Thanks to Joshua Kleinfeld for this suggestion.

[234] On April 11th, 2017, Alabama governor Kay Ivey signed into law a bill banning the judicial override for defendants convicted after April 11th, but the law does not apply retroactively to defendants convicted before April 11th. S.B. 16, Reg. Sess. 2017 (Ala. 2017). Judges remain free to exercise the override for pre-April 11th murder convictions. The legislation will not affect the 183 inmates currently on Alabama's death row. Alabama judges have overridden juries 112 times since 1976. *See Alabama Ends Death Penalty by Judicial Override*, ASSOCIATED PRESS, (April 11, 2017, 6:30 PM), https://www.usnews.com/news/best-states/alabama/articles/2017-04-11/alabama-ends-death-penalty-by-judicial-override.

scheme.[235] This scheme of capital sentencing seems headed for constitutional prohibition, given recent developments in the Court's death penalty jurisprudence, not limited to its special emphasis on the jury's fact-finding responsibility in capital trials.[236] The aim, in what follows, is to suggest that the constitutional question may turn on a proper understanding of the law/fact distinction.

A. The Supreme Court's Recent Death Penalty Jurisprudence and the Judicial Override

Capital sentencing schemes across all states that have the death penalty follow the same basic structure. They require three findings before a defendant can be lawfully sentenced to death: (i) a finding of aggravating factors in the defendant's case; (ii) a finding of mitigating factors; and (iii) a balancing of aggravating against mitigating factors based on the "weight" of each.[237] A convicted defendant can lawfully receive the death penalty *only if* the aggravating factors are found to outweigh the mitigating factors. Aggravating factors might include multiple victims killed or injured in the course of committing the murder, prior convictions, or a lack of remorse.[238] Mitigating factors may include mental impairment, childhood abuse or neglect, and remorse.[239] Judicial override provisions grant authority to the trial judge to override a jury's determination of aggravating and mitigating factors' existence or weight. Prior to *Hurst*, only three

[235] 136 S. Ct. 616 (2016).
[236] *See* discussion *infra* Part IV.A.
[237] Modern sentencing schemes stem from the requirements outlined in Gregg v. Georgia, 428 U.S. 153, 153 (1976). *See also* Abramson, *supra* note 113, at 153.
[238] *See, e.g.*, 18 U.S.C.A § 3592 (2006).
[239] *Id.*

states—Indiana, Delaware, and Alabama—had override provisions in their capital sentencing schemes.[240]

In *Hurst,* decided only last year, the Court held that Florida's override scheme at the time violated the crucial holding in *Ring*,[241] that in capital cases any fact that exposes the defendant to a greater punishment than that authorized by the jury's guilty verdict must be submitted to a jury.[242] The *Hurst* Court found that the existence of an aggravating factor is a "fact that exposes the defendant to a greater punishment" and so must be found by the jury.[243] The jury's merely "advisory verdict" on the existence of aggravating factors was deemed not to satisfy *Ring's* requirements.

Whereas Alabama's scheme of judicial overrides avoids the problem found in Florida's—namely, the judge's ability to independently find aggravating factors; it nevertheless empowers judges to have the final say in *weighing* the aggravating against the mitigating factors.[244] Judges, in other words, can override the jury on the ultimate determination of whether death is the appropriate sentence. Moreover, in Alabama, the judge's overriding discretion is much less constrained than in Florida.[245] The jury's findings and recommendations need not be given any particular weight by the sentencing

---

[240] *See* Michael L. Radelet, *Overriding Jury Sentencing Recommendations in Florida Capital Cases: An Update and Possible Half-Requiem,* 2011 MICH. ST. L. REV. 793, 818 (2011) (listing overrides in Indiana); *id.* at 822–23 (listing overrides in Florida); *id.* at 825–27 (listing overrides in Alabama).

[241] Hurst, 136 S. Ct. at 617.

[242] Ring v. Arizona, 536 U.S. 584, 609 (2002).

[243] Hurst, 136 S. Ct. at 616.

[244] ALA. CODE §§ 13A-5-39 to -59 (2012).

[245] The Florida Supreme Court articulated, in Tedder v. State, what has come to be known as the "Tedder standard," which requires that "[i]n order to sustain a sentence of death following a jury recommendation of life, the facts suggesting a sentence of death should be so clear and convincing that virtually no reasonable person could differ." 322 So.2d 908, 910 (Fla. 1975). Nothing comparable to the Tedder standard constrains judges in Alabama.

judge. All that the state requires is that the judge "consider" the advisory verdict.[246] Given this broad discretion afforded to judges, Alabama is the *only* state where judges continue to routinely override jury life sentences. Between 1981 and 2011, ninety-three defendants were sentenced to death after the jury recommended life imprisonment.[247] Since 2000, twenty-six of the twenty-seven life-to-death overrides in the United States have occurred in Alabama alone.[248]

On April 11th, 2017, Alabama governor Kay Ivey signed into law a bill banning the judicial override for defendants convicted after April 11th, but the law does not apply retroactively to defendants convicted prior to that date.[249] Judges remain free to exercise the override for prior murder convictions. Accordingly, the legislation will not affect the 183 inmates currently on Alabama's death row and those due to be sentenced based on pre-April 11th convictions.[250]

In *Harris v. Alabama*, decided prior to *Ring*, *Apprendi*, and *Hurst*, the Supreme Court explicitly upheld Alabama's override scheme.[251] However, there is increasing evidence that some justices are prepared to review the Court's earlier decision. While the justices recently denied certiorari in a 2016 appeal that would have allowed them to reconsider the constitutionality of Alabama's override in light of *Hurst*, Justices Ginsburg and Sotomayor wrote a concurrence clarifying that the denial, in their minds, was not

[246] *See* Radelet, *supra* note 240, at 809.
[247] *Id.* at 801–02.
[248] Woodward v. Alabama, 134 S. Ct. 405, 407 (2013) (Sotomayor, J., dissenting from denial of certiorari).
[249] S.B. 16, Reg. Sess. 2017 (Ala. 2017).
[250] *Alabama Ends Death Penalty by Judicial Override*, ASSOCIATED PRESS, (April 11, 2017, 6:30 PM), https://www.usnews.com/news/best-states/alabama/articles/2017-04-11/alabama-ends-death-penalty-by-judicial-override.
[251] 513 U.S. 504 (1995).

based on the merits of the issue but was instead due to procedural constraints in the case preventing relief.[252] Moreover, they openly suggested that Alabama's capital sentence scheme is due a rehearing based on recent developments in the Court's capital jurisprudence on jury fact-finding.[253] The justices noted that *Harris*, which upheld Alabama's override, was based on cases that *Hurst* overruled.[254]

B. Alabama's Judicial Override on the Weight of Aggravating & Mitigating Evidence

Can Alabama's override scheme withstand the ruling in *Hurst* requiring that "any fact that expose[s] the defendant to a greater punishment than that authorized by the jury's guilty verdict . . . must be submitted to the jury"? The issue is complicated by the fact that *Hurst* dealt specifically with the practice of judges independently finding *aggravating factors* and thus deciding the death-*eligibility* issue. But a determination that particular aggravating factors exist is distinct from a finding that the aggravating factors *outweigh* the mitigating factors—or, in other words, the ultimate finding that the defendant should receive the death penalty. The critical question, then, is whether in light of *Hurst* and related cases, judges can still independently weigh the aggravating and mitigating factors and make the ultimate sentencing decision over the jury.

Justices Sotomayor and Breyer have suggested that the answer is no—judicial determination of the ultimate sentence, whether based on the jury's advice or not,

[252] Brooks v. Alabama, 136 S. Ct. 708 (2016) (Sotomayor, J., concurring in denial of certiorari).
[253] *Id.*

[254] *Id.*

unlawfully infringes on the jury's power to decide questions of fact in capital cases.[255] They could not have stated their position more clearly than in their 2013 dissent from a certiorari denial. The case involved an Alabama judge overriding a jury's eight-to-four life sentence recommendation after independently finding that the aggravating factors outweighed the mitigating factors. The justices observed that

> [t]he statutorily required finding that the aggravating factors of a defendant's crime outweigh the mitigating factors is therefore necessary to impose the death penalty. It is clear, then, that this factual finding exposes the defendant to a greater punishment than he would otherwise receive: death, as opposed to life without parole. Under *Apprendi* and *Ring*, a finding that has such an effect must be made by a jury.[256]

Whether other members of the Court agree remains to be seen, but the position requires a more detailed examination, one that I provide below.

Quite apart from the law/fact issue, there are various constitutional principles that seem to be in tension with Alabama's sentencing scheme. The override is hard to square with the Court's broader death penalty jurisprudence, which emphasizes, among other things, (1) the need for death sentences to enjoy broad-based communal support in order

[255] Woodward v. Alabama, 134 S. Ct. 405, 410–11 (2013) (Sotomayor, J., dissenting from denial of certiorari).
[256] *Id.*

to be consistent with the Eighth Amendment,[257] and (2) the need for capital sentencing procedures to guard against arbitrariness in death sentencing.[258] Shannon Heery makes a compelling empirical case that judicial overrides make death sentencing more arbitrary than it would otherwise be in states like Alabama.[259] Precisely because the overall case against the constitutionality of the override seems so strong, a number of commenters expect the Court to strike it down.[260] But it is worth clarifying the rationale for doing so based on the jury's fact-finding responsibility in light of the law/fact distinction.

C. The Weighing of Aggravating and Mitigating Factors Raises a Convention-independent ("Factual") Normative Question

Justices Sotomayor and Breyer are quite right that the ultimate determination that the aggravating factors outweigh the mitigating factors is a question of fact for the jury. But this does not simply fall out of *Hurst's* requirement that all "facts" raising the likelihood of a severer sentence must be found by the jury. The Court emphasized in *Ring* that the scope of the Sixth Amendment jury trial right turns on the understanding of the jury's role that prevailed at the time of the Amendment's adoption.[261] It cited

---

[257] *See, e.g.*, cases cited *supra* note 72; *see also* Steve Semeraro, *Responsibility in Capital Sentencing*, 39 SAN DIEGO L. REV. 79, 144 n.232 (2002) ("[T]he case law as a whole indicates that communal values must play a role in capital sentencing.").

[258] *See* Gregg v. Georgia, 428 U.S. 153, 194–95 (1976) (noting that "the concerns expressed in Furman" included that the death penalty "not be imposed in an arbitrary or capricious manner").

[259] Shannon Heery, *If It's Constitutional, Then What's the Problem?: The Use of Judicial Override in Alabama Death Sentencing*, 34 WASH. U. J. L. & POL'Y 347, 348 (2010).

[260] *See, e.g.*, Radelet, *supra* note 240, at 817. ("Perhaps the uniqueness of the Alabama law and the increasing dissimilarity of its override cases to other death penalty cases will soon lead the Supreme Court to find that such overrides clearly violate evolving standards of decency and guarantee the power of the jurors to firmly reject the death penalty. Stay tuned.").

[261] Ring v. Arizona, 536 U.S. 584, 599 (2002).

approvingly the work of Welsh White, who observes that English legal scholars and the common law understanding of the distinction between legal questions and factual questions "undoubtedly influenced the framers."[262] White cites Blackstone, who explicitly appeals to the law/fact distinction in delineating the jury's authority in civil and criminal cases.[263] Moreover, the distinction's relevance is well-established under the Court's interpretation of the Seventh Amendment jury trial right. Jury trial rights under the Sixth and Seventh Amendments have been implemented in coordinate fashion and have been understood to serve similar societal functions.[264] Accordingly, a "finding of fact," under *Hurst*, can reasonably be interpreted as referring to the law's traditional understanding of factual as opposed to legal questions. Under this line of reasoning, even if the capital sentencing judge's final determination that the defendant deserves the death penalty obviously "increases the likelihood of a more severe penalty," it does not violate *Hurst* unless it also represents a "finding of fact" as opposed to a "finding of law."

Whether the final determination is factual is not obvious. For one, it rests on a normative evaluation—a weighing of aggravating and mitigating factors. The final weighing of the evidence represents an "ethical judgment," and in the "final analysis, capital punishment rests on not a legal but an ethical judgment—an assessment of . . . the 'moral guilt' of the defendant."[265] As previously discussed, there are many examples of

---

[262] *Id. See also* White, *supra* note 44, at 4–5.

[263] White, *supra* note 44, at 4 ("[T]he principles and axioms of law . . . should be deposited in the breasts of the judges . . . . But in settling and adjusting a question of fact . . . a competent number of sensible and upright jurymen . . . will be found the best investigators of truth and the surest guardians of public justice.") (quoting 3 W. BLACKSTONE COMMENTARIES * 379–80).

[264] *See* Kirgis, *supra* note 67, at 902–03. *See also* Cooper Indus., Inc. v. Leatherman Tool Grp., Inc., 532 U.S. 424 (2001).

[265] Spaziano v. Florida, 465 U.S. 447, 481 (1984) (Stevens, J., concurring in part and dissenting in part).

normative questions that have historically been treated as questions of *law*, as in contract law.[266] To consider an especially pertinent example, in *Cooper Industries*, the Court held that punitive damages findings in the context of unfair competition are intermediate between law and fact, emphasizing precisely the normative character of such findings.[267] Hence, the application of *Hurst*—and the *Ring-Apprendi* line of cases on which it builds—to the life-to-death judicial override turns crucially on whether the normative evaluation that ultimately determines a death sentence is factual as opposed to legal.

To put the issue in terms of the present framework, the question is whether the normative truths pertaining to the weight of the relevant evidence are convention-dependent or convention-independent. In the case of punitive damages in the economic context, it is very plausible to think that law-related conventions have a role to play in determining whether conduct should be punished financially. It is far less plausible to think that the question of whether a criminal defendant deserves the death penalty based on the facts about his crime and his circumstance turns in any significant way on our conventions—how judges treat similarly situated defendants, for instance.[268] Conventions cannot *favor* the imposition of the death penalty where the penalty is morally undeserved, given the gravity of the harm inflicted and the importance of protecting defendant's from undeserved execution.[269] An affirmative answer to the death penalty question must be settled by morally weighing the particular facts in the defendant's case and considering whether the death penalty would be consistent with the

[266] *See* discussion *supra* Part I.B, III.B.
[267] Cooper Indus., Inc., 532 U.S. 424.

[268] *See* Gregg v. Georgia, 428 U.S. 153, 182 (1976); discussion *supra* Part III.A.
[269] Gregg, 428 U.S. at 187 ("When a defendant's life is at stake, the Court has been particularly sensitive to insure that every safeguard is observed.").

fundamental rights of persons.[270] In other words, Alabama's override scheme is inconsistent with *Hurst* because it empowers the judge, alone, to rule on a convention-*independent* normative question, or a question of fundamental moral fact. Once the nature of the normative question is properly understood, Sixth Amendment jurisprudence and the logic of the law/fact distinction can be seen to support the conclusion that the life-to-death judicial override unconstitutionally impinges on the jury's fact-finding responsibility in criminal trials.

The Court might find it advantageous to rely on an apolitical rule in deciding the issue, given that capital sentencing continues to be morally and politically contested. As such an apolitical rule, the common law's distinguishing of legal from factual normative questions, combined with a straightforward application of the Court's recent holdings, seems a powerful basis on which to eliminate the life-to death override.

Conclusion

Leon Green once observed that "[n]o two terms of legal science have rendered better service than 'law' and 'fact,'" given the many uses to which they have been put, and, so, warned that "the man who could succeed in defining them would be a public enemy."[271] Green's warning, while partly ironic, touches on a legitimate concern one might have about attempts to define legal terms and distinctions. Definitions impose constraints which, when artificial, can rob legal distinctions of the flexibility that made

[270] *See* cases cited *supra* note 186, 191.
[271] GREEN, *supra* note 9, at 270.

them useful to judges in the first place. But the importance of objective constraints on judicial discretion that can be specified *ex ante* is equally worthy of emphasis: they are essential to the rule of law.[272] Walking the thin line between these two concerns, I have sought to offer not a full definition of 'law' and 'fact,' but an intuitive principle that can inform the classification of normative questions under the distinction. The principle should promote the integrity of law in this domain, while respecting the reality of judicial judgment.

