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

2023

Blame, Punishment, and Standing: Legal and Philosophical Dimensions

Preston, Samuel J.

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01qv33s089s

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

BLAME, PUNISHMENT, AND STANDING:

LEGAL AND PHILOSOPHICAL DIMENSIONS

Samuel J. Preston

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: Gideon Rosen, Philip Pettit, Michael Smith

January 2023

© Copyright by Samuel J. Preston, 2022. All rights reserved.

Abstract

Philosophers interested in blame typically focus on *blameworthiness*, i.e., the conditions under which an agent is an appropriate target of blame and punishment. But no less important to our understanding of blame is *standing*, i.e., the conditions an agent must meet to blame and punish appropriately. This dissertation argues that by focusing on standing, we can solve perennial problems in the theory of punishment and understand blame in a fuller light.

The dissertation is in two parts. The first, chapter 1, is targeted at legal as well as philosophical audiences. It argues that the standing to blame offers the best explanation for the Fourth Amendment exclusionary rule in American criminal procedure. This rule, which requires unconstitutionally obtained evidence to be excluded from criminal trial, is best understood as a mechanism to obviate criminal convictions that cannot support the state's standing to blame, and thus cannot serve as the basis for legitimate punishment. The chapter develops this argument into a comprehensive theory of the exclusionary rule that purports to better explain its doctrinal contours and more clearly elucidate its foundations than any extant, alternative account.

The second part, which spans chapters 2 through 5, is targeted at philosophers. Chapter 2 argues that many facts that can undermine blame's appropriateness have nothing to do with the target's blameworthiness. They include the blamer's hypocrisy or complicity in the wrongdoing at issue. A satisfactory account of blame must explain why *all* these considerations are relevant to blame's appropriateness, and no account that fixates on blameworthiness alone can do so.

Chapter 3 puts forward a novel view of blame that purports to meet this burden. The view, dubbed *Contractualism about Blame*, claims that blame's appropriateness turns on the contractualist justifiability of principles that permit the blamer to subject her target to a kind of punitive treatment. Chapter 4 then uses this framework to explain why hypocrisy can undermine the standing to blame, and chapter 5 uses the framework to identify the epistemic requirements a blamer must meet for her blame to be appropriate.

In loving memory of Holden T. Tanner, Yale Law School ’21.

Acknowledgments

Gideon Rosen has been a stellar main adviser and I owe him a great debt. Philip Pettit has been a constant source of wisdom, insight, and support. And Michael Smith has incisively challenged and encouraged me since he first taught me in 2014. I'm overwhelmingly grateful to all three of my advisers.

I also want to thank Liz Harman for her mentorship and the key role she played in the evolution of my thinking about blame; Gideon Yaffe for his generosity and support during my time at Yale Law School and my writing chapter 1 of this dissertation; and Akhil Amar, Carol Rovane, and Vicki Schultz for their years of feedback, counsel, and support.

Thank you, too, to my family: Tom, Eddie, Winnie, David, and most of all, my mother, Jane, whose love and aid have been boundless. And thank you to my friends who've helped me with this project, including Emad Atiq, Alisabeth Ayres, Calvin Baker, Jindai Joseph, Kyle Landrum, Cole Rianda, Aarthy Vaidyanathan, Melody Wang, and many others I've failed to mention, but to whom I'm truly grateful.

Finally, thank you to the Princeton Philosophy Department, the Princeton University Graduate School, and the Princeton University Center for Human Values, and thank you to Yale Law School, especially Dean Ellen Cosgrove and other administrators and staff, whose flexibility and support made this work possible.

Table of Contents

Abstract ..... *iii*

Acknowledgments ..... *v*

Table of Contents ..... *vi*

Note on Style and Citation Format ..... *viii*

Chapter 1: The Exclusionary Rule and the Standing to Blame *1*

I. The Exclusionary Rule: Doctrine and Explanandum **5**

II. Extant Accounts & Intractable Problems **13**

III. Punishment and the Standing to Blame **34**

IV. Giving Order to Exclusionary Rule Doctrine **53**

V. Conclusion and Future Work **89**

Chapter 2: Blame's Appropriateness and the Evaporation Heuristic *95*

I. The Plausibility of AIB **98**

II. The Debate and Piovarchy's Challenge **100**

III. The Evaporation Heuristic **102**

IV. The Standing to Blame **106**

V. Being Blameworthy vs. Deserving Blame **112**

VI. Conclusion **114**

Chapter 3: Contractualism about Blame *115*

I. A Theory of Blame Attitudes **116**

II. Contractualism in Outline **129**

III. Contractualism and Desert **135**

IV. The Practice of Punishing **142**

V. Conclusion **150**

Chapter 4: Hypocrisy and Superiority *151*

I. The Evaporation Heuristic and Pro Tanto Reasons for a Mixed View **153**

II. John 8's Principle **155**

III. Bell's Principle **157**

IV. A Provisional Hypothesis **162**

V. Wallace, the Principle of Equal Consideration, and the Private Blame Objection **165**

VI. Blamers' Objections **169**

VII. The Provisional Hypothesis Expanded **172**

VIII. Conclusion **180**

Chapter 5: The Epistemic Condition on Appropriate Blame *183*

I. A Glitch in the Evaporation Heuristic **185**

II. The Argument for an Epistemic Appropriateness Condition **187**

III. Followability and the Knowledge Principle **190**

IV. A Credence Principle? **194**

V. Evidential Principles **198**

VI. An Outstanding Issue: Moral Procedural Requirements? **206**

VII. Conclusion **210**

Note on Style and Citation Format

The first chapter of this dissertation is written as a philosophical law-review article whereas later chapters are written as analytic philosophy papers. I conform to the stylistic norms of law in the first chapter and the distinct stylistic norms of analytic philosophy in the remaining chapters. But because philosophy is more flexible about citation format than law, I opt to use the citation format required by law reviews—the "Blue Book" rather than APA or MLA format—throughout the dissertation for the sake consistency.

CHAPTER 1

THE EXCLUSIONARY RULE AND THE STANDING TO BLAME

Abstract: This chapter argues that the standing to blame offers the best explanation for the Fourth Amendment exclusionary rule in criminal procedure. When an officer obtains evidence of a party's criminal behavior only because the officer violated the party's constitutional rights, the exclusionary rule prohibits the state's use of the evidence at criminal trial. Critics object that the rule is unjustifiable, and proponents have struggled to articulate a basis for it that withstands scrutiny. This chapter is the first to argue that the exclusionary rule is best conceived as a mechanism to obviate criminal convictions that cannot support the state's standing to blame, and thus cannot serve as the basis for legitimate punishment. The chapter develops this argument into a comprehensive theory of the exclusionary rule that better explains its doctrinal contours and more clearly elucidates its foundations than any extant, alternative account. Before the rule is interpreted away, it is urgent that critics and proponents of the exclusionary rule grapple with the standing to blame.

INTRODUCTION

When an officer obtains evidence of a party's criminal behavior only because the officer violated the party's constitutional rights, the exclusionary rule ("ER") requires the evidence to be suppressed at criminal trial.[1] The rule's critics argue that it is legally unfounded, morally

[1] *See* Mapp v. Ohio, 367 U.S. 643 (1961) (incorporating the exclusionary rule); Weeks v. United States, 232 U.S. 383 (1914) (holding that the rule is required by the Federal Constitution); Boyd v. United States, 116 U.S. 616 (1886) (applying an inchoate form of the exclusionary rule for the first time). *See generally* WAYNE R. LAFAVE ET AL., 2 CRIMINAL PROCEDURE § 3.1 (4th ed. 2020) ("The Fourth Amendment Exclusionary Rule and Other Remedies"); John E. Theuman, *State Constitutional Requirements as to Exclusion of Evidence Unlawfully Seized*, 19 A.L.R.5th 470 (2020) (summarizing the Federal Constitutional rule and comparing it to state constitutions' exclusionary rules, insofar as the latter sometimes require exclusion in cases the former does not).

unsound, and wholly unprincipled[2]—qualities reflected in its ever-more-chaotic doctrine.[3] Yet its proponents have failed to articulate a unified explanation for the rule that justifies its existence in the face of critics' arguments.[4] And now, the rule is in peril.[5] This chapter makes an explanatory

---

[2] *See, e.g.*, Akhil Reed Amar, *Fourth Amendment First Principles*, 107 HARV. L. REV. 757, 770 *passim* (1994) (describing the rule as "unsupported"); *id.* at 761 ("The words of the Fourth Amendment . . . do not require – or even invite – exclusions of evidence."); *id.* at 799-800 ("The exclusionary rule renders the Fourth Amendment contemptible in the eyes of judges and citizens . . . [and] creates long-term instability, driving a wedge between We the People and Our Constitution"); H. Mitchell Caldwell, *Fixing the Constable's Blunder*, 2006 B.Y.U. L. REV. 1, 1 *passim* (2006) (asking how the rule survives despite its "wholesale damnation"); Margaret K. Lewis, *Controlling Abuse to Maintain Control*, 43 N.Y.U. J. INT'L L. & POL. 629, 642-43 (2011) (arguing that recent cases have drilled "another hole into the increasingly Swiss-cheese-like Fourth Amendment exclusionary rule" and signaled that the Supreme Court's stated rationale for the rule, deterrence, is no longer convincing); Lawrence Joseph Perrone, *Curing the Real Problem: Cleaning Up Fourth Amendment Jurisprudence by Altering Our Current Exclusionary Rule to Conform with an International Model*, 16 TULSA J. COMP. & INT. LAW 67, 67 (2008) ("The highly controversial automatic exclusionary rule has been called into doubt by many . . . [and] many instances [of its application] seem perverse to our own sense of justice."); *see also* Erid Kades, *Windfalls*, 108 YALE L.J. 1489, 1565 (1999) ("Perhaps the most well-known criminal windfall is the exclusionary rule.").