Simply put, normative questions that essentially depend on conventions—what we do around here—and the practices of legal officials in particular, are aptly described as "questions of law"; normative questions that are primarily independent of such conventions because they turn on what we *ought* to do as a matter of basic justice are better described as questions of fact. This principle explains settled law/fact classifications in a broad range of legal domains, including torts, contracts, First Amendment law, and criminal procedure. It also points towards a possible solution to a looming controversy over judicial involvement in capital sentencing.

* * *

ACKNOWLEDGMENTS. I am indebted to Erin Miller for her extensive and very helpful feedback on earlier versions of this paper. The paper was presented at law schools at Washington & Lee, Pennsylvania State University, University Park, Columbia

[272] To quote Justice Benjamin Cardozo, "[a] jurisprudence that is not constantly brought into relation to objective or external standards, incurs the risk of degenerating into . . . a jurisprudence of mere sentiment . . . ." BENJAMIN N. CARDOZO, THE NATURE OF THE JUDICIAL PROCESS 106 (1921).

University, University of Oregon, University of Idaho, Cornell University, Arizona State University, University of Southern California, George Washington University, and Fordham University, during the winter of 2017-18. I'm grateful to all who attended the workshops. I am especially grateful for the thoughtful comments I've received from Scott Altman, Akhil Amar, Alisabeth Ayars, Ian Ayres, Kevin Clermont, Chad Flanders, Stephen Galoob, Joshua Kleinfeld, Daniel Markovits, Jens Ohlin, Michael Pardo, Philip Pettit, Joseph Raz, Daria Roithmayr, Gideon Rosen, Nomi Stolzenberg, Nelson Tebbe, Michael Smith, Nat Tabris, Gideon Yaffe, and Benjamin Zipursky.

Chapter 7. The Limits of Law in the Evaluation of Mitigating Evidence

Emad H. Atiq & Erin L. Miller[1]

Abstract. Capital sentencers are constitutionally required to "consider" any mitigating evidence presented by the defense. Under *Lockett v. Ohio* and its progeny, neither statutes nor common law can exclude mitigating factors from the sentencer's consideration or place conditions on when such factors may be considered. We argue that the principle underlying this line of doctrine is broader than courts have so far recognized.

A natural starting point for our analysis is judicial treatment of evidence that the defendant suffered severe environmental deprivation ("SED"), such as egregious child abuse or poverty. SED has played a central role in the Court's elaboration of the "consideration" requirement. It is often given what we call "restrictive consideration" because its mitigating value is conditioned on a finding that the deprivation, or a diagnosable illness resulting from it, was an immediate cause of the crime. We point out, first, that the line of constitutional doctrine precluding statutory and precedential constraints on the consideration of mitigating evidence rests on a more general principle that "consideration" demands an individualized, moral—as opposed to legalistic—appraisal of the evidence. When judges restrict the moral principles under which they evaluate the mitigating weight of evidence on the basis of precedent or even judicial

[1] The authors are equally responsible for the ideas and writing within this article; the ordering of names is alphabetical.

custom, they fail to give a reasoned, moral response to the evidence. We articulate a three-factor test for when legalistic thinking of this sort prevents a judge from satisfying the constitutional requirement. Restrictive consideration of SED evidence, in many jurisdictions, is a product of legal convention and thus fails the test.

Second, we contend that, when the capital sentencer is a judge rather than a jury, she has a special responsibility to refrain from restrictive consideration of mitigating evidence. The Constitution requires that death sentences must be consistent with community values. Unrestricted consideration of evidence—evaluating its mitigating weight in light of a range of moral principles—ensures that the diverse moral views of the community are brought to bear on the capital question.

"The sentencer must . . . be able to consider and give effect to [mitigating] evidence in imposing a sentence, so that the sentence imposed . . . reflects a reasoned *moral* response to the defendant's background, character, and crime."[2]

Introduction

It is well-established law, since *Eddings v. Oklahoma*, that evidence of "severe environmental deprivation" (SED)—such as egregious child abuse, neglect, or poverty—must be "considered" by judges as a mitigating factor during the penalty phase of capital trials.[3] In *Smith v. Texas*, the Supreme Court found unconstitutional under *Eddings* a judicial practice of excluding SED from consideration as a potential mitigating factor unless the deprivation suffered met a narrowly defined condition.[4] Rather than broadening their review of SED, the judges who previously engaged in this practice of outright exclusion switched to a subtly different practice: when SED evidence is presented by the defense, judges declare that they are "considering" SED as a mitigating factor but assign it little to no mitigating weight unless it meets the very same condition.[5]

---

[2] Penry v. Johnson, 532 U.S. 782, 788 (2001) (citations omitted).

[3] 455 U.S. 104, 113 (1982); *see also id.* at 114–15 ("Just as the State may not by statute preclude the sentencer from considering any mitigating factor, neither may the sentencer refuse to consider, *as a matter of law*, any relevant mitigating evidence . . . . The sentencer . . . may determine the weight to be given relevant mitigating evidence. But they may not give it no weight by excluding such evidence from their consideration."). Henceforth, we use the same *Eddings* pin citation for all grammatical forms of "consider."

[4] Smith v. Texas, 543 U.S. 37, 45 (2004) (per curiam).

[5] *See* discussion *infra* Part I.

Mitigating factors that receive little to no weight make no difference, as far as we can tell, to the defendant's sentence.[6] The practice raises the question: is judicial treatment of SED evidence consistent with the kind of "consideration" the Constitution requires?

Courts of appeals declined to take a position on the issue—until the Ninth Circuit's ruling in *McKinney v. Ryan* in December 2015.[7] In *McKinney*, the Arizona Supreme Court had dismissed SED evidence as non-mitigating because it did not "causally contribute" to the capital crime, claiming that this counted as "consideration" under *Eddings*. A divided Ninth Circuit, sitting *en banc,* disagreed, finding that the state court failed "to evaluate and give appropriate weight" to that evidence, contrary to *Eddings*, because the causal prerequisite it invoked mirrored the one that, until *Smith*, it had used to wholly exclude most SED evidence from consideration.[8] In an impassioned dissent, Judge Bea described the notion that the Arizona Supreme Court "did not *really* consider [the evidence] even though it used the word 'considering'" as "nonsense."[9] He argued that "giving little or no weight to such evidence [after consideration] is perfectly permissible under *Eddings*."[10]

Two dueling approaches to the "consideration" of deprivation evidence underpin this dispute.[11] A fact offered as mitigating by the defendant can only be judged mitigating

---

[6] *See* discussion *infra* Parts I, II.

[7] McKinney v. Ryan, 813 F.3d 798, 802 (9th Cir. 2015) (en banc), *cert. denied*, 137 S. Ct. 39 (2016).

[8] *Id.* at 820, 823.

[9] *Id.* at 847 (Bea, J., dissenting) (emphasis in original).

[10] *Id.* at 843–44 (Bea, J., dissenting).

[11] *See* discussion *infra* Parts I, II.

based on a principle concerning moral responsibility or punishment. On some such principles, the fact might have greater mitigating value than on others. For instance, evidence of an act of kindness of the defendant might be mitigating given the principle that mercy is appropriate towards individuals of decent moral character or the principle that even murderers who may be rehabilitated should be spared execution. When a sentencer draws on just one normative principle, or an unduly restricted range of plausible principles, to explain the evidence's mitigating value, they engage in what we call *restrictive consideration*.[12] Restrictive consideration is not necessarily unlawful, but the consideration found inadequate in *McKinney* was both restrictive and unlawful. In that case, and many others in Arizona, the defendant's deprivation was deemed to have mitigating value *only if* it bore a very particular causal relation to the criminal act: namely, that the SED was an immediate or "specific" cause of the act. (For example, SED causes the crime in the relevant sense when it results in a psychological disorder like PTSD that results in an irresistible impulse or motive to commit the crime in question.) This restriction seems to rest on the principle that a defendant's prior deprivation only diminishes his punishment-worthiness when the deprivation directly causes his intention to commit the crime and negates his responsibility for the crime. The *McKinney* majority sought less restricted consideration, which would have appraised the mitigating significance of the deprivation based on alternative moral principles. In what follows, we demonstrate that numerous such alternative principles exist and are not only plausible but widely accepted.

[12]By plausible principles, we mean those that are believed by significant numbers of reasonable persons and should be known to the sentencer. *See* discussion *infra* Part I.

We welcome *McKinney* as a clarification of the *Eddings* consideration doctrine. We argue that implicit in *Eddings* and its progeny is the attractive ideal that it is unconstitutional for sentencers to limit the moral principles under which they consider mitigating evidence for *legalistic* reasons; in evaluating which moral principles bear on the mitigating significance of evidence presented by the defense, the sentencer should rely exclusively on moral reasoning. *Eddings* explicitly held that capital sentencers must not be constrained by legal norms from considering any relevant mitigating evidence.[13] The holding was an extension of *Lockett v. Ohio*, which held that statutes excluding any mitigating factors from the sentencer's consideration are unconstitutional.[14] *Eddings* elaborated that the sentencer's consideration can be unconstitutionally constrained not just by statutes but also other sources of law, like judicial custom. And legal rules can operate as unconstitutional constraints not just by requiring outright exclusion of mitigating factors from consideration, but by subtly pressuring judges to limit the conditions under which evidence counts as mitigating. A later case, *Tennard v. Dretke*, clarified that judge-made rules or conventions limiting when mitigating evidence can be considered also amount to unconstitutional constraints on consideration.[15] We argue that *Lockett*, *Eddings*, and *Tennard*, together, stand for the proposition that a practice of restrictive consideration of mitigating evidence where the restrictions are imposed because judges feel *bound by the law* (in a sense to be made precise) is unconstitutional.

[13] Eddings v. Oklahoma, 455 U.S. 104, 115 (1982).

[14] Lockett v. Ohio, 438 U.S. 586, 606 (1976).

[15] Tennard v. Dretke, 542 U.S. 274, 287 (2004).

McKinney took a step toward this broader doctrinal interpretation by finding an *Eddings* violation in restrictive consideration of SED induced by an informal judicial practice. However, because the Ninth Circuit based its decision on historical facts specific to the Arizona practice, it missed an opportunity to articulate a general rule for identifying when restrictive consideration counts as unconstitutionally induced by a legal custom or practice under *Eddings*. We seize the opportunity *McKinney* missed, offering a three-factor test for just this purpose that applies most obviously to the review of SED evidence and potentially to the review of mitigating evidence more broadly.

We also present an argument, grounded in an original interpretation of Supreme Court precedent, that restrictive consideration of mitigating evidence may be inherently or *per se* unconstitutional when the sentencer is a judge, even if the judge was not acting on the basis of any assumed legal rules. As the Court has repeatedly emphasized since *Gregg v. Georgia*, a death sentence cannot be constitutionally legitimate unless it enjoys broad-based communal support.[16] This is in part why juries—representing a cross-section of their community—are so extensively involved in the administration of capital

[16] *See* Gregg v. Georgia, 428 U.S. 153, 190 (1976) ("Jury sentencing has been considered desirable in capital cases in order 'to maintain a link between contemporary community values and the penal system—a link without which the determination of punishment could hardly reflect 'the evolving standards of decency that mark the progress of a maturing society.'" (citation omitted)); Woodson v. North Carolina, 428 U.S. 280, 297–98 (1976) (reflecting on the importance on the moral views of society in the administration of death penalty). *See also* McCleskey v. Kemp, 481 U.S. 279, 310 (1987) (outlining the capital jury's task of expressing "the conscience of the community on the ultimate question of life or death"); Spaziano v. Florida, 468 U.S. 447, 461 (1984) ("[T]he decision that capital punishment may be the appropriate sanction in extreme cases is an expression of the community's belief that certain crimes are themselves so grievous an affront to humanity that the only adequate response may be the penalty of death.") (quotation marks omitted) (citing *Gregg*, 428 U.S. at 184). *Accord* Steve Semeraro, *Responsibility in Capital Sentencing*, 39 SAN DIEGO L. REV. 79, 144 n.232 (2002) ("[T]he case law as a whole indicates that communal values must play a role in capital sentencing.").

punishment in nearly every jurisdiction in the United States legal system. We argue that, because of their comparative disadvantage at fulfilling this constitutional function, judges who issue death sentences have a unique responsibility to consider each piece of mitigating evidence in light of different moral theories that give it the broadest potential mitigating value; and to give significant weight to the evidence if, under some such theories, it has significant mitigating value. Doing so does a better job ensuring that the death penalty if issued would enjoy broad based communal support than a practice of restrictive consideration (even if morally motivated). Given the common and (we argue) reasonable belief that SED is inherently mitigating, judges should give unrestricted consideration to deprivation evidence and assign substantial (though not necessarily dispositive) mitigating weight to it.

While the active controversy over what "consideration" requires has centered in the Ninth Circuit, the question is even more pressing in other jurisdictions. Although most states have shifted to exclusive jury sentencing in capital cases, Alabama, until last year, continued to allow death sentencing by a single judge through a jury override provision, and required no deference to the jury's preference for life; over a hundred inmates on Alabama's death row were subject to this provision and might still appeal their sentences.[17] Alabama is in the Eleventh Circuit, which has shown no signs of following the Ninth's lead in giving teeth to *Eddings's* "consideration" requirement. Prior to a significant shift in Supreme Court doctrine in 2002, many other states also employed

[17] *See* ALA. CODE §§ 13A-5-39 to -59 (2012); *accord* FLA. STAT. ANN. § 921.141(3) (West 2001). The Supreme Court recently struck down the Florida override in part, but the rest survives intact. Hurst v. Florida, 84 USLW 4032 (2016).

judicial capital sentencing, and likely still have inmates on death row who were sentenced by judges under these older regimes.[18]

In Part I, we illustrate how restrictive consideration can become an entrenched judicial practice, using examples of SED review from Arizona, Alabama, and Florida. We attempt to understand the underlying moral principle, called here the "causal nexus theory," which treats SED as mitigating when it has effects at the time of the crime that undermine the defendant's control over his act, similar to those of a serious mental illness. We find, however, that judges in these districts offer no justification for ignoring all *other* moral principles under which SED could have mitigating value.

In Part II, we review recent work in moral philosophy on the mitigating significance of SED, which informs our argument that the causal nexus theory is neither the only nor the most charitable available theory of SED's mitigating value. We make a brief case for the plausibility of three theories that regard SED as mitigating without proof of direct and specific causation, as well as for their popularity among capital jurors.

In Part III, we provide a two-pronged constitutional rationale for appellate courts to scrutinize lower courts' restrictive consideration of SED evidence. First, we trace the Supreme Court's jurisprudence on what constitutes adequate "consideration" of mitigating evidence at trial, arguing that the thread that unifies the holdings in *Lockett*,

[18] Prior to the Supreme Court's decision in *Ring v. Arizona*, 536 U.S. 584 (2002), which required extensive jury involvement in capital sentencing, eight states in addition to Alabama—Arizona, Colorado, Delaware, Florida, Idaho, Indiana, Montana, and Nebraska—gave judges either exclusive authority to issue a death sentence or final authority with some level of input from the jury. *See generally* ARIZ. REV. STAT. ANN. § 13-703 (2001); COLO. REV. STAT. ANN. § 16-11-103 (West 2001); DEL. CODE ANN. tit. 11, § 4209 (West 2001); FLA. STAT. ANN. §921.141 (West 2001); IDAHO CODE ANN. § 19-2515 (West 2001); IND. CODE ANN. § xx-xx-x (West 2001); MONT. CODE ANN. § 46-18-301 (West 2001); NEB. REV. STAT. ANN. § 29-2520 (West 2001).

Eddings, and *Tennard* is the principle that the moral theories used by a sentencer to consider relevant mitigating evidence cannot be subject to legal constraint—whether statutory, precedential, or a matter of judicial custom. We articulate three factors for evaluating whether restrictive consideration of deprivation evidence violates this principle: (i) the court did not even attempt to justify or explain why the moral theory it used was the appropriate one to rely on, or why alternative theories were and should be dismissed; (ii) the same court, or other courts in its jurisdiction, have in the past routinely and without justification used the same theory—and only that theory—in considering mitigating evidence, while citing to precedent; and (iii) independent reasons exist for thinking that a substantial number of reasonable jurors would consider the evidence broadly mitigating on other moral grounds that the judge did not consider. *Second*, we make the case for sentencing judges at both the trial and appellate level having a unique responsibility to ensure that death sentences are issued only when they enjoy broad-based communal support.[19] Applied to the SED context, this means ensuring that SED evidence is given unrestricted consideration regardless of the judge's particular moral beliefs.