[3] Joseph R. Weisberger, *The Exclusionary Rule: Nine Authors in Search of a Principle*, 34 S.C. L. REV. 253, 253 (1982) ("Due to the philosophic disarray of its members, the Supreme Court has not demonstrated the ability to control the exclusionary rule, to define it, and to lend symmetry to its various applications and limitations. Consequently the exclusionary rule has developed a chaotic life of its own."); *see* Amar, *supra* note 2, at 759 (the rule is a "doctrinal mess"); *id.* at 785 ("The Court . . . has concocted the awkward and embarrassing remedy of excluding reliable evidence of criminal guilt, and has then tried to water down this awkward and embarrassing remedy in ad hoc ways."); Kit Kinports, *Culpability, Deterrence, and the Exclusionary Rule*, 21 WM. & MARY BILL RTS. J. 821, 821 (2013) ("By this point, the regrettable state of the Supreme Court's Fourth Amendment jurisprudence comes as no surprise to anyone. Academics and jurists of all stripes agree that the Court's case law in this area is a mess."); Christopher Slobogin, *Why Liberals Should Chuck the Exclusionary Rule*, 1999 U. ILL. L. REV. 363, 375 (describing the ER as "swiss cheese"); Potter Stewart, *The Road to* Mapp v. Ohio *and Beyond: The Origins, Development and Future of the Exclusionary Rule in Search-and-Seizure Cases*, 83 COLUM. L. REV. 1365, 1366 (1983) ("Looking back, the exclusionary rule seems a bit jerry-built—like a roller coaster track constructed while the roller coaster sped along."); *see also* Danielle Hayes, *He Says, She Says: Utah v. Strieff and the Role of Narrative in Judicial Decisions*, 61 HOWARD L. J. 611, 630 (2018) ("The numerous exceptions to the exclusionary rule undercut its ability to deter because they cause the rule to be applied inconsistently."); Katerina Kokkas, *Fourth Amendment Standing and the General Rule of Waiver*, 2018 U. CHI. LEGAL F. 309, 318-20 (describing "irreconcilable doctrinal tension" within the ER).

[4] *See infra*, Part II for an overview and critique of extant explanations for the rule.

[5] *See, e.g.*, Rhett DeHart, *Is the Exclusionary Rule Doomed?,* 26 S.C. LAWYER 40 (2015) (claiming "the tide has turned" against the rule); Tracey Maclin & Jennifer Rader, *No More Chipping Away: The Roberts Court Uses an Axe to Take Out the Fourth Amendment Exclusionary Rule*, 81 MISS. L.J. 1183 *passim* (2012) (arguing the Court intends to do away with the rule); Richard Re, *The Due Process Exclusionary Rule*, 127 HARV. L. REV. 1885, 1885 (2014) ("As the Supreme Court continues to cut back on and perhaps eliminate Fourth Amendment suppression, the exclusionary rule has entered a new period of crisis."); *id*. at 1885 n.1 (saying the rule is in "peril").

intervention: it is the first to argue that the ER must be understood in terms of the *standing to blame.*

Blame is a central part of American criminal punishment,[6] but for blame to be justified, it is not enough that the one blamed is a culpable wrongdoer or lawbreaker. The blamer, too, must have standing to blame.[7] Agents have standing to blame by default, but they can lose it through certain wrongs of their own, as in cases of culpable hypocrisy or complicity in the very wrongdoing for which they are blaming another.[8] This chapter argues that an agent can also lose the standing to blame when the agent has culpably violated her target's rights, the agent knows her target is a culpable wrongdoer only because of that violation, and the agent is at least as blameworthy for her own violation as her target is for her wrongdoing. The same applies when

[6] For arguments to this effect, *see* R.A. Duff, *Blame, Moral Standing, and the Legitimacy of the Criminal Trial*, 33 RATIO 123, 130 & n.16 (2010); Joel Feinberg, *The Expressive Function of Punishment*, 49 THE MONIST 397, 400 *passim* (1965); Douglas Husak, *"Broad" Culpability and the Retributivist Dream*, 9 OHIO ST. J. CRIM. L. 449, 449-50 (2011-12); sources cited *infra* notes 140-141.

[7] For arguments to this effect in the philosophical context, *see, e.g.*, G.A. Cohen, *Casting the First Stone: Who Can, and Who Can't, Condemn the Terrorists?*, 58 ROYAL INST. PHIL. SUPPLEMENT 113 *passim* (2006); Kyle G. Fritz & Daniel Miller, *Hypocrisy and the Standing to Blame*, 97 PAC. PHIL. Q. 118 (2015); Amy L. McKiernan, *Standing Conditions and Blame*, 32 SOUTHWESTERN PHIL. REV. 145 (2016); Angela Smith, *On Being Responsible and Holding Responsible*, 2 J. ETHICS 465, 466 *passim* (2007); Patrick Todd, *A Unified Account of the Moral Standing to Blame*, 53 NOÛS 347 (2019); R. Jay Wallace, *Hypocrisy, Moral Address, and the Equal Standing of Persons*, 38 PHIL. & PUB. AFFAIRS 307 (2010). For arguments that stress and rely on this point in the legal context, *see, e.g.*, Duff, *supra* note 6 *passim*; Victor Tadros, *Poverty and Criminal Responsibility*, 43 J. VALUE INQUIRY 391 *passim* (2009); Gary Watson, *A Moral Predicament in the Criminal Law*, 58 INQUIRY 168, 172 *passim* (2015); sources cited *infra* note 141. For a more skeptical take on the standing to blame, *see* Macalester Bell, *The Standing to Blame: A Critique*, *in* BLAME: ITS NATURE AND NORMS 263 (D. Justin Coates & Neal A. Tognazzini eds., 2013).

[8] The sources cited in *supra* note 7 all include hypocrisy or complicity in their discussions of standing to blame. For discussions of other conditions that can undermine the standing to blame, *see* D. Justin Coates, *The Epistemic Norm of Blame*, 19 ETHICAL THEORY & MORAL PRAC. 457 (2016) (standing requires knowledge of blameworthiness); Gabriel Mendlow, *The Moral Ambiguity of Public Prosecution*, 130 YALE L.J. 1146 (2021) (standing can be undermined when the state usurps the blamer role from the victim); Linda Radzik, *On Minding Your Own Business: Differentiating Accountability Relations Within the Moral Community*, 37 SOC. THEORY & PRAC. 574 (2011) (standing can be undermined by one's failure to mind one's own business); *infra* Section III.C (standing can be undermined by the wrongful and culpable acquisition of evidence essential to knowledge of blameworthiness).

the state is the agent, and the target is a criminal defendant.[9] Because the state cannot legitimately punish a defendant it cannot legitimately blame,[10] the ER protects the legitimacy of state punishment by demanding the suppression of evidence which, if essential to a guilty verdict, would undermine the state's standing to blame on the verdict's basis.

The standing to blame is thus the missing link in attempts to understand the exclusionary rule. This chapter develops a novel account of the rule in these terms and argues that it explains the rule's doctrinal complexities better than any extant alternative. The account also sheds light on a key problem within ER doctrine. We must modify the ER's good faith exception, along with the doctrine applying the rule to all constitutional violations, to preclude the rule's application in cases when a defendant's crime is more outrageous than the state's violation.[11] The account thus vindicates the intuition that the ER's application in those cases is unjust.[12]

Here's the plan. Part I reviews ER doctrine and clarifies the chapter's explanandum. Part II evaluates extant accounts of the ER and argues they cannot adequately explain the rule. Part III discusses the standing to blame and formulates this chapter's account of the rule. Part IV argues that this novel account is more explanatorily powerful than extant alternatives. It does so by illustrating the account's explanations and prescriptions for the rule's doctrinal features: in Section IV.A, the rule's application to criminal trials but not to civil, pre-trial, or post-conviction contexts; in Section IV.B, its good faith exception and application to all and only constitutional

[9] This is a rough statement of the view. For the precise, official formulation, *see* Section III.E.
[10] *See infra* Section III.D.
[11] *See infra* Section IV.B.
[12] *See* Amar, *supra* note 2, at 799 ("In the popular mind, the Fourth Amendment has lost its luster and become associated with grinning criminals getting off on crummy technicalities. When rapists are freed [on such bases], the people . . . lose respect for the Fourth Amendment."); sources cited and discussed *infra* note 184.

violations; and in Section IV.C, its causation requirements and the bar on vicarious assertions. Part V concludes and outlines key questions for future work.

I. THE EXCLUSIONARY RULE: DOCTRINE AND EXPLANANDUM

This Part provides key background about the exclusionary rule. Section I.A describes ER doctrine, and Section I.B argues that this doctrine is what a successful account needs to explain.

A. Doctrinal Overview

1. Exclusive to the Criminal Context

The ER is inapplicable outside the criminal context. It is not applied in any domain of civil law,[13] including deportation hearings and civil actions brought by governmental entities like the IRS.[14]

2. Exclusive to the Trial Phase

[13] *See* Penn. Bd. of Prob. & Parole v. Scott, 524 U.S. 357, 363 (1998) ("[W]e have repeatedly declined to extend the exclusionary rule to proceedings other than criminal trials."); Amar, *supra* note 2, at 792 ("[C]ivil exclusion is not the rule, never has been the rule, and shows little sign of becoming the rule.").
[14] *See* INS v. Lopez-Mendoza, 468 U.S. 1032 (1984) (deportation); United States v. Janis, 428 U.S. 433 (1976) (IRS civil action). Note also that violations of the ER in the Fourth Amendment context are categorically unreviewable on federal habeas. Stone v. Powell, 428 U.S. 465, 494 (1976) ("[A] state prisoner may not be granted federal habeas corpus relief on the ground that evidence obtained in an unconstitutional search or seizure was introduced at his trial."); *see also* BRIAN R. MEANS, FEDERAL HABEAS MANUAL § 1:60 (2020) ("*Stone* applies to State and Federal Prosecutions."). The *Stone* rule is surprising because though federal habeas is a civil action, it is undoubtedly part of the *criminal* process, for it reviews a criminal conviction and sentence. The *Stone* rule is also singular, and ER violations are the only kinds of violations subject to it. *See id.* § 1:59 (2020); *see also* Vikram David Amar, *Federal Practice and Procedure: Grounds for Writ—In General—The Rule of* Stone v. Powell § 4263.1 (2020) ("[T]he Supreme Court has not extended the *Stone* rule."). Moreover, the rule applies *only* when *Fourth* Amendment violations are the predicates for the ER error. *See* Withrow v. Williams, 507 U.S. 680 (1993) (allowing federal-habeas review for exclusionary-rule violations in the *Miranda* context); *see also id.* at 697, (O'Connor, J., concurring in part and dissenting in part) (arguing that *Withrow* is inconsistent with *Stone*).

The ER does not apply in grand jury proceedings, which occur prior to the initial charge,[15] nor does it apply in preliminary hearings where probable cause is determined or bail is set.[16] Nor, moreover, does the rule apply after conviction, either at the sentencing phase[17] or at future parole hearings.[18]

3. Usable in Impeachment

The ER does not apply to evidence used to impeach a defendant's testimony in a criminal trial.[19] This goes whether the impeachable testimony was given during direct or cross-examination.[20] But the prosecution cannot "'smuggle[] in' the impeaching opportunity in the course of cross-examination."[21] Rather, the defendant must first do something "to justify cross-examination in respect of the [tainted] evidence."[22]

If properly used to impeach the defendant's testimony, the tainted evidence will be admitted into the record.[23] However, the jury will be instructed that the evidence "should be considered only for impeaching [the defendant's] credibility," and not as direct evidence of guilt.[24]

4. Applied Across Constitutional Contexts

---

[15] United States v. Calandra, 414 U.S. 338 (1974); *see also* Massiah v. United States, 377 U.S. 201, 205 (1964) (stating that for purposes of exclusion in the context of Sixth Amendment violations, "the most critical period of the proceedings" begins at "the time of their arraignment" (citing Powell v. Alabama, 287 U.S. 45, 53 (1932))).
[16] *See* FED. R. CRIM. P. 5.1 & cmt. ("[T]he preliminary examination is not the proper place to raise the issue of illegally obtained evidence. This is current law." (citing Giordenello v. United States, 357 U.S. 480, 484 (1958)).
[17] *See* Del Vecchio v. Illinois Dep't of Corr., 31 F.3d 1363, 1388 (7th Cir. 1994) ("The exclusionary rule is generally inapplicable during sentencing."); State v. Nichols, 438 F.3d 437, 442 (4th Cir. 2006).
[18] Penn. Bd. of Prob. & Parole v. Scott, 524 U.S. 357 (1998) (holding that the ER is not applicable to parole revocation hearings).
[19] *See, e.g.*, United States v. Havens, 446 U.S. 620 (1980).
[20] *See id.* at 623-25.
[21] *Id.* at 625.
[22] *Id.* (citing Agnello v. United States, 269 U.S. 20, 35 (1925)).
[23] *Id.* at 623.
[24] *Id*.

The ER is often associated specifically with the Fourth Amendment (searches and seizures), but the rule also applies to Fifth Amendment contexts, including *Miranda* violations and illegally coerced confessions,[25] and to Sixth Amendment contexts, including violations of the right to counsel[26] and the confrontation clause.[27] In all these contexts, it is incorporated against the states.[28] Additionally, if a constitutional violation is established, the violation's severity plays no role in determining the ER's application; in this sense, the ER is "automatic" or "non-discretionary."[29]

The Court sometimes speaks as if there are "exclusionary rules,"[30] one for each constitutional context. But more often, it speaks as if there is a single rule, "[t]he exclusionary rule,"[31] that applies across constitutional contexts.[32] This chapter adopts the second convention, for though there are certain doctrinal distinctions between the rule's operations in different

[25] *See* Miranda v. Arizona, 384 U.S. 436, 444 (1966) ("[T]he prosecution may not use statements . . . stemming from custodial interrogation of the defendant unless it demonstrates the use of procedural safeguards effective to secure the privilege against self-incrimination."); *see also* Withrow v. Williams, 507 U.S. 680 (1993) (elaborating and applying this rule in the Federal Habeas context).
[26] *See* Massiah v. United States, 377 U.S. 201, 204-05 (1964).
[27] *See* Crawford v. Washington, 541 U.S. 36, 42, 50-51 (2004) (rejecting the view that the Confrontation Clause is satisfied "so long as [the hearsay] has adequate indicia of reliability," for such a view "would render the Confrontation Clause powerless").
[28] The rule was incorporated and applied to the states in the 1960's, first in the Fourth Amendment context, Mapp v. Ohio, 367 U.S. 643 (1961), then in the Sixth Amendment context, *Massiah*, 377 U.S. 201 (1964), and then in the Fifth Amendment context, *Miranda*, 384 U.S. 436 (1966). The exclusionary rule sets a floor for the states; they are free to provide rules that are more defendant friendly. *See* Mark E. Cammack, *The Rise and Fall of the Constitutional Exclusionary Rule in the United States*, 58 AM. J. COMP. L. 631, 631-33 (2010).
[29] *See, e.g.*, Craig M. Bradley, Mapp *Goes Abroad*, 52 CASE W. L. REV. 375, 382 n.55 (2001); James Stribopoulos, *Lessons from the Pupil: A Canadian Solution to the American Exclusionary Rule Debate*, 22 BOS. C. L. REV. 77, 84 (1999); Adam Liptak, *U.S. Is Alone in Rejecting All Evidence if Police Err*, N.Y. TIMES, July 19, 2008, https://www.nytimes.com/2008/07/19/us/19exclude.html.
[30] Lego v. Twomey, 404 U.S. 477, 488 (1972).
[31] Brown v. Illinois, 422 U.S. 590 *passim* (1975).
[32] *See* WILLIAM ANDREW KERR, 16A INDIANA PRACTICE SERIES, CRIMINAL PROCEDURE—TRIAL § 14.1 (2020).

constitutional contexts, they are peripheral and easy to highlight when relevant.[33] The one exception is the so-called "exclusionary rule" in the context of testimonial immunity for legally compelled (not illegally coerced) self-incriminating testimony, which operates *prospectively* to ensure no Fifth Amendment violation occurs to begin; it is not at issue in this chapter.[34]

5. Applied Only to Constitutional Contexts

Although the ER operates across constitutional contexts, it does not operate in non-constitutional contexts, from mere statutory violations[35] to violations of international treaties.[36] The Court recently held, for example, that even granting officers had acquired incriminating evidence against a suspect by violating his rights to consular notification and communication under Article 36 of the Vienna Convention on Consular Relations, the ER does not apply.[37]

---

[33] I primarily have in mind the fact that exclusionary rule violations can be reviewed on federal habeas if predicated on the Fifth or Sixth Amendment, but not on the Fourth. This is a real distinction, but it is predicated on idiosyncrasies of the collateral review context (e.g., comity and federalism). *See* Stone v. Powell, 428 U.S. 465 *passim* (1976).

[34] The kinds of evidence excluded by testimonial immunity and the kinds excluded by the ER are importantly different, especially as regards causation doctrine, for the rules themselves have different foundations. When a confession is illegally coerced during interrogation, the Fifth Amendment is violated, but when a witness is legally compelled to provide self-incriminating testimony, testimonial immunity ensures there is never a Fifth Amendment violation to begin. These are different topics, and when this chapter refers to the *exclusionary rule* or *ER*, it never refers to the requirements of testimonial immunity.

[35] In *Sanchez-Llamas v. Oregon*, 548 U.S. 331, 348 (2006), the Court explained,

> We have applied the exclusionary rule primarily to deter constitutional violations. . . . The few cases in which we have suppressed evidence for statutory violations . . . [were cases in which] the excluded evidence arose directly out of statutory violations that implicated important Fourth and Fifth Amendment interests. (citing Miller v. United States, 357 U.S. 301, 305 (1958), and McNabb v. United States 318 U.S. 332, 344 (1942)).

Hence, even when the rule is applied in statutory contexts, it is applied because these contexts are intimately connected to constitutional rights.

[36] *See Sanchez-Llamas*, 548 U.S. at 348.

[37] *Id.*; *see* Vienna Convention on Consular Relations & Optional Protocol on Disputes, T.I.A.S. No. 6820 (Dec. 14, 1969).

Note, for simplicity, I'll refer to the conjunction of doctrinal features 4 and 5 as the ER's application to "all and only" constitutional contexts.

6. Good Faith Exception

The constitutional violation that triggers the ER's application must be to some degree *culpable* for the ER to apply.[38] It must be "deliberate, reckless, or grossly negligent," or it must constitute "recurring or systemic negligence."[39] If, in contrast, "an officer [is] acting with objective good faith" in the legality of her actions, the exclusionary rule does not apply.[40] It follows that Cardozo's famous objection to the exclusionary rule—that it wrongly demands the criminal "go free because the constable has blundered"[41]—requires the qualification that a non-negligent or merely negligent blunder does not trigger exclusion at all.

7. Factual (But-For) Causation and the Inevitable Discovery Doctrine

The ER applies only if the constitutional-rights violation was a factual, but-for cause of the state's acquisition of evidence.[42] "If the prosecution can establish by a preponderance of the evidence that the information ultimately or inevitably would have been discovered by lawful means," then the "inevitable discovery doctrine" allows admission of the evidence despite the constitutional violation.[43]

8. Proximate Causation and the Attenuation Doctrine

---

[38] *See* United States v. Leon, 468 U.S. 897 (1984).
[39] Herring v. United States, 555 U.S. 135, 144 (2009).
[40] *Leon*, 468 U.S. at 920.
[41] People v. Defore, 242 N.Y. 13, 21 (1926).
[42] Hudson v. Michigan, 547 U.S. 586, 592 (2006). Much is said later about the difference between factual and but-for causation. *See infra* Section IV.C.2 and note 193. For now, I conflate them for ease of exegesis.
[43] Nix v. Williams, 467 U.S. 431, 444 (1984); *see* Steven P. Grossman, *The Doctrine of Inevitable Discovery: A Plea for Reasonable Limitations*, 92 DICK. L. REV. 313 (1988) (analyzing the inevitable discovery doctrine articulated in *Nix*).