I. A Troubling Case of Restrictive Consideration: The Causal Nexus Requirement for SED

Nearly all death penalty states require three findings before the issuance of the death penalty: a finding of "aggravating factors," a finding of "mitigating factors," and a

[19] *See, e.g.*, *Gregg*, 428 U.S. at 181–84 (1976); discussion *infra* Part III.

balancing of aggravating against mitigating factors based on the "weight" of each.[20] The weight of an aggravating or mitigating factor represents the degree to which it militates in favor of or against the death penalty. A death sentence is legally justified only if the aggravating factors outweigh the mitigating ones. Rules restricting the potential weight of relevant mitigating evidence can, therefore, make the difference between life and death for a defendant, at least in cases involving few or insignificant aggravating factors. The judicial custom of considering the mitigating value of SED evidence *only* on the causal nexus theory, prevalent in multiple jurisdictions, has been restrictive in precisely this way.

A. The Causal Nexus Requirement in Arizona

As mentioned earlier, the causal nexus theory once functioned as an exclusionary rule in Arizona. Under the old rule, SED evidence would be outright excluded from consideration unless the defendant was able to show that the deprivation "caused" the crime or "had an effect or impact on his behavior" at the time of the crime.[21] In practice,

---

[20] The current capital sentencing scheme in most states has emerged from the requirements articulated in *Gregg*, 428 U.S. at 189–96. *See also* Jeffrey Abramson, *Death-Is-Different Jurisprudence and the Role of the Capital Jury*, 2 OHIO ST. CRIM. L. J. 117, 153 (2004) (describing that scheme).

[21] *See, e.g.*, Poyson v. Arizona, 743 F.3d 1185, 1193 (9th Cir. 2013) (quoting the Arizona trial court's statement that "[t]he court finds absolutely nothing in this case to suggest that [the defendant's commission of the murder] was a result of his childhood"); State v. Phillips, 46 P.3d 1048, 1060 (Ariz. 2002) ("[A]lthough Phillips presented evidence of substance abuse and a difficult childhood, he did not offer any evidence that these factors caused him to commit the robberies." (citation omitted)); State v. Djerf, 959 P.2d 1274, 1289 (Ariz. 1998) ("[D]ifficult family background is not relevant unless the defendant can establish that his family experience is linked to his criminal behavior." (citation omitted)); State v. Mann, 934 P.2d 784, 795 (Ariz. 1997) ("Defendant did not show any [causal] connection."); State v. Towery, 920 P.2d 290, 311 (Ariz. 1996) (en banc) ("These events, however, occurred when Defendant was young, years before he robbed and murdered at the age of 27. They do not prove a loss of impulse control or

the rule demanded proof that the SED was a specific cause.[22] Accordingly, the test set a high bar for admission.[23] Few defendants could offer the required proof, for reasons we discuss below.[24]

Once the Supreme Court invalidated a similar exclusionary rule in the Fifth Circuit, judges switched from "excluding" SED evidence to "considering" it but assigning "little to no mitigating weight" unless the defendant could establish the required causal nexus.[25] The sentencing procedure was "indistinguishable" in practice "from an analytical 'screen' that excludes such evidence from consideration as a matter

---

explain what caused him to kill."); State v. Murray, 906 P.2d 542, 573 (Ariz. 1995) ("[D]ifficult family background is nonmitigating unless defendant can show that something in that background impacted his behavior in a way beyond his control." (citation omitted)). Some early cases added that the "effect or impact" had to be "beyond the defendant's control." *E.g.*, State v. Murray, 906 P.2d 542, 573 (Ariz. 1995); State v. Wallace, 773 P.2d 983, 986 (Ariz. 1989). There were (rare) exceptions. *See generally* State v. Herrera, 850 P.2d 100 (1993) (life sentence in part because of "dysfunctional family background"); State v. Rockwell, 775 P.2d 1069 (1989) (life sentence in part because of SED).

[22] Many of the cases suggested that the causal link they sought was at the moment of the crime, such as an impulse or mental health symptom. *See, e.g.*, State v. Hoskins, 14 P.3d 997, 1022 (2000) (en banc), *supplemented*, 65 P.3d 953 (2003) ("Where we determine questions of aggravation and mitigation in the sentencing process, the significant point in time for causation is the moment at which the criminal acts are committed. If the defendant's personality disorder or dysfunctional family background leads reasonable experts to conclude that the disorder in fact caused the crime, significant mitigation is established."); *Mann*, 934 P.2d at 795 ("An abusive background is usually given significant weight as a mitigating factor only when the abuse affected the defendant's behavior at the time of the crime.").

[23] In a review of cases since *Eddings*, we have found only two in which the court applied the causal nexus test but found the SED sufficiently mitigating to recommend against the death penalty. *See generally* State v. Trostle, 951 P.2d 869 (Ariz. 1989); State v. Bocharski, 189 P.3d 403 (Ariz. 2008).

[24] *See* discussion *infra* Part II.

[25] *See, e.g.*, State v. Prince, 250 P.3d 1145, 1170 (Ariz. 2011) (en banc) ("We consider [SED evidence from the defendant's childhood] in mitigation but give it little weight."); State v. McCray, 183 P.3d 503, 511 (Ariz. 2008) ("A difficult family history is considered in mitigating, but its strength depends on whether the defendant can show it has a causal connection with the crime." (citation omitted)).

of law."[26] In practice, the results of restrictive consideration and exclusion were the same. We have found no case in which SED evidence was treated as having "little or no weight" but in which the defendant was ultimately sentenced to life imprisonment.[27] Indeed, the evidence suggests that judges who assign SED "little to no" mitigating weight regard it as wholly non-mitigating.[28] These cases are now constitutionally suspect under the ruling in *McKinney*.[29] However, the Ninth Circuit's ruling was narrow: it placed substantial weight on the fact that the Arizona Supreme Court, despite claiming that it

---

[26] *Poyson*, 743 F.3d at 1205. *See also id.* ("Simply altering the label attached to an unconstitutional process does not magically render it constitutional.").

[27] A survey of Arizona capital cases makes clear that mitigating evidence given "little" or "slight" weight rarely, if ever, results in leniency. *See Prince*, 250 P.3d at 1170 ("little" weight); State v. Harrod, 183 P.3d 519, 534 (Ariz. 2008) (en banc) ("minimal weight"); *McCray*, 183 P.3d at 503 ("little weight in mitigation"); *Hoskins*, 14 P.3d at 1022 (trial court accorded the SED evidence "slight" weight). Numerous other cases say that the lack of a causal nexus merely "lessens" the mitigating value of the SED evidence. While we suspect—and believe that an appellate court could find—that these cases, too, give little to no mitigating weight to the SED presented, because they do not address other theories under which the SED could be morally relevant, we do not address them here. *See, e.g.*, State v. Hampton, 140 P.3d 950, 968 (Ariz. 2011) (en banc) ("[The defendant's] troubled upbringing is entitled to less weight as a mitigating circumstance because he has not tied it to his murderous behavior."). Evidence assigned little to no weight is often excluded from the judge's final list of mitigating factors.

[28] Before affirming a death sentence, Courts routinely attach "little" weight to all of the mitigating factors—as though to emphasize that the mitigating evidence, even cumulatively, could not be decisive. *See, e.g.*, State v. Armstrong, 189 P.3d 378, 392-93 (Ariz. 2008) (en banc) (state supreme court dismissed each of the following mitigating factors as having "little" weight: the negative impact of Armstrong's death sentence on his children, his "troubled and unstable upbringing," his mental health history, and his "compassionate nature" and then affirmed the death sentence); State v. Murdaugh, 97 P.3d 844, 860 (Ariz. 2004) (en banc) (naming five mitigating factors, all of which the trial court had assigned "little weight" before imposing the death sentence); State v. Moody, 94 P.3d 1119, 1168 (Ariz. 2004) (en banc) ("[The trial judge] gave little weight to the [four mitigating factors] . . . and concluded that they were insufficient to call for leniency."). Evidence assigned little to no weight is often excluded from the judge's final list of mitigating factors. *See e.g., Poyson*, 743 F.3d at 1210 ("For at the end of its opinion, the state court listed all of the mitigating circumstances it considered in its independent review of Poyson's death sentence. It omitted from this critical tally both Poyson's personality disorders and his abusive childhood.").

[29] *McKinney*, 813 F.3d at 798.

"considered" the evidence, included a pin cite to an older case that relied on the unconstitutional exclusionary rule.[30]

The nature of the causal nexus demanded by Arizona judges becomes clearer upon comparing cases in which SED was treated as mitigating with cases in which it was not. For instance, in the only recent case in Arizona where SED was given substantial weight, *State v. Bocharski*, a psychologist testified that events leading up to the murder triggered symptoms of the defendant's post-traumatic stress disorder, which stemmed from his childhood trauma.[31] In that case, the defendant had been "severely abused emotionally, physically, and sexually as a child" and had suffered from extreme neglect.[32] The court observed that "in assessing the quality and strength of the mitigation evidence" it looks to the "strength of a causal connection between the mitigating factors and the crime."[33] It noted as "evidence of a causal connection" the fact that the psychologist "testified that Bocharski's troubled upbringing helped *cause* the murder of [the victim]."[34] The following facts were cited as supporting that determination: that the murder occurred immediately after a conversation between the defendant and the victim about the defendant's childhood abuse; that one especially traumatizing facet of that

[30] *Id.* at 820.

[31] 189 P.3d 403, 423 (Ariz. 2013) (en banc). *Bocharski* also mentions that the defendant suffered problems with alcoholism from a young age, and that he was in an alcoholic state on the day of the murder that may have made it harder for him to control his actions. *Id.*

[32] *Id.* at 424–25 (listing childhood hardships that included abandonment; physical abuse and extreme neglect, including starvation, by his mother; squalid living conditions with little privacy; poverty that required foraging in garbage cans; exposure to drugs and sex at a young age; and repeated foster care).

[33] *Id.* at 426 (citing *Hampton*, 140 P.3d at 968).

[34] *Id.* at 426 (emphasis added).

abuse involved the malicious killing of the defendant's childhood pet animals; and that the victim mistreated her pets. The explanation that elicited a merciful response from the court was that the defendant's deprivation made him vulnerable to stressful emotions when confronted with animal mistreatment, and the circumstances leading up to the murder placed him in a disturbed emotional state in which he was less able to "control and manage his feelings and reactions."[35]

The Court's reasoning in *Bocharski* contrasts with its reasoning in a case decided that same year where SED was given no weight. In *State v. Ellison*, the defendant argued that the abuse he suffered as a child significantly impaired his capacity to make moral choices as an adult.[36] A psychologist testified that "for a person having experienced Ellison's upbringing [and] history of physical and sexual abuse . . . , the damage would carry on into adulthood and potentially destroy the individual."[37] Yet the court determined that the defendant's "childhood troubles deserve[d] little value as a mitigator," given that he had "not provided any specific evidence that his brain chemistry was actually altered . . . so as to *cause* or *contribute* to his participation in the murders."[38] Notably, the court conceded that the psychiatric testimony made it more than likely "that Ellison did suffer some mental or emotional damage due to his [SED]." However, it could not find in this fact any grounds for mitigation.[39]

---

[35] *Id.* at 423.

[36] 140 P.3d 899 (Ariz. 2006) (en banc).

[37] *Id.* at 928.

[38] *Id.* at 927-28 (emphasis added).

[39] *Cf.* State v. Anderson, 111 P.3d 369, 399 (Ariz. 2005) (en banc) (finding that the defendant's evidence of sexual abuse, low IQ, frequent moves between schools, and follower-type personality "do[es] not in any way explain his decision, decades later at age forty-eight, to kill three innocent

The different outcomes in *Ellison* and *Bocharski* seem to have turned on the different types of causal connections that the defendants drew between their childhood deprivations and their crimes. In *Ellison,* the nexus was a fairly general one: the defendant's emotional and mental traits, which were shaped by the SED, and, it could be inferred, played a role in his resort to crime. In *Bocharski,* the causal nexus was specific: the defendant's post-traumatic stress disorder, which was originally caused by his abuse and triggered by memories of the abuse at the time of the crime. In other words, the SED did not shape his moral and decision-making faculties themselves, but simply, via the PTSD, subverted them at the time of the murder. Because no such specific and direct causal link between the SED and the crime could be established in *Ellison,* the court deemed evidence that the defendant had suffered from much the same kind of extreme deprivation as Bocharski to be "not of such a quality or value as to warrant leniency."[40]

Similarly, in *State v. Prince* in 2011, the Arizona Supreme Court cursorily dismissed evidence of even severer deprivation—that the defendant's father was "an alcoholic, abusive to his wife and children and often on the run from law enforcement," and that as a child the defendant lived in an old barn that lacked adequate heat, running water, a kitchen, or a bathroom and then as a teenager with an adult male who molested and sexually abused him[41]—as having "little weight" because the defendant did not "establish[] a connection between his childhood trauma and the murder."[42] A lone

---

people to steal a pickup," as defendant was not mentally retarded and was able to tell right from wrong in making his own decisions).

[40] *Ellison*, 140 P.3d at 928.

[41] State v. Prince, 250 P.3d 1145, 1170 (Ariz. 2011) (en banc).

[42] *Id*. at 1170–71.

citation to *Bocharski* made clear that the court sought a causal "connection" of a very immediate sort, such as a mental illness traceable to the deprivation that prompted the defendant to commit the crime, or to lose control of his mental faculties.[43]

As far as we can tell from these cases, the causal nexus theory Arizona courts have used in considering SED evidence restrictively is similar to the causal nexus theory used to review mitigating evidence of mental illness. Mental illness is generally thought to be mitigating only if it undermines a defendant's control over her actions at the time of the crime.[44] One rationale for this rule is that defendants who lack control or free will when they commit crimes are not culpable, and the defendant's culpability is a critical factor in mitigation. Unfortunately, the opinions in cases like *Ellison* and *Prince* do not explain, in moral and legal terms, why the causal nexus theory is the *only* plausible

[43] The only circumstance in which courts will infer a nexus is if the SED occurred close in time to the murder. The rule is that the mitigating value of SED evidence diminishes as time passes between the deprivation and the murder, entailing that SED seldom serves as a mitigating factor for older defendants. *See, e.g., Prince*, 250 P.3d at 1170 ("Difficult childhood circumstances also receive less weight as more time passes between the defendant's childhood and the offense."); State v. McCray, 183 P.3d 503, 511 (Ariz. 2008) (en banc) ("[A] difficult childhood is given less weight when the defendant is older."); *Ellison*, 140 P.3d at 927–28 ("His childhood troubles deserve little value as a mitigator for the murders he committed at age thirty-three."); State v. Hampton, 140 P.3d 950, 968 (Ariz. 2006) ("Hampton was thirty years old when he committed his crimes, lessening the relevance of his difficult childhood."); State v. McGill, 140 P.3d 930, 944 (Ariz. 2006) ("[T]he impact of McGill's upbringing on his choices has become attenuated during the two decades between his reaching adulthood and committing this murder."); *Anderson*, 111 P.3d at 399 ("Anderson's childhood troubles do not in any way explain his decision, decades later at age forty-eight, to kill three innocent people."). It is clear from *State v. Mann* that the court was not looking for just any psychological connection, because a doctor in the case concluded that the defendant's childhood "directly contributed to Defendant's behavior because he lacked 'healthy socialization experiences.'" 934 P.2d 784, 795 (Ariz. 1997) (en banc).