Even when a constitutional violation is the but-for cause of the acquisition of evidence, the causal connection "can be too attenuated to justify exclusion."[44] In other words, the ER applies only if the constitutional violation was also a proximate cause of the state's acquisition of evidence. Proximate causation can be lacking in two ways: the first is "when the causal connection is remote,"[45] and the second is when, "even given a direct causal connection, the interest protected by the constitutional guarantee that has been violated would not be served by suppression of the evidence obtained."[46]

9. Remedy Doctrine: Guilty Verdict Obtained Because of Tainted Evidence

If a criminal court erroneously fails to apply the ER and the defendant is ultimately convicted, the conviction will stand unless it is plausible that the violation was 'prejudicial,' i.e., a but-for cause of the conviction. In doctrinal terms, the exclusionary rule is subject to the harmless-error doctrine applicable to non-structural, constitutional errors.[47] But for our purposes,

---

[44] *Hudson*, 547 U.S. at 592 ("[B]ut-for causality is only a necessary, not a sufficient, condition for suppression.").
[45] *Id*.
[46] *Id.* at 593. The Court elaborated,

> Even in the early days of the exclusionary rule, we declined to "hold that all evidence is 'fruit of the poisonous tree' simply because it would not have come to light *but for* the illegal actions of the police. Rather, the more apt question in such a case is "whether, granting establishment of the primary illegality, the evidence to which instant objection is made has been come at by exploitation of that illegality or instead by means sufficiently distinguishable to be purged of the primary taint.

Id. at 592 (quoting Wong Sun v. United States, 371 U.S. 471, 487-88 (1963)). For a sophisticated overview of the issues posed by this second kind of attenuation, *see* Eric A. Johnson, *Causal Relevance in the Law of Search and Seizure*, 88 BOS. U. L. REV. 113, 116 *passim* (2008).
[47] *See* Arizona v. Fulminante, 499 U.S. 279, 295-96 (1991). This means that on direct review, the *Chapman* standard (error harmless beyond a reasonable doubt) applies if the defense objected to the error below. *See* Chapman v. California, 386 U.S. 18, 21-24 (1967). If the defense did not object, the *Olano* test for plain error applies and the defense must prove *Strickland*-level prejudice ("reasonable probability" of a different outcome). *See* United States v. Olano, 507 U.S. 725, 733-35 (1993) (citing United States v. Bagley, 473 U.S. 667 (1985), and Strickland v. Washington, 466 U.S. 668 (1984)). On federal habeas, *Kotteakos*-level prejudice (substantial and injurious harm) must be shown for most ER violations. *See* Brecht v. Abrahamson, 507 U.S. 619, 619 (1993) (citing Kotteakos v. United States, 328 U.S. 750, 776

we need only understand that although different burdens of proof and production apply depending on the procedural posture, the inquiry is always whether the erroneous failure to apply the ER was a but-for cause of the conviction.

10. Bar on Vicarious Assertions[48]

The one to assert the ER must be the one whose constitutional rights have been violated.[49] If officers illegally search Ann's home and discover evidence incriminating her brother, Bob, who is not living in the home himself, the exclusionary rule will not protect Bob.[50] This goes whether or not the officers were specifically targeting Bob.[51] Ann cannot assert the rule on Bob's behalf, and Bob cannot assert it himself, for he "has not had any of his . . . rights infringed."[52]

B. Explaining the (Actual) Exclusionary Rule

The doctrine described above constitutes the *actual* American exclusionary rule. Because Justices are fallible and make mistakes, we should expect that the actual rule is not perfectly consistent with the rationales, justifications, and foundations that underly it. In other words, we should expect that the *actual* ER is not identical to what I'll call the *corrected* ER, where "corrected" means maximally consistent with its underlying foundations.

(1946)). But as discussed above, *supra* note 14, at Federal Habeas, no amount of prejudice is sufficient to reverse a conviction predicated on an ER violation in the Fourth Amendment context. Doubtless, the language in which these sundry standards are expressed is unhelpful. *See* United States v. Dominguez Benitez, 542 U.S. 74, 82 (2004) (Scalia, J., concurring in judgment) ("Such ineffable gradations of probability seem to me quite beyond the ability of the judicial mind (or any mind) to grasp.").

[48] Note the discussion in this subsection often goes under the heading of "standing doctrine." Because the sort of "standing" at issue in this subsection has nothing directly to do with the standing to blame, I avoid using that terminology here. I also avoid it because the Court has explicitly claimed that for "analytical purpose[s]," the bar on vicarious assertions is "theoretically separate" from the "invariably intertwined concept of standing." Rakas v. Illinois, 439 U.S. 128, 139 (1978).

[49] Jones v. United States, 362 U.S. 257, 261 (1960).

[50] What is it to "live in" the home? The line is generally drawn between a mere invited guest and an overnight invited guest. *See* Minnesota v. Carter, 525 U.S. 83, 89-90 (1998); *Jones*, 362 U.S. at 259.

[51] This is the specific holding of *Rakas*, 439 U.S. 128.

[52] *See id.* at 133-34 (1978) (citing Alderman v. United States, 394 U.S. 165, 174 (1969)).

This chapter's project is to expose what underlies the actual ER. To do so, in Part II, I'll analyze the consistency between the actual ER and a proposed account of what underlies it; in effect, I'll compare the corrected ER, according to the account at issue, with the actual ER. One minor inconsistency is not enough to dismiss an account out of hand because we cannot expect perfect consistency. And indeed, while inconsistency never makes an account a better *explanation*, some inconsistencies make an account a better *contribution*, for those inconsistencies are the pathways along which the doctrine should change and grow.[53]

There comes a point, however, at which inconsistencies are substantial enough to cast serious doubt on the idea that the proposed foundations explain the actual ER.[54] I try to reach that point for the accounts I discuss in Part II, but if and where I fail, I nonetheless claim that the account proffered in Part III and analyzed in Part IV is *more* consistent with the actual ER—a *better explanation* of its contours and character—than any extant alternative.

To be clear, this chapter's explanandum is the actual exclusionary rule. (And note that unqualified references to the ER are always references to the *actual* rule). But while I say this to underline the descriptive and explanatory nature of this project, I do not always focus centrally on the Justices' *actual stated rationales* for their holdings.[55] There are two main reasons for this.

The first is that the "standing to blame" is, in a sense, new. To be sure, the phenomenon predates America, and we all have an intuitive grip on it. But it is *not* the same as the standing to

---

[53] *See* Section IV.B.

[54] When the account cannot explain features of the actual rule, that counts against it. To make this a bit more precise: If the feature is general and often relevant to the rule's application, like the fact it applies only to the criminal domain, this failing counts against the account more heavily than if the feature is 'peripheral' (i.e., specific and restricted in scope, like the fact that Fourth Amendment ER violations don't admit of reversal on federal habeas). When an account's corrected ER *contradicts* a feature of the actual rule (e.g., the corrected rule would apply *more* widely in the civil domain than the criminal), this counts against it more heavily than an account's mere inability to explain the feature (e.g., the corrected rule would apply in both domains).

[55] And I certainly don't claim that this chapter's novel account can be found in the United States Reports.

bring suit, and as far as I am aware, the concept was not 'coined' in philosophical circles until the late twentieth century.[56] In short, we have only recently acquired the conceptual and linguistic apparatus necessary to talk about the standing to blame in a clear way. If the standing to blame guided the evolution of the ER, it did so implicitly, by guiding the Justices' intuitions about right outcomes and holdings. The fact those decisions don't successfully articulate what guided them does not mean they are bad decisions; on the contrary, my suggestion is that they were largely *correct* decisions.[57] But it does suggest that the *holdings*, which together constitute the actual ER, should be our focus.

It also suggests that the stated justifications for those holdings are likely be tortured, inconsistent, unclear, conclusory, and ultimately unhelpful. The fact that this suggestion is widely thought *correct* is the second reason I do not always focus centrally on the Justices' stated rationales.[58] After all, it is because these rationales do so poor a job explaining and unifying ER doctrine that the rule is widely thought to lack any coherent foundation;[59] that the rule is under serious threat of being tossed out altogether;[60] and that a truly *novel* theory of this rule, which has determined defendants' lives and deaths for over a century, is both an urgent and reasonable pursuit.

II. EXTANT ACCOUNTS & INTRACTABLE PROBLEMS

This Part examines extant accounts of the ER in order to show that a novel account is needed. Section II.A focuses on views that find the ER in the Constitution; Section II.B on the

[56] *See infra* note 123 and accompanying text.
[57] This is a key theme of *infra* Part IV.
[58] *See* sources cited *supra* note 3.
[59] *See* sources cited *supra* notes 2-3.
[60] *See* sources cited *supra* note 5.

so-called "Majestic" defenses, typified by the 'judicial integrity' account; and Section II.C on the deterrence rationale. Every account has unique problems, but they all fail to account for the ER's restricted domain of application, particularly the fact it applies just in the *criminal* domain. Overcoming this failure is a key desideratum for the novel account articulated in Part III.

A. Constitutional Defenses

The ER is not stated in the text of the Constitution. Moreover, there is not "a single major statement from the Founding—or even the antebellum or Reconstruction eras—supporting Fourth Amendment exclusion of evidence in a criminal trial."[61] Instead, "Tort law remedies were . . . the ones presupposed by the Framers of the Fourth Amendment and counterpart state constitutional provisions."[62] At the time of the Fourteenth Amendment's ratification, this had not changed: The first case announcing the exclusion of illegally seized property in a criminal trial came fully two decades later, in *Boyd v. United States* in 1886.[63] Because the ER is neither provided in the text of the Constitution nor supported by the history of its ratification, constitutional accounts face an uphill battle.

It's worth asking, for clarity: why don't the substantive rights provided by the Fourth, Fifth, and Sixth Amendments entail an ER? The answer is, first, rights and remedies are different things.[64] Second, nothing in the Constitution says, 'if one of the aforementioned rights is violated, the remedy shall be exclusion of evidence at criminal trial.' And third, once you've unreasonably searched and seized my property, you've *already* violated my Fourth Amendment right: exclusion comes too late to prevent it.