[44] *See, e.g.*, State v. Hoskins, 14 P.3d 997, 1022 (Ariz. 2000) ("Where we determine questions of aggravation and mitigation in the sentencing process, the significant point in time for causation is the moment at which the criminal acts are committed. If the defendant's personality disorder or dysfunctional family background leads reasonable experts to conclude that the disorder in fact caused the crime, significant mitigation is established.").

explanation of the mitigating potential of either mental illness or SED evidence.[45] We argue in the next section that, at least in the case of SED, this absence of a justification for restrictive consideration is troubling because there appear to be many (and more compelling) alternative explanations for why SED is mitigating.

B. Failure to Justify the Causal Nexus Requirement in Alabama

Alabama courts also frequently give restrictive treatment to SED evidence, on a similar causal nexus theory, though they are perhaps more likely to offer an explanation grounded in individual responsibility. No one who suffers from SED is *determined* to commit murder, they emphasize.

For instance, in *Philips v. State*, the trial court rejected the mitigating value of the repeated violence and neglect suffered by the defendant during his childhood on the basis that it did not directly cause the criminal act.[46] It went on to observe: "[t]his Court has heard hundreds if not thousands of cases of drug abuse, neglect, and domestic violence over the last 20 years, but Capital Murder does not naturally result . . . from a bad childhood."[47] The Court of Criminal Appeals affirmed.[48] Similarly, in *Stanley v. State*,

[45] For instance, the court in *Prince* stated without explanation that "[d]ifficult childhood circumstances . . . receive less weight as more time passes between the defendant's childhood and the offense." *Prince*, 250 P.3d at 1170.

[46] Phillips v. State, No. CR–12–0197, 2015 WL 9263812, at *83–85 (Ala. Crim. App. Dec. 18, 2015).

[47] *Id*. at *81

[48] *Id.* at *85.

the trial court dismissed evidence of a difficult family background as "not mitigating" because the defendant did not offer any "credible evidence that any of these factors influenced the commission of the crime."[49] The trial court emphasized the fact that the defendant's sisters had suffered the same deprivation but did not become criminals.[50] Even more explicitly, the court in *Thompson v. Alabama* observed that:

> [T]he necessity for every person being morally responsible for his or her own actions causes these environmental factors which are offered as mitigation to appear weak . . . . The argument that when a bad social environment produces bad people, that fact should in some way mitigate the punishment for these bad people, leads ultimately to the absurd conclusion that only people who come from an impeccable social background deserve the death penalty if they commit capital murder.[51]

The court here also appeared to be laboring under the misimpression that a disadvantaged background is an *automatic* grant of leniency, rather than one sentencing factor to be considered among many.

In the end these courts also treat SED's mitigating potential on this theory as an all-or-nothing affair: either the extreme deprivation suffered makes it impossible for the defendant to choose not to commit a criminal act (perhaps because of a temporary mental inability at the moment of the crime), in which case SED is mitigating; or the deprivation

[49] Stanley v. State, 143 So.3d 230, 330–32 (Ala. Crim. App. 2011).

[50] *Id*. at 331.

[51] Thompson v. Alabama, No. CR–05–0073, 2012 WL 520873, at *85 (Ala. Crim. App. 2012).

could be overcome, in which case it is assigned *no* mitigating value. Of course, however, most of the effects of SED *can* be overcome, so in effect this reasoning renders SED non-mitigating unless it results in an effect—like a mental illness—that is generally thought to be less subject to the individual's control. In short, Alabama courts, too, engage in restrictive consideration of SED.

Alabama courts see no deficiency in their consideration. The appellate court that reviewed *Stanley*, mentioned above, affirmed, arguing that the sentencing judge adequately "considered all the evidence offered . . . including [the defendant's] family circumstances [and] background."[52] It rejected "Stanley's argument . . . that a trial court's failure to *find* a mitigating circumstance based on certain mitigating evidence necessarily means that the trial court did not *consider* that mitigating evidence."[53] Similar cases abound.[54]

In Alabama, this restrictive consideration of SED has significant consequences not only for the weighing of evidence at sentencing but in other areas of criminal law:

---

[52] *Id.* at 332.

[53] *Id.* at 331.

[54] *See, e.g.*, Davis v. Allen, No. CV 07-S-518-E, 2016 WL 3014784, at *50–51 (N.D. Ala. May 26, 2016) (rejecting defendant's argument that lower court's "failure to give appropriate weight to the evidence of Davis's childhood abuse because it occurred years earlier than the crime" was unconstitutional); Thompson v. State, 153 So.3d 84, 189 (Ala. Crim. App. 2012) ("While *Lockett* and its progeny require consideration of all evidence submitted as mitigation, whether the evidence is actually [mitigating] is in the discretion of the sentencing authority") (quoting *Ex parte* Slaton, 680 So. 2d 909, 924 (Ala. 1996)); Waldrop v. State, 987 So.2d 1186, 1202 n.6 (Ala. Crim. App. 2007) (observing that defendant's counsel did not err in declining to present available SED evidence at sentencing, because the sentencing judge had "evidence of [a co-defendant's] abusive childhood and stated in his sentencing order that he afforded it little weight" because "[i]t would be ironic for the courts to determine that environmental factors which cause people to become violent offenders should then be taken into consideration to make these people less susceptible to the death penalty").

judges reject ineffective assistance of counsel claims that are based on counsel's failure to present SED evidence when the defendant cannot show a causal connection between the SED suffered and the crime[55]; and judges reject claims that juries were biased by prosecutorial suggestion that the SED evidence has no mitigating weight because of lack of a causal connection.[56]

C. Other Appearances of the Causal Nexus Requirement

Judges have appealed to the lack of a direct "causal connection" between SED and criminal conduct to justify giving SED no weight in a number of other jurisdictions. For instance, in a relatively recent Florida case, a trial court found that the defendant was emotionally and physically abused as a child, and yet gave those factors "little weight" because "there was no connection between Petitioner's alleged childhood emotional and physical abuse . . . and the murders."[57] The causal connection sought was, once again, a specific one; generally impaired moral and intellectual capacities due to extreme deprivation did not suffice. The Eleventh Circuit affirmed the trial court's treatment of the evidence, arguing that "[a]s long as the defense is allowed to present all relevant mitigating evidence and the sentencer is given the opportunity to *consider* it, there is no

[55] *See, e.g.*, Jenkins v. Allen, No. 4:08-cv-00869-VEH, 2016 WL 4540920, at *41 (N.D. Ala. Aug. 31, 2016) ("Any contention that a causal connection exists between the abuse allegedly suffered by Jenkins and the murder of Tammy Hogeland, is undercut by evidence within Jenkins's own family.").

[56] *See, e.g.*, Scheuing v. State, 161 So. 3d 245, 267-68 (Ala. Crim. App. 2013).

[57] Lynch v. Sec'y, Dep't of Corr., 897 F. Supp. 2d 1277, 1299 (M.D. Fla. 2012).

constitutional violation."[58] In another case from Florida, the state supreme court suggested that the defendant's childhood abuse could not have reduced his moral responsibility because "the defendant's sister, who had also been abused, including sexually abused by the same alcoholic father, proceeded to live a normal and productive life."[59] As before, SED's mitigating value was seen to turn on whether it rendered virtually impossible the defendant's ability to conform his conduct to the law. Because it is difficult to show that SED has any such effect, it is routinely dismissed when no direct causal connection between it and the crime is found.[60]

Courts do sometimes dismiss proffered SED evidence on factual grounds. If the record does not show that the defendant experienced truly *severe* deprivation or if it reveals that the defendant was rescued from his unenviable circumstances fairly quickly and led a relatively normal adult life after a short period of deprivation, judges reasonably

---

[58] *Id*. at 1339. *See also* Waldrop v. State, 987 So.2d 1186, 1202 n.6 (Ala. Crim. App. 2007) (observing that defendant's counsel did not err in declining to present available SED evidence at sentencing, because the sentencing judge had "evidence of [a co-defendant's] abusive childhood and stated in his sentencing order that he afforded it little weight" because "[i]t would be ironic for the courts to determine that environmental factors which cause people to become violent offenders should then be taken into consideration to make these people less susceptible to the death penalty").

[59] Douglas v. State, 878 So.2d 1246, 1260 (Fla. 2004).

[60] *See, e.g.*, Callahan v. Campbell, 427 F.3d 897, 923 (11th Cir. 2005) (rejecting IAC claim for failing to investigate mental health and abuse, noting that "no causal connection between the alleged abuse Callahan suffered as a child and the crime he committed, which were separated by 23 years"); Davis v. Scott, 51 F.3d 457, 461-62 (5th Cir. 1995), *overruled in part by* Tennard v. Dretke, 542 U.S. 274 (2004) (evidence of child abuse, alone, without demonstrating any link to the crime, does not constitute "constitutionally relevant" mitigating evidence); Madden v. Collins, 18 F.3d 304, 308 (5th Cir. 1994) (evidence of troubled childhood not constitutionally relevant mitigating evidence when not linked in any way to the crime); Barnard v. Collins, 958 F.2d 634, 638-39 (5th Cir. 1992) (rejecting a *Penry* claim where the crime was not attributable to the proffered evidence of troubled childhood); Hines v. State, 856 N.E.2d 1275, 1283 (Ind. Ct. App. 2006) ("the trial court was not obliged to afford any weight to [the defendant's] childhood history as a mitigating factor in that [he] never established why his past victimization led to his current behavior.").

find that the alleged SED remains unproven.[61] We have no quarrel with this practice. Our concern is exclusively with the narrow scope of the mitigating analysis once it is recognized that the defendant did in fact suffer from especially severe neglect, abuse, and/or poverty.

II. Unrestricted Consideration of Deprivation Evidence: Theory and Practice

Recent work in moral philosophy and psychology indicates a renewed interest in the reasons why severe environmental deprivation mitigates the punishment a defendant deserves. We briefly review some of this work, much of which forms a key part of the literature on retributive justice, to show that causal analysis plays a limited-to-non-existent role in prominent theories of SED's mitigating force. In addition, we try to show that such theories that support unrestricted consideration of SED are widely embraced, including by a great many judges and jurors.[62] Both the intuitiveness of such theories and

[61] *See, e.g.*, State v. Kuhs, 224 P.3d 192, 204 (Ariz. 2010) (defendant grew up in a poverty and was abused at least once); State v. Kiles, 213 P.3d 174, 191 (Ariz. 2009) (mixed evidence, because some witnesses testified that Kiles's family life as "ordinary"); State v. Dann, 207 P.3d 604, 628 (Ariz. 2009) (no evidence of child abuse other than spankings with a belt that his father later viewed as child abuse). We emphasize, throughout this article, that the environmental deprivation we reference is of an especially severe sort. The effects of SED we discuss may or may not be fairly inferable from milder forms of deprivation.

[62] *See, e.g.*, Williams v. Taylor, 529 U.S. 362, 370-71 (2000) (concluding that the defendant's attorney had fallen "below the range expected of reasonable, professional competent assistance of counsel" for failing to investigate and present at his sentencing trial "documents prepared in connection with Williams' commitment when he was 11 years old that dramatically described mistreatment, abuse, and neglect during his early childhood" and "repeated head injuries"—evidence the Court described as "significant" mitigating evidence).

their wide appeal will feature critically in the constitutional arguments we go on to offer in Part III.

A. The Defendant's Diminished Moral Capacities & Culpability

It is not just a scientific platitude but a matter of common sense that the development of key behavioral capacities is critical to pro-social decision-making.[63] These include emotional capacities, like the capacity to empathize with others or to form human attachments, and capacities for self-regulation, including impulse control and anger management. Still others involve basic executive brain function, such as working memory and the capacity to think through the consequences of one's actions.[64]

The development of these capacities, critical as they are to the process of becoming morally mature, is impaired by severe emotional, psychological, and sexual abuse.[65] The psychological evidence is extensive, and often presented at trial by

---

[63] Moral capacities are generally seen as being influenced by all three of these elements. *See* THOMAS KEENAN & SUBHADRA EVANS, AN INTRODUCTION TO CHILD DEVELOPMENT 297-98 (2009).

[64] *See* Tina Malti & Sophia F. Ongley, *On Moral Reasoning and Relationship with Moral Emotions*, *in* HANDBOOK OF MORAL DEVELOPMENT RESEARCH 166-69, 171-72 (Melanie Killen & Judith G. Smetana eds., 2d ed. 2014) (reviewing the relationship between moral emotions and moral reasoning, and the connection between empathy/sympathy and higher levels of other-oriented moral reasoning and prosocial moral reasoning); Roy F. Baumeister & Julie Juola Exline, *Self-Control, Morality, and Human Strength*, 19 J. SOC. & CLINICAL PSYCH. 29 (2000) ("Self-control refers to the self's ability to alter its own states and responses, and hence it is both key to adaptive success and central to virtuous behavior, especially insofar as the latter requires conforming to socially desirable standards instead of pursuing selfish goals.").

[65] Darcia Narvaez & Daniel Lapsley, *Becoming a Moral Person—Moral Development and Moral Character Education as a Result of Social Interactions*, *in* EMPIRICALLY INFORMED ETHICS: MORALITY BETWEEN FACTS AND NORMS 227 (Markus Christen et. al eds., 2014).

experienced defense counsel in the form of expert testimony. Childhood abuse or neglect is associated with decreased levels of empathy and altruism, and increased levels of aggression and antisocial behaviors, well into adulthood.[66] Extreme poverty, too, is significantly correlated with increased levels of depression, low self-esteem, and diminished impulse control in children.[67] Darcia Narvaez and Daniel Lapsley explain that children who have been subject to regular threats, violence, and deprivation are more likely to develop a "survival-first" mindset—a persistent physical and mental state of "high alert"—that "subverts the more relaxed states that are required for positive prosocial emotions and sophisticated reasoning."[68] When a child's own caregivers are the source of threats and deprivation, the child can miss crucial opportunities to develop interpersonal trust and receive affection from others. These "disruptions and deviations in socialization" can seriously undermine later attempts to form relationships in adolescence and adulthood, and are linked to subsequent emotional and behavioral problems among

---

[66] *Id.* at 228. *See also* Joanna Cahall Young & Cathy Spatz Widom, *Long-term Effects of Child Abuse and Neglect on Emotion Processing in Adulthood*, 38 CHILD ABUSE NEGLECT 1369 (2014) (the "effects of childhood abuse/neglect on emotion processing extend until middle adulthood" though it would be worthwhile to have multiple assessments over time); Michael D. De Bellis & Abigail Zisk, *The Biological Effects of Childhood Trauma*, 23 CHILD ADOLESCE. PSYCHIATR. CLIN. N. AN. 2014 185 (2014) ("the data to date strongly suggests that childhood trauma is associated with adverse brain development in multiple brain regions that negatively impact emotional and behavioral regulation, motivation, and cognitive function"); Anthony Nazarov et al., *Moral Reasoning in Women with Post-Traumatic Stress Disorder Related to Childhood Abuse*, 7 EUR. J. PSYCHO-TRAUMATOLOGY 2016 (altruism); Paul A. Miller & Nancy Eisenberg, *The Relationship of Empathy to Aggressive and Externalizing/Antisocial Behavior*, 103 PSYCH. BULLETIN 324 (1988) (Childhood abuse is associated with low levels of empathy/sympathy, which are in turn associated with aggression and antisocial, externalizing behaviors).

[67] David T. Takeuchi et al., *Economic Distress in the Family and Children's Emotional and Behavioral Problems*, 53 J. MARRIAGE & FAM. 1031, 1037-39 (1991) (reporting that economic stress significantly impacts children's emotional and behavioral problems, often resulting in higher levels of depression, antisocial behavior, and diminished impulse control).

[68] Narvaez & Lapsley, *supra* note 65, at 228-29.

abused children.[69] Studies also show that these factors more generally limit the development of basic brain functions, including planning skills, inhibitory control, working memory, cognitive focus, and reward processing.[70] The younger the child is at the time of the severe abuse, and the more sustained the deprivation, the worse and more long-lasting are the cognitive, emotional, and behavioral effects.[71] Each one of these developmental deficits is individually linked to physical abuse, sexual abuse, and extreme neglect in childhood, and many capital defendants have experienced more than one of these deprivations.[72]

---

[69] DAVID A. WOLFE, CHILD ABUSE: IMPLICATIONS FOR CHILD DEVELOPMENT AND PSYCHOPATHOLOGY 35-36 (2d ed. 1999).

[70] *See* Nicolas Berthelot et al., *Childhood Abuse and Neglect May Induce Deficits in Cognitive Precursors of Psychosis in High-Risk Children*, 40 J. PSYCHIATRY NEUROSCIENCE 336 (2015) (finding much lower IQ and poorer cognitive performance in visual episodic memory and in executive functions of initiation); DeBellis et al., *Neuropsychological Findings in Pediatric Maltreatment: Relationship of PTSD, Dissociative Symptoms, and Abuse/Neglect Indices to Neurocognitive Outcomes*, 18 CHILD MALTREATMENT 171 (2013) (maltreated persons performed significantly lower on IQ, academic achievement, and nearly all of the tested neurocognitive domains); Kathryn L. Hildyard & David A. Wolfe, *Child Neglect: Developmental Issues and Outcomes*, 26 CHILD ABUSE & NEGLECT 679 (2002) (even neglect alone can have "more severe cognitive and academic deficits, social withdrawal and limited peer interactions, and internalizing [] problems" than physically abused peers); Roy F. Baumeister & Julie Juola Exline, *Self-Control, Morality, and Human Strength*, 19 J. SOC. & CLINICAL PSYCH. 29 (2000) (neglected children can have difficulties predicting the consequences of their behavior); William B. Harvey, *Homicide Among Black Adults: Life in the Subculture of Exasperation*, *in* HOMICIDE AMONG BLACK AMERICANS 153 (Damell F. Hawkins ed., 1986) (describing how numerous social pressures, including a pervasive sense of hopelessness, contribute to high crime rates among impoverished African American communities within the inner city).

[71] Raquel A. Cowell et al., *Childhood Maltreatment and Its Effect on Neurocognitive Functioning: Timing and Chronicity Matter*, 27 DEV. PSYCHOPATHOLOGY 521 (2015) (children who suffered maltreatment as infants or chronically had higher deficits in working memory and inhibitory control); *see also* Hildyard & Wolfe, supra note 70, at 679.

[72] Gwendolyn M. Lawson et al., *Socioeconomic Status and Neurocognitive Development: Executive Function*, *in* EXECUTIVE FUNCTION IN PRESCHOOL AGE CHILDREN: INTEGRATING MEASUREMENT, NEURODEVELOPMENT AND TRANSLATIONAL RESEARCH (J.A. Griffin et al. eds., 2016); Kimberly G. Noble et al., S*ocioeconomic Gradients Predict Individual Differences in Neurocognitive Abilities*, 10 DEVELOPMENTAL SCI. 464 (2007).

Adults with histories of childhood deprivation and maltreatment are almost twice as likely to have been incarcerated than those without such histories, and significantly more likely to have been arrested for a violent crime.[73]

These facts about the link between childhood deprivation and psychological development are close to common knowledge in the judicial system. As Justice Rehnquist observed in *Santosky v. Kramer*, "[a] stable, loving homelife is essential to a child's physical, emotional, and spiritual well-being."[74] Judges also routinely take "judicial notice" of the fact that extreme neglect and sexual abuse "increases the probability of [maladjustment and mental] problems."[75]

Poverty, under-education, and immersion in a culture of violence similarly distort a person's moral compass even later in life. A number of theorists have argued that chronic stressors and high levels of psychological distress due to consistent economic deprivation severely erode "self-esteem and the sense of mastery, control, and personal efficacy."[76]

---

[73] Hyunzee Jung et al., *Does Child Maltreatment Predict Adult Crime? Reexamining the Question in a Prospective Study of Gender Differences, Education, and Marital Status*, 30 J. INTERPERSONAL VIOLENCE 2248 (2015); Izabela Milaniak & Cathy Spatz Widom, *Does Abuse and Neglect Increase Risk for Perpetration of Violence Inside and Outside the Home?*, 5 PHYSICAL VIOLENCE 246, 250 (2015); Phyllis L. Crocker, *Childhood Abuse and Adult Murder: Implications for the Death Penalty*, 77 N.C. L. REV. 1143, 1154 (1998) (noting the "strong evidence . . . that a person who was abused as a child is at risk of suffering long-term effects that may contribute to his violent behavior as an adult").

[74] 455 U.S. 745, 788-9 (1982) (Rehnquist, J., dissenting).

[75] Bouchillon v. Collins, 907 F.2d 589, 590 n.2 (5th Cir. 1990). *See also* Russell v. Collins, 998 F.2d 1287, 1292 (5th Cir. 1993) (acknowledging that child abuse as "generally understood" would "have the tendency to affect the child's moral capacity by predisposing him or her toward committing violence").

[76] Mary Keegan Eamon, *The Effects of Poverty on Children's Socioemotional Development: An Ecological Systems Analysis*, 46 SOCIAL WORK 257, 258 (2001); *see also id.* (citing psychological

What is the mitigating upshot of the fact that SED causes such general impairment in the development of critical moral and behavioral capacities? Courts who engage in the restrictive consideration of SED assume that deprivations can only be mitigating if they entirely undercut the defendant's ability to conform to the law. Accordingly, judges look for evidence that the SED directly and specifically caused the criminal act. Interestingly, a similar view informed a seminal article by Judge David Bazelon in the 1970s that was highly *sympathetic* towards SED sufferers. Judge Bazelon likened "mental impairments associated with social, economic, and cultural deprivation" to mental diseases that undermine the defendant's free will, and argued that such deprivation provides grounds for *excusing* the defendant.[77] Courts reasonably resisted such arguments, sometimes pointing to socially well-adjusted siblings of capital defendants, like the Alabama courts cited above.[78] Indeed, none of the studies we have come across suggest that extreme deprivation *destines* persons to lead criminal or immoral lives—which it obviously does not.

---

research on the impact of poverty on moral development); Richard Lipke, *Social Deprivation as Tempting Fate*, 5 CRIM. L. & PHIL. 277, 283-84 (2011) (contending social deprivation reduces the incentives for self-control and may work to stunt its development, thereby reducing the culpability of the defendant).

[77] David L. Bazelon, *The Morality of the Criminal Law*, 49 S. CAL. L. REV. 385, 394 (1976). *See also* Richard Delgado, *"Rotten Social Background" Should the Criminal Law Recognize a Defense of Severe Environmental Deprivation?* 3 L. & INEQUALITY. 9, 23-34 (1985) (arguing that, in some cases, a propensity toward crime arising from deprivation is so strong as to render the individual not responsible for their crimes).

[78] *See, e.g.*, Douglas v. State, 878 So.2d 1246, 1260 (Fla. 2004) (noting the diminished mitigating value of SED evidence where "the defendant's sister, who had also been abused, including sexually abused by the same alcoholic father, proceeded to live a normal and productive life"). Prosecutors often also present such evidence to persuade courts. *See, e.g.*, State v. Hester, 324 S.W.3d 1, 84 (Tenn. 2010) ("The State also presented evidence that Mr. Hester's other siblings, including a sister who had been sexually abused by her father, had managed to grow up in the same house with the same parents without having become killers.").

But the sentencing question is not whether the defendant should be altogether excused. The question is whether he deserves to be held *fully* responsible and maximally punished. Accordingly, while the search for a causal nexus seems sensible in the context of evaluating questions of guilt and excuse at the trial stage, it is far from adequate in the context of mitigation once the defendant has already been convicted. Thus, modern theorists of SED's moral significance for punishment are less inclined to treat it as an excuse, and instead regard it in terms of the intuitive notion that moral responsibility comes in degrees.[79] Even a person who *could* have chosen to lead a law-abiding life, and is therefore culpable for his wrongful choices, can, by virtue of the extreme challenges he faced in achieving moral maturity, be *less than fully responsible* and/or deserving of less than maximal punishment.

Arguably the most well-known and influential contemporary moral philosopher, Thomas Scanlon, articulates the moral intuitions underlying this theory of "diminished responsibility" as follows. He argues that a wrongdoer's liability for punishment depends on the adequacy of his "opportunity to avoid" committing the wrongful act and thus suffering the associated punishment. A person's opportunity to avoid making a certain choice "depends on the conditions under which the choice is made: the quality of information that the person has, the absence of competing pressures, the attractiveness of the available alternatives, and so on."[80] In his discussion of a wealthy individual who compares himself to one living in poverty, Scanlon contends that the wealthy person's

---

[79] *See* D. Justin Coates & Philip Swenson, *Reasons-Responsiveness and Degrees of Responsibility*, 165 PHIL. STUDIES 629 (2013).

[80] THOMAS M. SCANLON, MORAL DIMENSIONS: PERMISSIBILITY, MEANING, BLAME 204-05 (2010).

claim that he "chose" to use his opportunities better than the impoverished person is "weakened by our supposition that the conditions under which the poor man chose—and might have chosen differently—did not provide him with adequate opportunity [to achieve the same results]."[81] Note that "inadequate opportunity" is not equivalent to "no opportunity." The diminished opportunities that SED sufferers have for cultivating their moral capacities and avoiding punishment under the law, accordingly, limits the extent to which we can hold such persons responsible for their actions.[82]

Judges often appeal to the idea that moral responsibility and culpability come in degrees. Justice O'Connor opined, concurring in *California v. Brown*, that "evidence about the defendant's background and character is relevant [in mitigation] because of the belief, long held by this society, that defendants who commit criminal acts that are attributable to a disadvantaged background . . . may be *less culpable* than defendants who have no such excuse."[83] Writing for the majority a couple of years later in *Penry v. Lynaugh*, O'Connor confirmed that, "[b]ecause Penry was mentally retarded . . . and because of his history of childhood abuse," a rational juror "could conclude that Penry was less morally culpable than defendants who have no such excuse."[84]

---

[81] *Id.*

[82] MANUEL VARGAS, BUILDING BETTER BEINGS: A THEORY OR MORAL RESPONSIBILITY 245 (2013) (arguing that the "moral ecology" in which a person comes to make his choices—including whether or not he has been "trained up" with the resources to respond to moral considerations in the way we see fit—is relevant to whether or not that person can be thought to be a responsible agent); Lipke, *supra* note 76, at 287 (contending social deprivation reduces the incentives for self-control and may work to stunt its development, thereby reducing the culpability of the defendant).

[83] 479 U.S. 538, 545 (1987) (O'Connor, J., concurring) (emphasis added).

[84] 492 U.S. 302, 322-23 (1989), *abrogated on other grounds by* Atkins v. Virginia, 536 U.S. 304 (2002).

Note that on the diminished responsibility theory we are expounding, extreme deprivation's mitigating weight does not turn on any proof of immediate or specific causation of any *particular* crime. It turns on the fact, inferable from established SED evidence, that the deprivation impaired the defendant's capacities, which made it *generally* harder for him to live a law-abiding and decent life.

Many "death-eligible" jurors—that is, jurors who are not in principle opposed to the death penalty—are sympathetic to this theory and are less likely to vote for death because of it. Using data from the Capital Juror Project, Stephen Garvey finds that of 153 capital jurors interviewed who were presented with evidence of extreme poverty and "circumstances over which the defendant had no control [but] that may have helped form (or misform) his character," roughly 32% were less likely to sentence the defendant to death.[85] If a third of a capital jury refused to issue a death sentence, in a state where juries rather than judges control the ultimate sentence, the result would be a life sentence.

B. The Defendant's Suffering & the "Whole Life" View of Retributive Justice

Even if the defendant emerged from childhood trauma with critical behavioral capacities largely intact, the suffering inherent in experiencing severe deprivation can be directly relevant in mitigation. Physical, emotional, and sexual abuse combined with

[85] Stephen P. Garvey, *Aggravation and Mitigation in Capital Cases: What Do Jurors Think?* 98 COLUM. L. REV. 1538, 1565 (1998).

extreme poverty in childhood almost always means not just great physical and psychological pain at the time of the deprivations, but harmful ripple effects throughout a person's life. As Craig Haney observes, capital defendants have often "confronted chronic poverty, extraordinary instability, and, for some, almost unimaginably brutal and destructive mistreatment over which, for most of their lives, they have been granted little or no control."[86] On "whole life" views of retributive justice, such facts about the overall suffering experienced by a person over the course of his life are intrinsically relevant to what punishment the person deserves when he acts wrongfully.

Traditional retributive theories of punishment took a very restricted view of the times relevant to deciding what a wrongdoer deserves. The key animating principle behind such theories was, roughly, that the suffering a wrongdoer inflicts on others *must* be matched by his equivalent suffering in the future, regardless of what had already happened to him in the past: "[t]hose who perform specific criminal acts deserve specific punishments . . . largely independently of their acts or happiness at other times."[87] An eye can be taken for an eye, even if the wrongdoer already lost an eye a long time ago.

By contrast, on what is now called the "whole life approach" or the "life-cycle" view of retributive justice, what wrongdoers deserve cannot be decided without considering previous suffering and unhappiness. As Shelly Kagan, one of the leading proponents of this view, observes, "time drops out from further consideration: we look at lives *as a whole*, to see what one deserves (overall), and whether one has received it

---

[86] Craig Haney, *Violence and the Capital Jury: Mechanisms of Moral Disengagement and the Impulse to Condemn to Death*, 49 STANFORD L. REV. 1447, 1565 (1997).

[87] Thomas Hurka, *Desert: Individualistic and Holistic*, *in* SERENA OLSARETTI, DESERT AND JUSTICE 45, 52 (2003) (describing the view that he critiques).

(overall)."[88] According to such theorists, the relevant question that the sentencer should be asking in capital cases is whether the defendant, in light of his criminal conduct *and* all of the suffering he has so far endured in his life, deserves so much additional suffering that he should be executed. The sentencer should treat the defendant as substantially less deserving of the harshest and ultimate sentence if the defendant has already experienced incredible suffering in life, as SED sufferers undoubtedly have.[89]

One way of motivating this picture is by appeal to an intuitive principle (a kind of side-constraint on punishment): there is a limit to the amount of suffering we should expect any one person to bear in a lifetime. The need to ensure that no one suffers beyond tolerable levels militates against the execution of SED sufferers—those who have already suffered enough in life. The fact that the suffering happened in the past does not make it any less bad for the person. Defense attorneys routinely appeal to such considerations and judges give voice to them as well. The Court in *Eddings*, for instance, observed that the defendant's terrible family background was relevant to the sentencing decision, because

---

[88] SHELLY KAGAN, THE GEOMETRY OF DESERT 11 (2012) (emphasis added); *see also* DAVID ROSS, THE RIGHT AND THE GOOD 58 (2d ed. 2003).

[89] *See, e.g.*, Knight v. Dugger, 863 F.2d 705 app. at 749 (11th Cir. 1988) (describing the defendant's "impoverished home" as abusive and lacking supervision); Mathis v. Zant, 704 F. Supp. 1062, 1065 (N.D. Ga. 1989) (noting that the defendant was repeatedly verbally abused by his chronically alcoholic father, missed school one-third of the time, was ridiculed because he was slow, and dropped out in fifth grade; thereafter, he spent most of his time in prisons), *vacated and remanded*, 975 F.2d 1493 (11th Cir. 1992); Phillips v. State, 608 So. 2d 778, 782 (Fla. 1992) (stating that the defendant grew up in poverty and his parents were migrant workers "who often left the children unsupervised"); State v. Murphy, 605 N.E. 2d 884, 909 (Ohio 1992) (Moyer, C.J., dissenting) (stating that trial testimony established that the defendant was raised in "desperate poverty"; had an "unloving, unsupportive, and abusive family"; lived in a home described as a shack with no hot water or plumbing; lived on public assistance; and had a father who was an alcoholic).

of its "potential for *evoking sympathy*" for the defendant.[90] Arguably, the reason why such facts of deprivation evoke sympathy is that we recognize a duty to help those who have suffered too much in life. One way in which we help is by exercising mercy in sentencing.