---

[61] Amar, *supra* note 2, at 786.
[62] *Id*.
[63] 116 U.S. 616. The Fourteenth Amendment is sometimes thought to have "updated" the amendments it incorporated against the states, hence why its timing in proximity to *Boyd* is relevant.
[64] *See infra* Section II.B.1 for discussion of attempts to derive the ER from the proposition that rights require remedies.

Compare: the Fifth Amendment right against self-incrimination plausibly does *entail* a kind of 'exclusion' (not at issue in this chapter[65]). After you seize my diary, a criminal court's *later act of admitting it* into evidence is arguably a violation of my Fifth Amendment right against self-incrimination because "introduc[ing] a man's diary as evidence against him is perilously close to forcing him to take the stand himself."[66] Assuming that's so, the court can and should suppress the diary to prevent the Fifth Amendment violation. In contrast, application of the ER cannot prevent a Fourth Amendment search and seizure violation or a Fifth Amendment *Miranda* violation. Indeed, this is why the ER is difficult to explain, and why a constitutional defense of the ER must explain what *other* part of the Constitution requires the exclusion of evidence after the Fourth, Fifth, or Sixth Amendment violation has already been completed.[67]

The most viable attempt points to the due process clauses.[68] On Richard Re's view, "the introduction of illegally obtained evidence . . . [is a] violation of a defendant's right not to be deprived of liberty without due process."[69] So in a paradigm case, the acquisition of evidence violates the Fourth Amendment, and the use of that evidence to obtain a conviction violates the Fourteenth: two distinct violations, the second predicated upon the first.

But views like Re's struggle to provide an independently viable account of due process that fits ER doctrine. The ER has never applied in the civil domain, after all, yet the Fourth Amendment has always been thought to be part of civil procedure, and the due process clause

---

[65] *See supra* note 34 and accompanying text.
[66] Amar, *supra* note 2, at 788.
[67] Amar argues that the ER rests on the mistake of taking the limited role for exclusion in the Fifth Amendment self-incrimination context, confusing and forgetting its constitutional basis, and then applying it baselessly in broader constitutional contexts. *Id.* at 787-89. This is less a way of explaining the rule's justificatory basis than of explaining that it has none.
[68] *See* U.S. Const. amends. V, XIV § 1.
[69] Re, *supra* note 5, at 1917.

applies to civil deprivations of property.[70] The ER has never applied outside the context of constitutional violations,[71] moreover, yet criminal procedure encompasses far more than the Fourth, Fifth, and Sixth Amendments. And illegally obtained evidence can be an essential predicate for a grand jury indictment, a probable cause finding, or a more severe sentence; these are all parts of the process that results in the deprivation of liberty, yet the ER applies to none of them.[72]

Re's account also struggles to explain *why* the introduction of evidence obtained illegally violates due process. It is equally conceivable that the standards of due process require, for example, admitting any and all reliable evidence, for only this provides litigants with the *accurate* verdict *that they are due*. Mysteriously, Re invokes the language of *authority*, suggesting that to admit tainted evidence is to "deprive[] the jury of *authority* to enter a conviction . . . [and] deprive[] the government of *authority* to defend the conviction on appeal."[73] But he does not explain what this means—and that is the whole ball game.

The account this chapter offers in Parts III and IV can be seen understood to articulate the relevant "authority" in terms of the standing to blame.[74] Charitably, Re might have had this same idea in mind, but no way to articulate or defend it, because he was bound by the due process clause, whose breadth of coverage left him with an overinclusive account and no coherent way to

[70] As Re notes, the Fourth Amendment was originally understood to be exclusively part of civil (tort) process, and civil proceedings "can obviously result in deprivations of liberty or property." *Id.* at 1938-39. This suggests civil exclusion should have been the rule *long before* criminal exclusion. Moreover, Re acknowledges, "the Due Process Clauses were originally understood to command adherence to all positive-law procedures for depriving persons of life, liberty, or property," making the ER's non-application in the civil context all the more inexplicable. *Id.* at 1941.
[71] *See supra* Section I.A.4-5.
[72] *See also* Albert W. Alschuler, *Regarding Re's Revisionism*, 127 HARV. L. REV. F. 302 *passim* (2014) (raising related concerns about Re's view).
[73] Re, *supra* note 5, at 1941 (emphases added).
[74] *See also infra* Section V.A.4 (discussing whether the novel account offered in Parts III and IV might have constitutional implications, perhaps in conjunction with something like Re's view).

place blame or punishment at its center. The problem for constitutional accounts of the ER is quite simply that the ER is not given to us by the Constitution, so accounts like Re's are ineluctably bound by an inadequate explanatory basis.

B. Majestic Accounts

The so-called "majestic" accounts claim that principles of justice or integrity require the ER. The Constitution's importance may be part of the reason justice or integrity demands we pay the ER's high price, but the Constitution itself is not what explains the rule. This Section details and criticizes four of the most plausible majestic accounts, omitting discussion of their weaker variations.[75]

1. The Only Suitable Remedy

Anglo-American jurisprudence has long endorsed the principle that rights require remedies.[76] But when an officer violates an individual's constitutional rights, civil damages and criminal liabilities are generally blocked by qualified immunity doctrine.[77] The remedial account

---

[75] Many of these variations take the form of slogans that state obvious truths but fall short of coherently arguing for the ER, e.g., *no one is above the law*. Consider how such an argument would go: When the government or its agents violate a person's constitutional rights, they break the law. The government and its agents are not above the law. So—what?—the admission of evidence obtained through such a violation must be proscribed? That is a non-sequitur. The lacuna could be filled in in a number of ways—*see infra* Section II.B.4—but by itself, the slogan is not an argument. *See also infra* note 88 (discussing other non-arguments for the ER).

[76] *See* Mapp v. Ohio, 367 U.S. 643, 655 (1961) ("[Without the exclusionary rule,] the assurance against unreasonable federal searches and seizures would be 'a form of words', valueless and undeserving of mention."); Amar, *supra* note 2, at 812 ("[F]or every right there must be a remedy."); Donald H. Zeigler, *Rights Require Remedies: A New Approach to the Enforcement of Rights in the Federal Courts*, 38 HASTINGS L.J. 665, 667 (1987) ("The ancient idea that rights must have remedies has played a significant role in English and American legal history."); *id*. at 678 ("[A] right without a remedy is not a legal right; it is merely a hope or wish.").

[77] *See* GREGORY M. CASKEY, CALIFORNIA SEARCH & SEIZURE § 1.1 (2020) ("[G]iven the expansion of qualified immunity for judges and law enforcement officers, the remedy of a civil action for damages would appear illusory in all but the most egregious cases."); Kit Kinports, *The Supreme Court's Quiet Expansion of Qualified Immunity*, 100 MINN. L. REV. 62, 64 (2016) ("[T]he Supreme Court's qualified immunity opinions . . . have made a *sub silentio* assault on constitutional tort suits."); *see also* John A. Braden, *Constitutional Tort Liability*, MICH. B.J., 7-97, at 680 (1997) ("The last few decades have seen a veritable deluge of tort immunity laws.").

draws on these facts: There must be a remedy for state violations of Fourth, Fifth, and Sixth Amendment rights, and in light of qualified immunity, the ER is the best—and perhaps *only*—practically available option. If the ER seems an extreme and unusual remedy, this is only because of the profundity of individuals' constitutional rights and the absence of available alternatives.

But if the remedial account were correct, then we'd have an ER in the civil context, where Fourth, Fifth, and Sixth Amendment violations have no remedy by the account's own lights. After all, commentators and the Court have long said that the ER imposes such substantial costs because it requires the release of dangerous, guilty *criminals*.[78] If the principle that rights require remedies is important enough to require the release of such criminals, then it is important enough to dismiss civil actions for lack of admissible evidence.

What's more, if the remedial account were correct, the ER would be a *bad* remedy. For the ER does not—perhaps cannot—apply when a person's constitutional rights are violated and no incriminating evidence turns up. This will happen especially often when a suspect is innocent, so the innocent will be disproportionately left without a remedy.[79] The ER is no remedy for a

[78] *See, e.g.*, Herring v. United States, 555 U.S. 135, 141 (2009) ("The principal cost of applying the rule is, of course, letting guilty and possibly dangerous defendants go free—something that 'offends basic concepts of the criminal justice system.'" (quoting United States v. Leon, 468 U.S. 897, 908 (1984))); *Leon*, 468 U.S. at 907 ("The substantial social costs exacted by the exclusionary rule . . . have long been a source of concern."); United States v. Janis, 428 U.S. 433, 448 (1976) ("Jurists and scholars uniformly have recognized that the exclusionary rule imposes a substantial cost on the societal interest in law enforcement."); Harry M. Caldwell & Carol A. Chase, *The Unruly Exclusionary Rule: Heeding Justice Blackmun's Call to Examine the Rule in Light of Changing Judicial Understanding About Its Effects Outside the Courtroom*, 78 MARQ. L. REV. 45, 49 (1994) ("Let there be no question that the costs of the Exclusionary Rule are high. The costs run the gamut, from lost prosecutions, to public hostility directed at the entire criminal justice system, to monetary burdens due to the increased cost of litigation, to occasional police coverups for their misconduct, and finally to the prostituting impact the Rule has on the Fourth Amendment."); Luke M. Milligan, *Police Transparency and the Exclusionary Rule*, 58 U. LOUISVILLE L. REV. 467, 471 (2020) ("No one denies the exclusionary rule imposes high costs on society.").

[79] *See* Amar, *supra* note 2, at 796-97 (stressing this point).

black homeowner whose home is searched and ransacked because he is black, for instance. Yet if the principle that rights require remedies is important enough to require an exclusionary rule, we should wonder why it is not important enough to find a remedy for those whose violations only evince their innocence.

An advocate for the remedial account would insist that the ER is the only practically available remedy and that it's better some people have a remedy than no one does. But we already know there is another option, one presupposed by the Founders: civil damages. The fact that the courts refuse to take that option, and instead employ an exclusionary rule that they've long described as overwhelmingly costly, suggests that the remedial account is mistaken about what's driving the ER. More generally, if the principle that rights require remedies were so important to our law, we would be able to exact civil damages. That we cannot suggests the principle is not as important as the remedial account is committed to claiming. And this is no big surprise, for the law seems to violate the principle in a wide variety of contexts.[80]

These are already serious problems for the remedial account. But its most serious fault is that exclusion does not fit the legal concept of a *remedy*.[81] To see why, imagine that qualified immunity did not exist and civil damages suits were viable. An officer illegally breaks down a suspect's door and seizes incriminating evidence. The evidence is used to sentence the suspect to a decade of prison.

---

[80] Qualified immunity is but one of many doctrines that put technically available remedies out of reach. *See* Zeigler, *supra* note 76, at 665-66 ("[T]he chief legacy of the Burger Court may be the creation of impediments to the enforcement of rights. Cases that restrict access to the federal courts for vindication of constitutional rights against state official action are clear examples of such obstructions but they are only part of a broader pattern of decisions that impede enforcement of statutory and common-law rights. In most instances, the Court has not diluted the content or substance of rights directly; instead it has created procedural barriers that leave victims without a practical remedy.").

[81] The points made in this and the next paragraph are drawn in part from Amar, *supra* note 3, at 794-96.

Now, whenever the suspect sues the officer in civil court, he can expect to be compensated for the value of the broken door. But the value of the incriminating evidence? Perhaps if it was a family heirloom—but not if it was methamphetamine or bodily organs worth millions on the black market. And here we have our first mismatch between the concept of a remedy and the actual ER. The ER draws no distinction between the heirloom and the contraband; it excludes them whether or not their loss would be remediable in a civil action.

Could the suspect expect to be compensated for the value of the 10 years he's forced to spend in jail? That seems odd, and not just from a moral perspective. Consequential damages for constitutional violations generally turn on whether the consequential damages at issue constitute the kind of "risk" that the right is meant to protect against.[82] So under the remedial principles we'd expect to apply, the suspect will be compensated for the 10 years just in case *imprisonment* is the kind of risk that the Fourth Amendment is meant to protect against. And no one thinks that, least of all the Court.[83] The Fourth Amendment protects privacy and the sanctity of the home,[84] not the freedom of those who hide evidence there, so the suspect would receive nothing for his ten years. The ER, in contrast, gives him those ten years back in full.[85]

---

[82] John C. Jeffries, Jr., *Damages for Constitutional Violations: The Relation of Risk to Injury in Torts*, 75 VA. L. REV. 1461, 1474 (1989).

[83] *See* Hudson v. Michigan, 547 U.S. 586, 594 (2006) (describing the interests protected by the Fourth Amendment, which "do not include the shielding of potential evidence from the government's eyes" or "one's interest in preventing the government from seeing or taking evidence").

[84] *See id*.

[85] Amar, *supra* note 2, at 795, makes a similar point:

> Not all but-for consequences of an illegal search are legally cognizable. If the police buy fancy surveillance equipment from HiTek, Inc. to conduct illegal searches, and competitor Low-Tech Corp. is thereby driven into bankruptcy, Low-Tech's *factual* harm—though (by stipulation) a but-for consequence of the illegal searches—does not constitute *legally* cognizable injury. . . . [This] is the residue of a two-century tradition of civil-damage actions in America.

As a remedy, the exclusionary rule should do less: it should not compensate for the lost contraband or the ten-year sentence. Yet as we've seen, it should also do much more: it should cover civil violations, and if it is indeed *the* remedy for Fourth, Fifth, and Sixth Amendment rights violations, it should not serve only a small subset of victims. However much the exclusionary rule resembles a remedy on the surface, it is not well understood *as* a remedy, and the principle that rights require remedies is not its explanation.

2. *Restoration of the Status Quo Ante*

Another majestic account suggests that the ER operates to restore the state of affairs that would have obtained had the constitutional violation never occurred. This 'restoration' account is similar to the remedial account, but it does not suggest the ER *is* a remedy. This is an advance, in one respect: just as the ER excludes more than what's normally remedial, so restoration of the status quo ante requires more than merely remedying the violation.

But many of the remedial account's problems remain. First, if restoration of the status quo ante is so important in the criminal context, why not the civil? The same Amendment is violated in the same way regardless of what the evidence is used for. Second, our innocent black homeowner still has no recourse for his violation—restoration of the status quo ante is important enough to require the release of guilty criminals but not, apparently, to fix his door. This is silly, and it underlines the fact that our law does not take restoration of the status quo ante to be so important that it requires we pay the high cost of the ER. And third, just as the remedial account struggles with contraband evidence, so the restoration account would require returning it.

The restoration account also faces additional problems *because* restoration goes beyond remedy. Restoring the status quo ante would entail returning not only the cocaine and organs, but hostages too, so the advocate of this defense simply cannot stand by the restoration account as

stated.[86] Moreover, the breadth of restoration entails that the tainted evidence should be inadmissible at every point in the criminal process. The account thus cannot explain the ER's non-application in pretrial, impeachment, sentencing, and probation-revocation contexts. As an alternative to the remedial view, this defense trades one set of problems for another, equally damning.

3. The Government Should Not Profit from Its Wrongs

An alternative 'no-profit' account rests on the principle that the state should not profit from its wrongdoing. It adds, plausibly enough, that the government profits when it achieves guilty verdicts. The ER blocks the state from achieving guilty verdicts on the basis of evidence it acquired through its wrongdoing, and thus ensures "that the government would not profit from its lawless behavior."[87]

Initially, this account seems to have some hope of explaining the ER's non-application in the civil context. When Joe obtains damages against Sue using evidence the state acquired by wronging Sue, it's *Joe* who profits, not the state, so the ER need not apply. But this does not get us far, for when the state is the plaintiff in a civil suit, it surely profits when its agents illegally seize evidence on the basis of which it receives a favorable verdict and a large pay-out. If the concern were with the government literally *profiting* from its own wrongs, this should be a paradigmatic case in which exclusion is applied, not one in which it isn't.

A similar problem arises from the fact the state can use tainted evidence in grand jury proceedings and at sentencing. If the government profits from a conviction, why does it not profit when it wins a favorable grand jury decision, or when the judge sentences the defendant to

[86] The question then becomes where to draw the line between the bloody knife and the hostage (and why the knife must be excluded if exclusion will lead to more hostages).
[87] Herring v. United States, 555 U.S. 135, 152 (2009) (Ginsburg, J., dissenting).

the long sentence the government fought for? We might ask similar questions in the contexts of impeachment, parole hearings, and deportation hearings in which the government uses tainted evidence against a party to achieve the result it favors. And we might also ask why the ER applies only in the context of constitutional violations. Oregon doubtless wronged Sanchez-Llamas when it interrogated him in violation of the Vienna Convention, after all. But Oregon was allowed to 'profit from that wrong,' just like any governmental entity when the wrong is not constitutional in nature.

As a final nail in the coffin of this account, consider the bar on vicarious assertions. If the government illegally searches Ann's home and finds evidence that incriminates her, the ER will stop the government from profiting from this wrong. But if the government illegally searches Ann's home and finds evidence that incriminates Bob, no one can raise the exclusionary rule to suppress that evidence. But the government profits from its own wrongdoing just as much whether it uses its fruits against Bob or against Ann. There is no relevant difference in terms of its profit, yet the doctrinal difference is manifest.

The principle underlying the no-profit account is plausible. But even if it could explain *an* ER, it does not explain *our* ER, whose doctrinal features it leaves unexplained and inexplicable.

4. Judicial Integrity

Judicial integrity is a vague concept that can support an array of bad or question-begging arguments for the ER.[88] But in its strongest form, the 'judicial integrity account' begins by

[88] Justice Brandeis suggested several of them when he spelled out the sense in which the ER "preserves the judicial process from contamination":

> Decency, security, and liberty alike demand that government officials shall be subjected to the same rules of conduct that are commands to the citizen. In a government of laws, existence of the government will be imperiled if it fails to observe the law scrupulously. Our government is the potent, the omnipresent teacher. For good or for ill, it teaches the whole people by its example. Crime is contagious. If the government becomes a

noting that the government and its agents often violate suspects' Fourth, Fifth, and Sixth Amendment rights in the course of working toward conviction. These violations often produce incriminating evidence that serves the government's goal. If courts systematically accept this evidence in criminal trials, they give the government and its agents reason to continue their violations: the courts *make it the case* that the government's "willful disobedience of [the] Constitution"[89] is a *success*. And this means two things:

First, it means the courts are *accessories*. Courts know about the government's systemic lawbreaking. And they know it is aimed at obtaining convictions. Courts could avoid

---

> lawbreaker, it breeds contempt for law; it invites every man to become a law unto himself; it invites anarchy. To declare that in the administration of the criminal law the end justifies the means-to declare that the government may commit crimes in order to secure the conviction of a private criminal-would bring terrible retribution. Against that pernicious doctrine this court should resolutely set its face.

Olmstead v. United States, 277 U.S. 438, 485 (1928), overruled in part by Berger v. State of N.Y., 388 U.S. 41 (1967), and overruled in part by Katz v. United States, 389 U.S. 347 (1967).

We might extract a number of arguments from this text. A first is that without exclusion, "the government becomes a lawbreaker." But unless this is a point about deterrence, it is straightforwardly question begging—it presupposes that admitting the evidence is lawbreaking. A second is that without exclusion, the government will violate citizens' constitutional rights *more* (and so society will crumble, etc.). But this is certainly a version deterrence rationale (*cum* catastrophizing) discussed below, so if this is what judicial integrity comes to, all the worse for judicial integrity. *See infra* Section II.C.

A third argument is that without exclusion, the government "declare[s]" that "the end justifies the means . . . that the government may commit crimes in order to secure the conviction of a private criminal." But this cannot be taken literally, else it is confused. The question is what to do with reliable evidence obtained *after and because* the government has *already* committed a crime. It is not whether "the government may commit crimes" and violate citizens' constitutional rights "in order to secure the[ir] conviction." No one thinks it may commit crimes in order to secure criminals' convictions.