As before, the whole life view favors looking beyond the causal nexus theory when considering SED. It regards SED as mitigating with no causal analysis. The morally relevant question is simply: how severe and injurious to the defendant was the deprivation suffered? The whole life view of SED's mitigating significance explains why jurors treat the factor as significantly mitigating on its own, without *any* causal connection to the crime or the defendant's capacities. In a study of juror receptivity to mitigation evidence based on 400 mock jurors, Mona Lynch and Craig Haney observed that childhood abuse history and bad family background were regularly treated as significant in mitigation without any indication of its relationship to the crime or the defendant's later life.[91]

C. The Diminished Societal Standing to Punish

We consider one final alternative theory of SED's mitigating value before turning to the constitutional argument. As before, the focus is not so much on proving that these

[90] California v. Brown, 479 U.S. 538, 548 (O'Connor, J., concurring) (emphasis added).

[91] Mona Lynch & Craig Haney, *Discrimination and Instructional Comprehension: Guided Discretion, Racial Bias, and the Death Penalty*, 24 L. & HUMAN BEHAVIOR 337 (2000).

theories are correct from the moral point of view but, rather, on making vivid their plausibility and the unreasonableness of restrictive consideration based on the causal nexus theory alone.

A number of theorists have articulated SED's moral significance for criminal justice in terms of the state's "standing to punish." Such theorists take for granted that society has an obligation to provide a minimally decent quality of life for all of its citizens.[92] What constitutes a minimally decent quality of life is disputed, but it is generally agreed that it involves safety from physical abuse and access to basic necessities, including food, clothing, and shelter.[93] Accordingly, these theorists argue that our failure to mitigate extreme poverty and its effects diminishes our standing to punish those who have suffered from extreme poverty to the maximum extent allowable by retributive principles.[94]

An individual can lose standing—or moral authority—to *hold* another person wholly responsible for a wrongful act, even if the wrongdoer bears full moral responsibility for the act. This happens when the individual himself has "unclean hands" with respect to the act. One source of society's unclean hands when it comes to criminals

---

[92] *See, e.g.*, Philippe Van Parijs, *Why Surfers Should be Fed: the Liberal Case for an Unconditional Basic Income*, 20 PHIL. & PUB. AFFAIRS 101 (1991); *see also* Emad H. Atiq, *How Folk Beliefs About Free Will Influence Sentencing: A New Target for the Neuro-Determinist Critics of Criminal Law*, 16 NEW CRIM. L. REV. 449 (2013); Daniel Markovitz, *How Much Redistribution Should There Be?*, 112 YALE L.J. 2291 (2003).

[93] *See, e.g.*, HENRY SHUE, BASIC RIGHTS: SUBSISTENCE, AFFLUENCE, AND U.S. FOREIGN POLICY (1996); Erwin Chemerinsky, *Making the Case for a Constitutional Right to Minimum Entitlements*, 44 MERCER L. REV. 525, 529, 531 (1993) (food, shelter, medical care, housing).

[94] For a discussion of this principle, *see* THOMAS M. SCANLON, WHAT WE OWE TO EACH OTHER 256-67 (2000). *See also* Atiq, *supra* note 92; Jeffrie G. Murphy, *Marxism and Retribution*, 2 PHIL. & PUB. AFFAIRS 317 (1973).

is its moral failure to ensure an adequate safety net that protects everyone from severe environmental deprivations. As Victor Tadros writes, "[b]y perpetrating distributive injustice against the poor, we lose standing to hold them responsible for what they have done."[95] Another reason for the collective's "unclean hands" concerns the collective's complicity in the wrongdoer's conduct. Tadros observes:

> There are different explanations of how our standing to hold others responsible may be eroded but two are most important, One is grounded in hypocrisy: the fact that one person commits the same kinds of wrong as someone else deprives the one of standing to hold the other person responsible for his wrongs. The other [reason] is complicity: the fact that one person participates in the wrong of someone else deprives the one of standing to hold the other person responsible for the wrong. A person cannot act as judge when he ought to be a co-defendant.[96]

Tadros views the collective as complicit in the crimes of SED sufferers because we know—or at least ought to know—that extremely poor socioeconomic conditions result in crime, and that we have an obligation to alleviate those conditions. Yet we deliberately choose to invest our resources in causes other than poverty relief, even at the cost of higher crime rates. By so choosing, we are complicit in each crime that we could have prevented had we helped the worst-off. As Tadros put it, "distributive injustice is criminogenic. In perpetrating distributive injustice, the state shows itself to have

---

[95] Victor Tadros, *Poverty and Criminal Responsibility*, 43 J. VALUE INQUIRY 391, 393 (2009).

[96] *Id.* at 394. *See also* G.A. Cohen, *Casting the First Stone: Who Can, and Can't, Condemn the Terrorists?*, 81 ROYAL INST. OF PHIL. SUPPS. (2006).

insufficient concern for the victims of crime."[97] Such rationales for limiting how much we punish SED sufferers may be esoteric, but their logic is compelling.

Judge Bazelon echoes a similar sentiment:

> [I]t is simply unjust to place people in dehumanizing social conditions, to do nothing about those conditions, and then to command those who suffer, 'Behave—or else!' The overwhelming majority of violent street crime, which worries us so deeply, is committed by people at the bottom of the socioeconomic-cultural ladder . . . . We cannot produce a class of desperate and angry citizens by closing off, for many years, all means of economic advancement and personal fulfillment for a sizeable part of the population, and thereafter expect a crime-free society.[98]

Bazelon argues that our "unclean hands" are driven not just by our complicity in the criminal wrongdoing (given its predictability) but also our failure to give the wrongdoer his due: an adequate social safety net.

How is the collective's diminished standing to punish relevant in mitigation? Showing mercy at sentencing is one way of recognizing the collective's diminished standing to punish. The reasons for exercising mercy, again, do not turn on the causal connectedness between the deprivation suffered and the crime. While the standing view

---

[97] For a defense of the two premises that poverty is criminogenic and that the collective has a responsibility to alleviate criminogenic conditions, *see* Tadros, *supra* note 95 (arguing that "the state [is] complicit in the crimes of the poor" and thus the poor have a moral claim "for the state to refrain from holding them responsible for their crimes, even if they are in fact responsible for them, which involves diminished blame").

[98] Bazelon, *supra* note 77, at 401-02.

is less obviously embraced by jurors, it is a common strategy of defense counsel to portray the defendant as a "victim" of societal ills. We have found at least one attorney and psychologist, Deena Logan, who concludes, based on an analysis of 31 closing arguments at death sentencing trials, that effective characterization of the defendant as a victim by appeal to his poverty, diminished mental capacity, and deprived social background elicits mercy from juries.[99]

III. The Constitutional Argument Against Restrictive Consideration of Deprivation Evidence

In this section, we argue that restrictive consideration of SED evidence warrants constitutional scrutiny. While it is often assumed that judges' weighing of mitigating factors is unreviewable, two strands of constitutional doctrine suggest otherwise. The first is found in a long line of cases identifying certain constraints on the "consideration" of mitigating evidence as unconstitutional.[100] The second is evident in the Court's refrain that the death penalty must not be issued unless it enjoys broad-based community approval. Our elaborations of these two lines of precedent, in combination with the evidence discussed in the previous section of the intuitiveness and broad-based appeal of

[99] Deana Logan, *Pleading for Life: An Analysis of Themes in 21 Penalty Arguments by Defense Counsel in Recent Capital Cases*, 4 CAL. DEATH PENALTY DEF. MANUAL 2SN-19 (1982); *see also* Deana Logan, *Why This Man Deserves to Die: Themes Identified in Prosecution Arguments in Recent Capital Cases* (1983) (unpublished manuscript).

[100] Eddings v. Oklahoma, 455 U.S. 104, 115 (1982).

the moral theories on which SED has mitigating weight absent a causal nexus with the crime, offer grounds for scrutinizing and invalidating restrictive consideration of deprivation evidence.

A. "Consideration" Requires a "Reasoned Moral Response," Not Legal Formalism

It is a bedrock principle of Eighth Amendment jurisprudence that a death sentence must be based on "individualized consideration" of any mitigating circumstances.[101] The case establishing the principle, *Lockett v. Ohio*, found that a statute prohibiting capital juries from taking into account any mitigating factors other than three specifically mentioned violated the individualized consideration requirement.[102] We think the holding rests on a more general principle, which we defend below: that individualized *moral* consideration of mitigating factors requires that the sentencer's reasoning not be cabined by artificial *legal* constraints. The Court has spent three decades elaborating what counts as a legal constraint preventing individualized consideration, and SED evidence has played a central role in its elaboration.

Ten years after *Lockett*, the Court prohibited not just statutory limitations on what mitigating factors can be considered, but judge-made rules limiting the conditions under

[101] Lockett v. Ohio, 438 U.S. 586, 606 (1978) (sentencers must "treat each defendant in a capital case with the degree of respect due the uniqueness of the individual").

[102] *Id.* at 593-94.

which a mitigating factor can be considered. In *Eddings v. Oklahoma*, the trial judge ignored evidence offered by the defendant of his youth and turbulent family history, stating that he could not "in following the law" consider such evidence unless it "tended to provide a legal excuse from criminal responsibility."[103] The court of criminal appeals affirmed the resulting death sentence. The Supreme Court expressed some uncertainty as to which *law* the trial judge was referring to. But he seemed to be alluding to a M'Naghten-style test for legal insanity, which gives the defendant a full defense if he lacked the capacity to know "the difference between right and wrong."[104] No Oklahoma statute at the time required sentencers to use the insanity defense standard in evaluating mitigating evidence presented at the penalty phase of a trial.[105] The Supreme Court concluded that, by excluding relevant mitigating evidence from consideration out of a sense—correctly or incorrectly—that the law requires it, the trial court and the highest state court had violated *Lockett*. As the Court explained, a judge has discretion to assign weight to a mitigating factor, but "may not give it no weight by excluding such evidence from their consideration."[106]

The Court further clarified the *Eddings* rule in a later case, *Tennard v. Dretke*, which held that judicial precedent—like a statute or a vague sense of what the law

---

[103] *Eddings*, 455 U.S. at 113.

[104] *Id.* at 109.

[105] *Id.* at 118. While the Oklahoma Supreme Court cited to an earlier decision, *Gonzales v. State*, for the test of criminal responsibility in the state, its use of the test as a means for weighing mitigating evidence was a judicial innovation. Eddings v. State, 616 P.2d 1159, 1170 (1980) (citing Gonzales v. State, 388 P.2d 312 (Okla. Crim. App. 1964)).

[106] *Eddings*, 455 U.S. at 115.

demands—cannot cabin a sentencing agent's "consideration" of mitigating evidence.[107] Again, the case involved SED evidence. *Tennard* reviewed the Fifth Circuit's use of a "constitutional relevance" test in determining whether to grant certificates of appealability for *Penry* claims—defendants' claims that jury instructions at sentencing improperly reduced the effect of their mitigating evidence.[108] The Fifth Circuit's test required that the evidence in question represent a "uniquely severe permanent handicap" that bears a "nexus" to the crime.[109] The Fifth Circuit refused to grant a certificate in Tennard's case on the grounds that his evidence of a low IQ and childhood abuse failed the test. The Supreme Court held that the court of appeals was wrong to condition its review on whether the mitigating evidence met a judge-made legal standard.[110]

Eddings and *Tennard* indicate that judges cannot limit their own moral consideration of relevant mitigating evidence out of a sense that the law—whether statute or judicial precedent—requires it. These cases are a logical application of the *Lockett* holding that capital sentencing requires individualized consideration of mitigating factors. Implicit in these cases is an important general principle that has yet to be fully articulated: that a judge's consideration of relevant mitigating evidence is unconstitutionally narrow when it involves assessing the evidence relative to a limited

---

[107] Tennard v. Dretke, 542 U.S. 274 (2004).

[108] Penry v. Johnson, 532 U.S. 782 (2001) [hereinafter *Penry II*].

[109] *Tennard*, 542 U.S. at 274.

[110] We actually think that *Smith* and *Tennard* are best understood as applications of *Eddings*, though the Court did not discuss them that way. Why didn't the Court come out and explain that more directly? Because the Fifth Circuit was not in the business of weighing mitigating evidence; that task was left for the jury. The court of appeals was merely reviewing whether the SED evidence was relevant in order to decide *whether it should hear the case*.

set of moral principles out of a sense that *legal* rules demand it (where 'it' refers to the limitation on the moral principles by which the evidence is judged). We do not intend to offer an analysis here of what it means to follow a rule or practice *because it is the law*. But it is easy to identify paradigmatic cases of legalism or legal rule following. For example, a judge might follow a rule out of a sense that it is binding precedent or because other judges have an informal convention of following the norm. When restricted consideration of mitigating evidence is the result of judges imposing restrictions legalistically, this violates the principle implicit in the *Lockett* line of cases.

The key to our interpretation is that the individualization principle of *Lockett* has its roots in the distinction between moral reasoning and legal reasoning. In *Lockett*, *Eddings*, *Tennard*, and *Smith*, the Court did not decide in an ad hoc way that particular sorts of legal rules may not constrain the capital sentencer's moral consideration of mitigating evidence. The Court was concerned with eliciting moral consideration from sentencers by removing legal constraints on their ability to consider the evidence from a purely moral point of view. This is why the Court has emphasized time and again that a capital sentence must reflect a "reasoned moral response to the defendant's background, character, and crime."[111] The Court itself has acknowledged that the "reasoned moral response" principle "first originated" in *Lockett* and *Eddings*.[112] As Justice Stevens once

---

[111] *Penry II*, 532 U.S. at 788 (citations omitted).

[112] Brewer v. Quarterman, 550 U.S. 286, 289 (2007) ("[W]e have long recognized that a sentencing jury must be able to give a 'reasoned moral response' to a defendant's mitigating evidence—particularly that evidence which tends to diminish his culpability—when deciding whether to sentence him to death. This principle first originated in *Lockett v. Ohio* and *Eddings v. Oklahoma,* in which we held that sentencing juries in capital cases "must be permitted to consider *any* relevant mitigating factor."); *see also* Abdul-Kabir v. Quarterman, 550 U.S. 233, 264 (2007) ("Our cases following *Lockett* have made clear that when the jury is not permitted to

wrote, "in the final analysis, capital punishment rests on not a legal but an ethical judgment—an assessment of . . . the moral guilt of the defendant."[113] In a precursor case to *Lockett*, the Court explained that capital sentencing requires "particularized consideration of relevant aspects of the record of each convicted defendant" lest defendants be treated as a "faceless, undifferentiated mass to be subjected to the blind infliction of the penalty of death."[114]

In *McKinney v. Ryan*, the Ninth Circuit appears to have implicitly relied on something like this insight in finding unconstitutional the longstanding practice among Arizona judges of considering a defendant's SED to have appreciable mitigating value only if it "caused" his crime.[115] As explained in Part I, judges who engaged in this practice did not view themselves as following binding judicial precedent, as they did when they relied on the old exclusionary rule invalidated by *Tennard*. Rather, they appeared to be following an informal custom amongst judges who had previously applied the exclusionary rule. Judicial customs can, of course, give rise to informal norms and rules that judges follow out of habit or a sense of their legality and the obligations of the judicial office. The practice of giving SED effectively no weight absent a causal nexus

---

give meaningful effect or a 'reasoned moral response' to a defendant's mitigating evidence . . . the sentencing process is fatally flawed.").

[113] Spaziano v. Florida, 465 U.S. 447, 469 (1984) (Stevens, J., concurring in part and dissenting in part). *See also* McCleskey v. Kemp, 481 U.S. 279, 310 (1987) (outlining the capital jury's task of expressing "the conscience of the community on the ultimate question of life or death"); Woodson v. North Carolina, 428 U.S. 280, 297-98 (1976) (reflecting on the importance on the moral views of society in the administration of death penalty); Witherspoon v. Illinois, 391 U.S. 510, 519-20 (1968) ("[A] jury that must choose between life imprisonment and capital punishment can do little more—and must do nothing less—than express the conscience of the community on the ultimate question of life or death.").

[114] *Woodson*, 428 U.S. at 303.