A fourth argument is that, however confused the thought may be, citizens think that to forgo the ER is "[t]o declare . . . that the ends justify the means." And because the government is the "potent . . . ever present teacher" that "teaches the whole people by example," people will take this declaration seriously. But this is an empirical prediction that's hard to swallow given that some countries do not impose exclusion and that the United States imposes it in only a limited set of situations in which the government violates citizens' rights.

Ultimately, Brandeis's exposition of "judicial integrity" is a set suggestive claims that do not by themselves add up to a viable justification for the rule. A key worry for any invocation of "judicial integrity" is that it will amount to just that: a high-minded phrase that sounds in profundity but cannot support a sound and valid argument.

[89] Elkins v. United States, 364 U.S. 206, 223 (1960) (A conviction predicated on tainted evidence "cannot be allowed to stand without making the courts themselves accomplices in willful disobedience of law.").

contributing to this criminal state of affairs by refusing to accept illegally obtained evidence. Their refusal would break the means-ends connection between the government's lawbreaking and its pursuit of convictions.[90] Or, alternatively, courts could perpetuate the lawbreaking by accepting the illegally obtained evidence. They would therefore ensure that lawbreaking remains instrumentally rational for the state, and that every state agent will have a good reason to violate suspects' rights—namely, and perversely, the *pursuit of justice*.

Second, it means that courts *express condonation* of the government's constitutional violations. To condone is not to approve or to declare that the government is above the law. To condone, rather, is to overlook or impliedly forgive the offense. It's to express something like, 'constitutional rights are not so important, and violations of constitutional rights are not so serious, that they require the courts to *do anything* about this state of affairs it contributes to.'

So understood, the judicial integrity account makes a good case. It provides a plausible answer to the question why the admission of evidence is a wrong that goes beyond the initial constitutional violation, namely, that admitting the evidence knowingly perpetuates and expressively condones the state's systemic pattern of violating constitutional rights in pursuit of convictions. The account also has some resources to explain the ER's domain of application: it applies specifically to constitutional violations, for example, because those rights are of such great importance that the cost of condoning their violation outweighs the cost of releasing guilty criminals.

But the account is ultimately unsatisfactory. First, there is no reason judicial integrity is sufficiently weighty for exclusion *only* in the guilt phase of criminal trial, so no clear explanation

[90] Or, at least, it would go some way toward breaking that connection. Even an ER that is vastly broader than our own cannot obviate the rationality or efficacy of the state's deceiving the courts about where they got the evidence.

for the ER's non-application to grand-jury, probable-cause, impeachment, and probation/parole-revocation contexts. The situation is worse for the ER's non-application at sentencing, which is almost entirely within the *judiciary's* purview and thus where considerations of judicial integrity should be at their strongest.

Second, the integrity account might be able to explain why the ER does not apply in civil tort actions, where the state cares little about who wins. But it does remarkably poorly in other civil contexts, like deportation, where the same state agencies conduct searches and seizures of suspects *and* fight in court to have them deported.[91] The need to break the state's means-ends connection between constitutional violations and *convictions* is no less strong when we replace convictions with *deportations*, yet the ER applies only in the criminal context.

Third, the bar on vicarious assertions is inexplicable on the judicial integrity account. If courts are accessories that express condonation of the governments' violations when they accept tainted evidence, this should go whether or not the evidence incriminates the same person who was violated. The overarching problem for the judicial integrity account is that ER doctrine does not track the contexts in which the judiciary's integrity is more or less under threat. The account tells us the rule is a tool to protect the courts' virtue, but the doctrine—especially the bar on

[91] *See* I.N.S. v. Lopez-Mendoza, 468 U.S. 1032, 1053 (1984) (White, J., dissenting) ("The majority correctly acknowledges that the 'primary objective' of the INS agent is 'to use evidence in the civil deportation proceeding.' . . . The Immigration and Naturalization Service likewise concedes that INS agents are 'in the business of conducting searches for and seizures of illegal aliens for the purpose of bringing about their deportation.'" (citations omitted)); CARDONZO IMMIGRATION JUSTICE CLINIC, CONSTITUTION ON ICE: A REPORT ON IMMIGRATION HOME RAID OPERATIONS 1 (2009) ("[F]requent accounts in the media and in legal filings have told a similar story of constitutional violations occurring during ICE home raids—a story that includes ICE agents breaking into homes and seizing all occupants without legal basis."); *id*. at 11 (arguing that "the data is extremely consistent and suggestive of widespread Fourth Amendment violations by [ICE] agents" from the two states analyzed in the report, New York and New Jersey).

vicarious liability—tells us that the *victim* and what we owe to *her in particular* is far more central to this story.

A successful account of the ER must not simply adduce a lofty ideal; it must also explain doctrinal reality. The judicial integrity account, whatever we might say for it as a justification for *some* ER, is an inadequate explanation for *the* ER. Indeed, this explanatory failing is common across the extant majestic accounts.

C. The Deterrence Rationale

The final extant rationale for the ER is *deterrence*, according to which the ER exists to deter state violations of constitutional rights. The rule's contours are determined by balancing its deterrent value (fewer violations of constitutional rights) with the costs of its application (more criminals going free).[92] Today, deterrence is the Court's sole doctrinal rationale for the exclusionary rule[93]—and it is likely the weakest rationale we've considered.

The first premise of the deterrence rationale is that state agents are conviction-driven: motivated to commit constitutional violations by their desire to obtain guilty verdicts against criminals. If an officer considers illegally searching a house because he expects to find evidence that can be used to convict, his knowledge of the ER should deter him. In principle, if he goes through with his plan and finds the evidence he seeks, the ER will suppress it. There is, in turn, *no point* for the conviction-driven officer to commit the violation.[94]

---

[92] *See* Davis v. United States, 564 U.S. 229, 237-38 (2011) (outlining this balancing regime).
[93] *Id*. at 236-37 ("The rule's sole purpose, we have repeatedly held, is to deter future Fourth Amendment violations.").
[94] *See, e.g.*, Elkins v. United States, 364 U.S. 206, 217 (1960) ("[T]he rule is calculated to prevent . . . . Its purpose is to deter—to compel respect for the constitutional guarantee in the only effective way—*by removing the incentive to disregard it*." (emphasis added)).

Consider how the Court uses these ideas to explain why the good faith exception draws a line between gross and mere negligence.[95] The cost of applying the ER is the same whether the officer acted with gross or mere negligence, for a criminal's culpability has nothing to do with the officer's negligence. But the deterrent value is different. Grossly negligent conduct is "sufficiently deliberate that exclusion can meaningfully deter it," whereas merely negligent is not.[96] (That previous sentence might reflect confusion.[97] But it's plausible on the surface.) And since the ER's cost is fixed here, it's possible that the scales tip toward its application in cases of gross but not mere negligence. Perhaps not a gold-star explanation, but it could be worse.

Deterrence as a rationale for the ER operates quite unlike deterrence as a rationale for punishment. In the latter context, deterrence is primarily about *imposing cost*, not *removing benefit*. But if the exclusionary rule is so costly[98]—vicious criminals going free on bases unrelated to their culpability—why *not* deter constitutional violations by imposing costs on violators, instead? We already know the answer: these are not real options because of qualified immunity doctrine and restrictive readings of criminal laws applicable to state officers.[99]

Even if we grant this, however, the problem for the deterrence rationale is that it fails abjectly to explain the rule we have. In what follows, I illustrate the rationale's explanatory

---

[95] *See Davis*, 564 U.S. 229; Herring v. United States, 555 U.S. 135 (2009). *But see* Kit Kinports, *supra* note 3 (arguing that the deterrence rationale cannot explain the good faith exception).
[96] *Herring*, 555 U.S. at 144.
[97] *See* MODEL PENAL CODE § 2.02(d) (defining negligence in terms of lack of awareness). If the Court's idea is that grossly negligent conduct is more likely to involve awareness that the ER might apply, it's in tension with the MPC's definition of negligence. *But cf.* Arizona v. Evans, 514 U.S. 1, 29 n.5 (1995) (Ginsburg, J., dissenting) ("It has been suggested that an exclusionary rule cannot deter carelessness, but can affect only intentional or reckless misconduct. This suggestion runs counter to a premise underlying all of negligence law—that imposing liability for negligence, i.e., lack of due care, creates an incentive to act with greater care."). Of course, this dictum from the dissent in *Evans* replaces one problem for another by suggesting that *mere* negligence can *also* be deterred.
[98] *See* sources cited *supra* note 78.
[99] *See* 18 U.S.C. § 242: Deprivation of Rights Under Color of Law; sources cited *supra* note 77 (discussing qualified immunity); *see also infra* note 111 (explaining why it does not follow from this answer that the Court can legitimately incorporate an ER against the states.).

ineptitude with respect to its bar on vicarious assertions, its inevitable discovery doctrine, and its inapplicability to the civil domain.

Start with vicarious assertions.[100] If Officer Owen considers breaking into Ann's house because he thinks he might find incriminating evidence against her, he should pause: if he's successful, exclusion applies. But if Owen considers breaking into Ann's house because he thinks he might find incriminating evidence against Bob, who does not live there, the ER shouldn't cross his mind: if he's successful, exclusion does not apply. Yet Owen commits exactly one Fourth Amendment violation, of the same magnitude, with the same degree of culpability, and against the same person, either way. So if the ER aims to deter Owen from committing *precisely that kind of violation*, what explains the asymmetry? There is no reason to think the deterrent value of exclusion depends on whether the one violated is also the one inculpated, and Owen violates Ann's Fourth Amendment rights to the same extent whether he aims to find evidence against her or against Bob. Nor is there reason to think that the Bobs of the world are more culpable, and so more costly to free, than the Anns.