[115] McKinney v. Ryan, 813 F.3d 798 (9th Cir. 2015).

bore all the earmarks of such a judicial custom The court of appeals deemed the practice unconstitutional under *Eddings*. Unfortunately, *McKinney*'s decision remains unnecessarily localized, given the *en banc* court's decision to focus not on the existence of an entrenched judicial practice of restricted consideration but on the practice's historical link to the old exclusionary rule.[116] As previously mentioned, the court emphasized the Arizona Supreme Court's pin citation to the old rule.[117] Because of this choice of emphasis, the Ninth Circuit missed an opportunity to articulate a general test for identifying when the improper influence of a legal practice or custom makes a court's consideration of evidence inadequate under *Eddings.*

We offer a three-factor test for this purpose, drawn from cases—such as those reviewed in Arizona and Alabama—in which an entrenched judicial practice clearly seems to have induced restrictive consideration of relevant evidence. Appellate courts have grounds for finding an *Eddings* violation when all three of the following facts concerning a lower court's sentencing analysis obtain: (i) the court did not even attempt to justify or explain why restrictive treatment of the mitigating evidence was morally appropriate, or why alternative theories of the moral significance of the evidence should be rejected; (ii) the same court, or other courts in its jurisdiction, have in the past routinely appraised the evidence according to the same circumscribed set of moral principles while citing to prior precedent; (iii) independent reasons exist for thinking that a substantial number of reasonable jurors would consider the evidence mitigating based on principles that the court did not even consider. The combination of these factors

[116] *Id.* at 813-18.

[117] *Id.* at 814.

suggest that the court did not engage in a careful, individualized *moral* assessment of the mitigating evidence, but instead simply followed an entrenched legal practice or custom. The first factor suggests an absence of moral analysis; the second indicates that the court was following a legal convention; and the third factor indicates that if the court had considered alternative, widely endorsed moral principles, then it would have reached a different conclusion about the evidence's mitigating value.

Of course, other factors might supplement an appellate court's review. For instance, it would undoubtedly be relevant if, as is in Arizona, a statute or precedent had previously demanded the same limited consideration.[118] The Arizona Supreme Court had, before *Tennard*, interchangeably described SED evidence lacking the requisite causal connection as "irrelevant" and as having "little to no weight," and switched to exclusive use of the weighing language only after *Tennard*.[119] This suggests that the court saw the outright exclusion of SED evidence from consideration and the denial of weight to it after restrictive consideration as equivalent.

Our test applied in the SED context suggests that restrictive consideration of SED evidence by judges is frequently unconstitutional. To approach the analysis in reverse, consider the third factor. We offered arguments in Part II in support of the notion that

---

[118] State v. Hoskins, 14 P.3d 997, 1022 (Ariz. 2003).

[119] The Ninth Circuit has expressed some confusion about the Arizona Supreme Court's application of the causal nexus exclusion rule, stating that "Arizona's case law in this regard is conflicting," and citing interchanging examples of the state supreme court saying that it was either (a) considering evidence without a causal nexus but giving it no weight or (b) altogether refusing to consider such evidence. Towery v. Ryan, 673 F.3d 933, 946 (9th Cir. 2012); *see also* Lopez v. Ryan, 630 F.3d 1198, 1200 (9th Cir. 2011) (noting that "Arizona has a checkered past" with respect to using the causal nexus test as a clearly illegal screening mechanism and as a weighing mechanism). This mixed record might be explained if the state court saw no difference between the two rules.

SED evidence is mitigating irrespective of its exact causal relationship with the crime—arguments concerning the defendant's moral capacities and culpability, the defendant's prior suffering, and the state's moral standing to punish. We also referred to studies demonstrating that a substantial number of jurors tend to treat SED evidence as *inherently* mitigating.[120]

Now consider the first and second factors. We have struggled to find instances—in any American jurisdiction—where a court made a serious attempt to explain why from the moral point of view SED can only be mitigating if the causal nexus with the crime obtains, as discussed in Part I. Indeed, the opinions we have reviewed rarely if ever provide any rationale for the limitation. Instead, courts tend to cite earlier cases where a judge relied on restrictive consideration—and not as *persuasive* authorities, because the cited cases rarely include an explanation of the moral grounds of the causal nexus requirement.

In the rare instances in which judges attempt to critique alternative approaches to SED evidence, they critique caricatures of them. For instance, in one case the Alabama state court of criminal appeals stated that "[t]he argument that when a bad social environment produces bad people, that fact should in some way mitigate the punishment for these bad people, leads ultimately to the absurd conclusion that only people who come from an impeccable social background deserve the death penalty if they commit capital murder."[121] We are unaware of any judge or scholar who has argued either that

[120] *See* discussion *infra* Part II.

[121] Thompson v. Alabama, No. CR–05–0073, 2012 WL 520873, at *85 (Ala. Crim. App. 2012).

mild deprivations are mitigating, or that even severe deprivations automatically disqualify a defendant from receiving the death penalty.[122] Certainly the arguments we consider in Part II do not have either of these implications. In another case, the Florida Supreme Court suggested that the defendant's childhood abuse could not have reduced his moral responsibility because "the defendant's sister, who had also been abused, including sexually abused by the same alcoholic father, proceeded to live a normal and productive life."[123] But on most theories of SED's mitigating value, as discussed in Part II, the deprivation need not *determine* a person's wrongful acts in order to diminish his punishment-worthiness.

Admittedly, the application of *Eddings* to the practice of restrictive consideration of SED evidence is imperfect, a point that the dissent in *McKinney* was eager to emphasize.[124] The court of appeals in *Eddings* explicitly stated that it was using the "legal test of criminal responsibility" to *exclude* the SED evidence as "non-mitigating." By contrast, courts that give restrictive consideration to SED's mitigating value do not claim to be "following the law," and they tend to give SED "little to no" mitigating weight rather than none at all. As we explained above, however, we think that *Eddings* rests on a broader principle: that a sentencing judge should not limit their moral evaluation of mitigating evidence based on any legal custom or authority, even if the custom is never expressly acknowledged or even recognized by the judge. A test for SED's mitigating

[122] Of course, mere humanity might be thought to be automatic disqualification—but this would apply to all persons and not just SED sufferers.

[123] Douglas v. State, 878 So.2d 1246, 1260 (Fla. 2004).

[124] McKinney v. Ryan, 813 F.3d 798, 843-44 (9th Cir. 2015) (Bea, J., dissenting).

value that is applied in customary fashion, one that drastically and counter-intuitively limits the deprivation's mitigating weight, is inconsistent with such a principle.

Moreover, as explained in Part I, in jurisdictions that favor restrictive treatment, "little to no" mitigating weight is equivalent, at least in effect, to excluding the evidence outright. A survey of Arizona capital cases, for example, makes clear that mitigating evidence given "little" or "slight" weight rarely, if ever, results in leniency.[125] Before pronouncing a death sentence, courts often cursorily attach "little" weight to *all* of the mitigating factors in the case—indicating both that the "little" modifier is meant as a dismissal, and that mitigating factors of "little" weight do not warrant a lighter sentence *even when considered in aggregate*.[126] In other cases, mitigating evidence assigned little to no weight is so far from the sentencing judge's mind that it is excluded from her final list of mitigating factors.[127]

Accordingly, appellate courts have sound basis to find a failure to "consider" SED evidence under *Eddings* whenever lower courts routinely rely on restrictive consideration of deprivation evidence without explanation or defense, especially in light of the strong reasons for thinking that SED is mitigating in the absence of any causal connection with the crime. Restrictive consideration *may* pass constitutional muster when a sentencing judge offers some explanation or justification for taking a markedly limited view of

---

[125] *See supra* note 27.

[126] *See supra* note 28.

[127] *See, e.g.,* Poyson v. Ryan, 743 F.3d 1185, 1210 (9th Cir. 2013) ("For at the end of its opinion, the state court listed all of the mitigating circumstances it considered in its independent review of Poyson's death sentence. It omitted from this critical tally both Poyson's personality disorders and his abusive childhood.").

SED's mitigating value, and there are indications that the judge is engaging in independent moral analysis. In general, however, restrictive consideration of SED evidence in the jurisdictions we have studied appears to be the product of an entrenched judicial practice or custom that has artificially cabined the individualized moral inquiry that *Lockett* and its progeny demand.

B. Communal Endorsement & the Constitutional Importance of Evaluating Mitigating Evidence Under a Range of Reasonable Moral Principles

Judges could simply consider SED's mitigating value on the basis of a variety of different moral perspectives and principles. For example, instead of considering whether SED is mitigating based on the impaired capacities/responsibility theory alone, they might also consider it's mitigating weight on the basis of the whole-life view of retributive justice. This would obviate the need to justify a restrictive view of SED's mitigating value in the sentencing decision. More importantly, it would be consistent with a line of Supreme Court precedent since *Gregg*, emphasizing that the death penalty depends for its constitutional legitimacy on its link with community values.[128]

[128] *See, e.g.*, McCleskey v. Kemp, 481 U.S. 279, 310 (1987) (outlining the capital jury's "task of express[ing] the conscience of the community on the ultimate question of life or death" (citation omitted)); Gregg v. Georgia, 428 U.S. 153, 184 n.30 (1976) ("Punishment is the way in which society expresses its denunciation of wrong doing: and, in order to maintain respect for law, it is essential that the punishment inflicted for grave crimes should adequately reflect the revulsion felt by the great majority of citizens for them."(citation omitted)); *id*. at 181 (reflecting on the importance of maintaining a link between contemporary community values and the penal system (citation omitted)); Woodson v. North Carolina, 428 U.S. 280, 297–98 (1976) (reflecting on the

Whereas our argument above emphasized the constitutionally suspect nature of the practice of taking a restrictive view of SED's mitigating weight for granted and without explanation, here we argue that judges may be constitutionally obliged to give unrestricted consideration to SED evidence: that is, consideration based on a number of different moral principles that are sufficiently plausible. Unrestricted consideration is *inclusive*: it incorporates a diversity of perspectives on the mitigating potential of SED; and sole-sentencing judges have a special responsibility to ensure that the defendant is sentenced to death only if such a penalty would enjoy broad-based communal support.

The importance of broad-based communal support to the constitutionality of capital sentencing schemes is well established. The death penalty must be tested against the "conscience of the community,"[129] and "one of the most important functions" of the sentencing agent in a capital trial is to "maintain a link between community values and the penal system."[130] It is, indeed, no coincidence that the constitutionality of the death

---

importance on the moral views of society to the administration of death sentences (citation omitted)); Taylor v. Louisiana, 419 U.S. 522, 530 (1975) (noting that the capital sentencing jury as a representative of a criminal defendant's community provides him with "diffused impartiality" (citation omitted)); Witherspoon v. Illinois, 391 U.S. 510, 519 (1968) (describing the sentencer's task as that of "express[ing] the conscience of the community on the ultimate question of life or death"). *See also* Steven Semeraro, *Responsibility in Capital Sentencing*, 39 SAN DIEGO L. REV. 79, 144-45 n.232 (2002) (reciting evidence that the "the case law as a whole indicates that communal values must play a role in capital sentencing").

[129] *Witherspoon*, 391 U.S. at 519 ("[A] jury that must choose between life imprisonment and capital punishment can do little more—and must do nothing less—than express the conscience of the community on the ultimate question of life or death."). *See also* Spaziano v. Florida, 468 U.S. 447, 469 (1984) (Stevens, J., dissenting) (a death sentence "is ultimately understood as an expression of the community's outrage—its sense that an individual has lost his moral entitlement to live"); *id.* at 483 ("But more important than its procedural aspects, the life-or-death decision in capital cases depends on its link to community values for its moral and constitutional legitimacy.").

[130] *See Gregg*, 428 U.S. at 181 ("[O]ne of the most important functions any jury can perform in making . . . a selection (between life imprisonment and death for a defendant convicted in a

penalty, in light of the Eighth Amendment's familiar prohibition against "cruel and unusual" punishments,[131] turns on the contemporary moral values of the public. The Amendment "draw[s] its meaning from the evolving standards of decency that mark the progress of a maturing society,"[132] and in the past fifteen years alone, the Supreme Court has held that capital punishment is unconstitutional for the mentally handicapped,[133] for minors,[134] and for crimes other than murder and treason[135]—all to bring our sentencing practices into alignment with the evolving moral standards of the citizenry. The fact that the death penalty is ever on the verge of being cruel and unusual by contemporary standards underscores the fact that capital sentencing depends for its ongoing legitimacy on the people's approval.[136]

---

capital case) is to maintain a link between contemporary community values and the penal system." (citation omitted)).

[131] U.S. CONST. amend. VIII.

[132] *Gregg*, 428 U.S. at 173 (citation omitted). *See also* Kennedy v. Louisiana, 554 U.S. 407, 419 (2008); Woodson v. North Carolina, 428 U.S. 280, 301 (1976) (plurality opinion); Furman v. Georgia, 408 U.S. 238, 242 (1972) (Douglas, J., concurring); *id*. at 274-79 (Brennan, J., concurring); *id*. at 309-10 (Stewart, J., concurring); *id*. at 332 (Marshall, J., concurring); *id*. at 382-84 (Burger, C.J., dissenting); *id*. at 409 (Blackmun, J., dissenting); *id*. at 429-30 (Powell, J., dissenting); Trop v. Dulles, 356 U.S. 86, 101 (1958) (plurality opinion).

[133] Atkins v. Virginia, 536 U.S. 304 (2002) (abolishing the death penalty for the mentally retarded).

[134] Roper v. Simmons, 543 U.S. 551 (2005) (abolishing the death penalty for individuals under the age of eighteen at the time of their capital crimes).

[135] *Kennedy* 554 U.S. at 407 (abolishing the death penalty for rape where the death of the victim was neither the result nor the intent).

[136]While it is not our concern in this article to defend this conception of capital sentencing, we take the rationale to be fairly obvious. The state, acting on society's behalf, needs to earn its moral approval before it inflicts such a grave harm on a person as death; in a pluralistic society, this means ensuring that a death sentence has been tested against as many of the dominant moral views of a community as possible.

The importance of broad moral approval is also apparent in the near-universal state legislative preference for jury-based capital sentencing. Even before a Supreme Court ruling in the last decade constitutionally mandated jury participation in capital sentencing, 33 of 38 death penalty states already required it.[137] In 27 of the current 31 death penalty states, the jury's decision to sentence a defendant to life imprisonment is final and cannot be overridden by a trial judge.[138] The case law and academic commentary explain this legislative preference in terms of the jury's perceived status as an especially reliable indicator of the "conscience of the community."[139] The twelve-

[137] Ring v. Arizona, 536 U.S. 584 (2002), requires that juries find all aggravating factors in death penalty cases, so juries must be involved at least to that extent. The only state in which the jury continues to be formally uninvolved in capital sentencing is Montana, which issued its last death sentence in 1996, prior to *Ring*. *See* MONT. CODE ANN. § 46-18-301 (2013). Even before *Ring*, only four other states—Arizona, Colorado, Idaho, and Nebraska—used exclusively judicial capital sentencing.

[138] Woodward v. Alabama, 134 S. Ct. 405, 407 (2013) (Sotomayor, J., dissenting). The only states in which jury decisions are not final are Delaware, where only one jury life sentence has been overridden in favor of death, and that was overturned by the state supreme court; and Indiana, where the judge may decide the sentence if the jury cannot reach a unanimous sentence, 2002 Ind. Acts 1734.

[139] *See* Atkins v. Virginia, 536 U.S. 304, 323 (2002) ("[the jury] . . . is a significant and reliable objective index of contemporary values") (quoting Coker v. Georgia, 433 U. S. 584, 596 (1977) (plurality opinion) (quoting Gregg v. Georgia, 428 U.S. 153, 181 (1976))); *id.* (noting the jury's function of "maintain[ing] a link between contemporary community values and the penal system"(citation omitted)); Witherspoon v. Illinois, 391 U.S. 510, 519 n.15 (1968) (quoting Trop v. Dulles, 356 U.S. 86, 101 (1958) (plurality opinion)) (arguing that juries preserve the essential link between capital punishment and communal values). *See also* Stephen P. Garvey, "*As the Gentle Rain from Heaven": Mercy in Capital Sentencing*, 81 CORNELL L. REV. 989, 1003 n.56 (1996) ("Capital sentencing juries are said to represent the 'conscience of the community.' However, they 'represent' the community only because they are members of the community, not because they discern and then apply community standards."); Stephen Gillers, *Deciding Who Dies*, 129 U. PA. L REV. 1, 101-19 (1980) (arguing that the jury, as representative of the community, is more likely to accurately measure the offense against community outrage); Michael Mello & Ruthann Robson, *Judge Over Jury: Florida's Practice of Imposing Death Over Life in Capital Cases*, 13 FLA. ST. U. L. REV. 31, 48 (1986) (arguing that "the requirement that a capital sentencing jury consist of twelve persons as compared with a solitary person acting as judge also contributes to the prospect that a cross section of the community will be making the sentencing decision").

person capital sentencing jury is selected to approximate a random cross-section of the community, one believed to be significantly more likely than a sole sentencing judge to bring a diversity of moral perspectives to bear on the sentencing decision.[140] The jury's unanimity—required for the imposition of the death penalty in every state save Florida and Alabama[141]—makes even likelier that each death sentence will enjoy widespread public support. Evidence that would mitigate the defendant's punishment-worthiness in the eyes of a substantial portion of the community is less likely to be overlooked by multiple jurors than by a judge acting as the sole sentencer—or so the advocates of jury sentencing argue.[142]

Until recently, two states allowed trial judges to independently issue death sentences, even when it meant overriding a jury's recommendation of life imprisonment. Yet even while doing so, Florida gave privileged status to the jury verdict, because of the jury's ability to represent communal sentiment. In that state, the trial judge could not impose death over a jury's recommendation of life unless "the facts suggesting a sentence

---

[140] *Id*. Of course, a single jury may not fully reflect dominant community sentiment insofar as voir dire challenges can skew a jury's cross-sectional character. *See* Williams v. Florida, 399 U.S. 78, 102 (1970) ("Even the 12-man jury cannot insure representation of every distinct voice in the community, particularly given the use of the peremptory challenge."); Gary Goodpaster, *Judicial Review of Death Sentences*, 74 J. CRIM. L. & CRIMINOLOGY 786, 798 (1983) ("[W]hile the jury role is essential to ensure expression of present and developing community sentiment there is a risk that individual juries may not reflect that sentiment.").

[141] FLA. STAT. ANN. § 921.141(3) ("Notwithstanding the recommendation of a majority of the jury, the court, after weighing the aggravating and mitigating circumstances, shall enter a sentence of life imprisonment or death.").

[142] *See, e.g.*, Ballew v. Georgia, 435 U.S. 223, 233-34 (1978) (surveying the empirical data and concluding that a greater number of decision makers increases the likelihood of approximating "the common sense of the community," and that "the smaller the group, the less likely it is to overcome the biases of its members to obtain an accurate result" (citation omitted)). *See also* SCOTT E. SUNDBY, A LIFE AND DEATH DECISION: A JURY WEIGHS THE DEATH PENALTY 125 (2005) (describing the difference in moral perspectives of pro-life vs. pro-death jurors).

of death [were] so clear and convincing that *virtually no reasonable person could differ*."[143] Florida's specific override provision was overturned in 2016 as a violation of the Sixth Amendment right to have all critical findings necessary to impose the death penalty decided by jury, because the jury in Florida issued no factual findings with its recommended verdict.[144] Although the Supreme Court has previously approved Alabama's judicial override, which did not require deference to the jury but did require the jury to find aggravating factors, doubts about the constitutionality of the practice linger. Earlier this year, the Alabama governor signed legislation banning the override for defendants convicted after April 11th.[145] The constitutional question is not entirely moot, however, because Florida may still rewrite its judicial override scheme and the recent Alabama legislation left the 183 inmates already on the state's death row unaffected.[146]

One of the central doubts animating resistance to judge-determined death sentences regards the trial judge's capacity to adequately embody the "conscience of the

---

[143] Tedder v. State, 322 So. 2d 908, 910 (Fla. 1975). The Supreme Court recently, in Hurst v. Florida, 136 S.Ct. 616 (2016), invalided an iteration of Florida's override because it allowed the judge to override the jury not just on the overall weight of the aggravating factors against the mitigating factors but on the initial finding of aggravating/mitigating factors as well.

[144] *Hurst*, 136 S.Ct. at 624 (2016).

[145] "Alabama Ends Death Penalty by Judicial Override," Associated Press (Apr. 11, 2017),https://www.usnews.com/news/best-states/alabama/articles/2017-04-11/alabama-ends-death-penalty-by-judicial-override; Harris v. Alabama, 513 U.S. 504, 515 (1995). *See also* Brooks v. Alabama, 2016 WL 266239, at *1 (U.S. Jan. 21, 2016) (Sotomayor, J., concurring in denial of cert.) ("This Court's opinion upholding Alabama's capital sentencing scheme was based on *Hildwin v. Florida*, and *Spaziano v. Florida*, two decisions we recently overruled in *Hurst v. Florida*.").

[146] "Alabama Ends Death Penalty by Judicial Override," Associated Press, Apr. 11, 2017, https://www.usnews.com/news/best-states/alabama/articles/2017-04-11/alabama-ends-death-penalty-by-judicial-override.

community" in sentencing.[147] Justice Stevens, dissenting in *Harris*, where the majority approved Alabama's capital sentencing scheme, observed that, "an unfettered judicial override of a jury verdict for life imprisonment cannot be taken to represent the judgment of the community. A penalty that fails to reflect the community's judgment that death is the appropriate sentence constitutes cruel and unusual punishment under our reasoning in *Gregg*."[148] His dissent argued that:

> [T]he men and women of the jury may be regarded as a microcosm of the community, who will reflect the changing attitudes of society as a whole to the infliction of capital punishment, and that there could therefore be no more appropriate body to decide whether the fellow-citizen whom they

[147] Witherspoon v. Illinois, 391 U.S. 510, 519 (1968). *See also* Williams v. New York, 337 U.S. 241, 253 (1949) (Murphy, J., dissenting) ("In our criminal courts the jury sits as the representative of the community. Its voice is that of the society against which the crime was committed. A judge, even though vested with statutory authority to do so, should hesitate indeed to increase the severity of such a community expression."); Scott E. Erlich, *The Jury Override: A Blend of Politics and Death*, 45 AM. U. L. REV. 1403, 1431 (1996) (noting that a judicial override is problematic because "it tends to dilute the community's voice as represented by the collegial body—the jury"); *id*. at 1434 ("[T]his deficiency has created a situation in which the conscience of the community—the jury—has been all but removed from Alabama's capital sentencing process."); Stephen Gillers, *The Quality of Mercy: Constitutional Accuracy at the Selection Stage of Capital Sentencing*, 18 U.C. DAVIS L. REV. 1037 (1985) (arguing that the constitutional law demands a sentencing body that has competency to decide the death question fairly); Shannon Heery, *If It's Constitutional, Then What's the Problem?: The Use of Judicial Override in Alabama Death Sentencing*, 34 WASH. U. J. L. & POL'Y 347, 392 (2010) (noting the jury's role to represent the community); Michael Mello & Ruthann Robson, *Judge Over Jury: Florida's Practice of Imposing Death Over Life in Capital Cases*, 13 FLA. ST. U. L. REV. 31, 47 (1986) ("Given that the purpose of a death sentence is to reflect community standards, judges should be denied the power of the override unless or until we are willing to evaluate prospective judges as to their propensity to embody communal consciousness.") .

[148] Harris v. Alabama, 513 U.S. 504, 525 (1995) (Stevens, J., dissenting).

have found guilty of murder should . . . [die] or receive a lesser punishment.[149]

More recently, Justice Sotomayor, dissenting from the Court's decision not to hear a case that would have provided an occasion to reconsider the constitutionality of Alabama's judicial override, observed that, "[b]y permitting a single trial judge's view to displace that of a jury representing a cross-section of the community, Alabama's sentencing scheme has led to curious and potentially arbitrary outcomes."[150] Notably, these justices perceived a tension between the majority's tolerance for judicial overrides in *Harris* and the Court's earlier precedent, in cases like *Gregg*, emphasizing the need for death sentences to be issued *only if* they would enjoy broad-based communal support.[151]

Setting aside the question of the constitutionality of judge sentencing in the capital context, we think that, at the very least, the importance of ensuring broad-based communal support for the death penalty militates strongly in favor of unrestricted consideration of deprivation evidence whenever the judge is the sole sentencer, precisely because such consideration involves assessing the evidence based on a diverse range of

---

[149] *Id.* at 517 (quoting Royal Commission on Capital Punishment 1949-1953, Report 200 (1953)).

[150] Woodward v. Alabama, 134 S. Ct. 405, 409-10 (2013 (Sotomayor, J., dissenting)) ("For example, Alabama judges frequently override jury life-without-parole verdicts even in cases where the jury was unanimous in that verdict. In many cases, judges have done so without offering a meaningful explanation for the decision to disregard the jury's verdict. In sentencing a defendant with an IQ of 65, for example, one judge concluded that "[t]he sociological literature suggests Gypsies intentionally test low on standard IQ tests." Another judge, who was facing reelection at the time he sentenced a 19–year–old defendant, refused to consider certain mitigating circumstances found by the jury, which had voted to recommend a life-without-parole sentence. He explained his sensitivity to public perception as follows: "'If I had not imposed the death sentence, I would have sentenced three black people to death and no white people.'"(citations omitted)).

[151] Gregg v. Georgia, 428 U.S. 153, 181 (1976) (quoting *Witherspoon*, 391 U.S. at 519 n.15).

moral perspectives on SED's mitigating value. In fact, we think our argument generalizes to all mitigating evidence: judges should embrace unrestricted analysis whenever they and not the jury decide the death penalty. The consistency of the Supreme Court's death penalty jurisprudence would be well served by a more explicit acknowledgment of this fact.

The sole sentencing judge does not enjoy the benefits of multiple voices participating in the sentencing process. If she brings only her *own* private moral beliefs to bear on the sentencing decision, the likelihood becomes high that any death sentence she issues will reflect only her private, as opposed to a communal, moral response. To guard against that risk, the sentencing judge, unlike the individual juror, needs to take seriously moral principles endorsed by her fellow citizens that assign significant weight to relevant mitigating evidence that she may not ultimately be persuaded by.[152] If some factor would be deemed, for plausible reasons, to be substantially mitigating by a significant number of reasonable judges and jurors, the sentencing judge should regard it as such even if she is ultimately unconvinced of its mitigating worth.[153]

Accordingly, sole sentencing judges should embrace unrestricted consideration of SED evidence. Unrestricted consideration incorporates the view that SED is mitigating

---

[152] In other words, the judge, as sentencer, needs to be a more self-conscious representative of public morality than the individual juror in a twelve-person jury. Feminist approaches to the role and responsibilities of the judge have been especially clear on the importance of "communal modes of decision-making" and the need to consult multiple, competing perspectives. *See, e.g.*, Judith Resnik, *On the Bias: Feminist Reconsiderations of the Aspirations for Our Judges*, 61 S. CAL. L. REV. 1877, 1924-26 (1988).

[153] To be clear, we are not agreeing with Justice Stevens's view in his *Spaziano* dissent that only the jury should be permitted to impose the death penalty. *See Spaziano v. Florida*, 468 U.S. 447, 490 (1984) (Stevens, J., dissenting). We argue only that the judge ought to emulate the jury when performing a function traditionally—and for good reason—left to juries.

when it impairs the defendant's ability to control his conduct and thereby limits his culpability. It recognizes the life-cycle view of retributive justice and the constraints on inflicting excessive suffering on persons who have led miserable lives. It considers the state's diminished standing to punish individuals who have been left behind. In other words, unrestricted consideration involves recognizing SED's substantial mitigating significance in the absence of demonstrable causal connections with the crime, and thereby ensures that serious deprivation has the effect at sentencing that it would have had it been considered by a representative collection of members of the community. As discussed in Part II, the treatment of SED as inherently mitigating is based on moral considerations that are substantively reasonable and enjoy wide-appeal. If one of the most important functions that the sentencer can serve in capital cases is ensuring that the death penalty is only issued if would enjoy broad-based communal support, sole sentencing judges should embrace unrestricted consideration of SED's mitigating value (and of mitigating evidence more generally).[154]

[154] A few caveats are in order. Our argument may seem as relevant to evidence offered in aggravation as it is to evidence offered in mitigation. After all, the sentencing agent must aim to capture the community's outrage as well as its compassion. Does this not entail that, if a great many reasonable persons believe that SED mitigates only if it was a specific cause of the crime, judges should give *less* weight to such SED? The simple answer is no. Structurally, the capital sentencing process is designed to be more responsive to the compassionate side of the community's moral response than to its vindictive side. By requiring jury unanimity for death sentences, most states tilt the scales in favor of the community's mercy. A single holdout vote for a life sentence generally has decisive power on a jury, whereas a single vote for the death penalty is powerless. Moreover, while the Supreme Court prohibits any constraints on the sentencing agent's authority to assess factors as mitigating, it has imposed constitutional constraints on which factors may be regarded as aggravating. *See* Penry v. Lynaugh, 481 U.S. 279, 304 (1987) ("In contrast to the carefully defined standards that must narrow a sentencer's discretion to impose the death sentence, the Constitution limits a State's ability to narrow a sentencer's discretion to consider relevant evidence that might cause it to decline to impose the death sentence."); Lockett v. Ohio, 438 U.S. 586, 604 (1978) (noting that a greater degree of reliability is required precisely on the issue of death-deservingness). Indeed, the scope of potentially aggravating evidence must be narrowly defined by statute. Zant v. Stephens, 462 U.S. 862, 877

Finally, it should be noted that our argument only extends to the *moral* or *normative* evaluation of mitigating evidence. There remains substantial room for judicial discounting of SED evidence on *empirical* grounds, and judges are under no obligation to consider communal values when reviewing the empirical facts. As mentioned above, the factual record may sometimes lead a judge to reasonably question whether claimed environmental deprivation actually occurred. In such cases, proffered SED evidence may well be properly dismissed by the judge before the question of moral significance even arises.

Conclusion

At critical junctures throughout the sentencing process, individual actors are tasked with making moral determinations. Yet very little attention has been paid to when this moral discretion is exercised correctly. That seems to be changing, at least in the capital sentencing context, with appellate courts being more willing to scrutinize sentencing decisions for failures to properly "consider and give effect to" relevant mitigating evidence. We have attempted to provide some clarity to this area of jurisprudence by closely examining the nature of the moral consideration of mitigating evidence that is required under constitutional law. Using the unusually restrictive

(1983) ("[To avoid a constitutional flaw] an aggravating circumstance must genuinely narrow the class of persons eligible for the death penalty and must reasonably justify the imposition of a more severe sentence on the defendant compared to others *found* guilty of murder."). Moreover, whereas the imposition of death must enjoy broad moral approval in order to be legitimate, the Supreme Court has never indicated that such broad appeal is necessary for a life sentence. *See, e.g.*, Furman v. Georgia, 408 U.S. 238 (1972). Our argument accordingly requires judges to take greater care in giving effect to the community's compassion than to its vengeance. Although we have not discussed it here, there may be further reason for judicial deference to merciful moral concerns discoverable in the fact that there are many members of society who do not favor the death penalty under any circumstance.

treatment of severe environmental deprivation evidence in some jurisdictions as our starting point, we have devised a three-factor test for determining when restrictive treatment of such evidence—the conditioning of deprivation's mitigating potential on restrictive conditions like its being a specific cause of the crime—represents an *Eddings* violation. Our test is based on the principle, drawn from a long line of Supreme Court rulings, that the sentencer cannot artificially limit her consideration of the mitigating weight of evidence presented by the defense using legal rules, whether those rules are derived from statute, prior case law, or judicial custom. Additionally, we have argued that in light of the importance of ensuring that the death penalty is sanctioned by communal values, sole sentencing judges have an obligation to consider all of the possible ways in which SED might be seriously mitigating—at least those that many reasonable jurors and judges would endorse. In other words, unrestricted or broad consideration of deprivation evidence is in general mandatory under constitutional law. Between these two independent lines of constitutional argument, appellate courts have more than enough basis for review of cases in Arizona, Alabama, and wherever else restricted consideration of severe deprivation evidence by sentencing judges has unfairly and unlawfully prejudiced defendants convicted of capital crimes.

* * *