On the contrary, there *is* reason to think that creating such an asymmetry will reduce the rule's deterrent force.[101] It's reasonable to think, after all, that with every inexplicable distinction, the rule becomes more complicated and thus more muddled in the minds of state

[100] For a related but different argument that the bar on vicarious assertions is incompatible with the deterrence rationale, *see* Sherry F. Colb, *Standing Room Only: Why Fourth Amendment Exclusion and Standing No Longer Logically Coexist*, 35 RUTGERS L. SCH. (NEWARK) FAC. ARTICLES 1, 18 (2006) (arguing that the question whether the one who raises an ER motion is also the one whose rights were violated "does not rest on the accessibility of predicate facts to police considering a search," which makes it irrelevant to deterrence).

[101] *See* Danielle Hayes, *supra* note 3, at 626 *passim* (2018) (describing "studies suggest[ing] that, after the exclusionary rule was applied under a deterrence rationale and exceptions were created, officers were confused about what was constitutionally required of them, which inhibited the rule's ability to deter misconduct").

actors. These are largely police officers with little legal training who must act in the heat of the moment and who cannot be expected to keep track of all the rule's permutations.

What's more, it is reasonable to suspect that by allowing Owen to get dirt on Bob by violating Ann's rights, Owen will have more reason to violate Ann's rights even when no particular "Bob" is his target. If Owen suspects Ann of criminality, he may well think it likely that she possesses evidence that inculpates her co-conspirators, especially friends and family.[102] Owen would not be irrational to think, first, there's always a chance the evidence he finds against Ann will be admitted in violation of the rule, whether by legal mistake or by judge who can't countenance applying the ER in the case at bar;[103] second, there's always a chance that some seemingly arbitrary exception or another will keep the rule from applying to the evidence he finds; and third, there's always a chance he'll find evidence that inculpates Ann's friends and family, leading to more convictions and, indeed, additional leverage against Ann. The bar on vicarious assertions contributes to Owen's second thought and gives him his third, directly vitiating the rule's deterrent force even when no particular "Bob" is in the picture.

The most plausible reply from the defender of deterrence would seem to cite *administrability costs*. When the one violated is also the one inculpated, the one violated will already be in court, ready, willing, and able to plead the facts of the violation. But when the one violated (Ann) is not also the one inculpated (Bob), Bob will have to haul Ann into court, compounding the intrusions into her privacy. Poor Ann! First her house is searched illegally, then she has to come to court to testify about it. (Of course, none of this has to do with deterrence. But it's fair game: just as the value of deterrence may be weighed against the

[102] *Cf.* CARDONZO IMMIGRATION JUSTICE CLINIC, *supra* note 91, at 11 ("[O]nly approximately one-third of all home raid arrests [by ICE] are of targets. The remaining two-thirds of the arrests are of civil immigration violators who ICE happens to encounter during home raid operations.").
[103] *See infra* note 183 and accompanying text.

disvalue of freeing the guilty, so it may be weighed against the disvalue of administrative difficulty.[104])

Even still, the administrability reply is tenuous—as if allowing the defendant to subpoena Ann were so very difficult for the court, or for Ann, that it outweighs the cost of drawing a muddled distinction that vitiates the deterrent force of the exclusionary rule. Moreover, given that probable cause is a matter of the officer's subjective standpoint, Ann's testimony will often be irrelevant: it is the officer's point of view that matters.[105] What's more, by placing a bar on vicarious assertions, the courts complicate every exclusion inquiry with the additional question of whether the one violated is the one inculpated. That is not a simple inquiry; in the case of homes, for instance, the line is drawn between a mere invited guest and an *overnight* invited guest.[106] And unlike the requisite inquiry into the author's subjective standpoint, this inquiry will almost necessarily require the homeowner's involvement, for she alone will authoritatively know if the defendant was invited and the invitation's length. In sum, the ER would not just be a better deterrent mechanism without its bar on vicarious assertions; it would be *more administrable*, too. The deterrence rationale simply has no decent story for the bar on vicarious assertions.

And things get worse. Consider the inevitable discovery doctrine. If the purpose of the exclusionary rule is to deter constitutional violations, what does it matter whether the evidence found in an illegal search would have been found anyway? An officer considering committing such a violation is rarely in a position to know the double counterfactual—whether the evidence he *would* find if he searches is evidence that *would be found* even if he doesn't. (Permutations upon permutations for the officer, who must act quickly.)

---

[104] *See* Hudson v. Michigan, 547 U.S. 586, 593-94 (2006) (listing various interests that might be considered).
[105] *See* Colb, *supra* note 100, at 15-19.
[106] *See* Minnesota v. Carter, 525 U.S. 83, 89-90 (1998); Jones v. United States, 362 U.S. 257, 259 (1960).

But what of the officers who think they *do* know the answer to the double counterfactual? These officers very well might be influenced by the existence of an inevitable discovery doctrine. Indeed, if the doctrine *prohibited* the admission of unconstitutionally obtained evidence *despite* its inevitable discovery, this would be a powerful disincentive for the conviction-driven officer to commit his violation: by going through with the search, he might effectively *eliminate* evidence that would otherwise be found legally and used successfully to convict. His violation could be the reason the guilty defendant goes free.

Yet the inevitable discovery doctrine is *precisely the opposite*. If the conviction-driven officer thinks the fruit of his unconstitutional search will be found inevitably, then there no risk if he acquires it *now* through an unconstitutional search: the evidence will come in either way. On the other hand, if he thinks the evidence will not inevitably be found, then his search may be the only chance for the evidence to come into play. It follows that the inevitable discovery doctrine tells the conviction driven officer that the benefit of his search *will* be realized unless it was *never* going to be realized. On the deterrence rationale, the ER is meant to tell Officer Owen that there is nothing to gain by violating the suspect's rights; instead, the inevitable discovery doctrine tells him there is nothing to *lose*. Deterrence cannot explain the inevitable discovery doctrine.[107]

Consider, finally, the rule's restricted application to the criminal domain. The exclusionary rule is thought to be so costly because it demands that rapists and murderers go

[107] *See* Harold S. Novikoff, *The Inevitable Discovery Exception to the Constitutional Exclusionary Rules*, 74 COLUM. L. REV. 88, 99-100 (1974) ("[The inevitable discovery doctrine] can only encourage police shortcuts whenever evidence may be more readily obtained by illegal than by legal means. This, of course, is the opposite of the purpose of the exclusionary rule: to deter law enforcement officers from using illegal methods to procure evidence.").

free—that *justice* be undone. This suggests that exclusion should be the rule in the civil context *at least* if it's a rule in the criminal.

So why isn't it? The Court has suggested that officers are not so concerned to see the fruits of their searches used against tortfeasors in the civil setting.[108] That might be true. And it *might* follow that the deterrent effect is weakened so much that the cost of exclusion, though itself discounted in the civil context, is no longer worth paying. But even if we accept this rationale in the context of tort suits, it certainly doesn't work for the INS and ICE, for whom deportation is the goal, and dubious midnight raids are the means to succeed.[109] If what explains the ER's non-application in the private tort context is the fact that violating officers don't care about private victims receiving compensation, the ER should certainly apply when the INS or ICE violates the constitution to obtain deportations. The fact it does not apply there *either* suggests that the Court's wildly speculative balancing of costs and deterrent benefits is, at best, post hoc justification for something whose true explanation lies elsewhere.[110]

We could continue through the deterrence rationale's explanatory failings, but what's been said should be sufficient to conclude that deterrence does not explain the ER.[111] More

---

[108] *See* United States v. Janis, 428 U.S. 433, 458 (1976) ("[A civil action] falls outside the offending officer's zone of primary interest.").

[109] *See* sources cited notes 91.

[110] *See also* WAYNE R. LAFAVE ET AL., 2 CRIMINAL PROCEDURE § 3.1(g) (4th ed. 2020) ("An extreme and fundamentally unsound cost-benefit analysis was utilized by the majority in *Immigration and Naturalization Service v. Lopez-Mendoza*.").

[111] And supporters of the rule should be glad of that. For if deterrence did explain the rule, the ER would be an abuse of power. We need no empirical evidence to see why. On the story of deterrence, the Court noticed that many states and state actors were violating the laws of the land, so it *created a new law*—the ER—in order to *ensure that the laws of the land are followed*. In other words, it usurped the legislative power to usurp the executive function. It then incorporated its new law against every state using the authority of the Constitution, undermining any possible defense that it was merely exercising its supervisory power over the federal courts. This is a scandalous story—and slanderous because false. With this in mind, it is remarkable in fact that the Court claims deterrence is the sole rationale for the exclusionary rule. It is so remarkable, in fact, that its sincerity is suspect. More likely, what many believe, both on the Court and off, is that the rule is a mistaken mess all the way down. By claiming that

broadly, what's been said in Part II should be sufficient to conclude that no extant account provides an explanatory adequate theory of the rule we actually have. Given this state of affairs and the ER's widely acknowledged costs, it is little wonder that the rule is in peril.

III. PUNISHMENT AND THE STANDING TO BLAME

With the failings of extant accounts of the ER in mind, this Part presents the chapter's novel account of the rule. It does so by offering a "bottom-up" explanation, showing how the rule follows from foundational premises about blame and punishment. This bottom-up explanation clarifies the account and complements the "top-down" argument in Part IV, which aims to show that the novel account provides the best available explanation for the rule.

Section III.A discusses blame and introduces the standing to blame. Section III.B explains the standing to blame in detail. Section III.C introduces what I call the *epistemic illicitness* condition on the standing to blame. Loosely: you lack standing to blame me for my culpable wrongdoing if you are aware of my culpable wrongdoing only because you culpably wronged me and your wrongdoing is at least as blameworthy as my own culpable wrongdoing. Section III.D argues that the state lacks the standing to *punish* a wrongdoer if it lacks the standing to *blame* her. Finally, Section III.E explains that the epistemic illicitness condition requires an ER to block the state from illegitimately punishing a guilty criminal without the standing to do so. It also provides the official formulation of the chapter's novel account.

A. Blame as the Missing Link

The accounts of the ER discussed in Part II all failed to explain why the rule is exclusive to the criminal domain. We should therefore consider whether it is something about the criminal

deterrence justifies the rule, the Court moves ever closer to the conclusion that the mistake is indefensible, the mess is irreparable, and the ER is *overruled*.

domain in particular that gives rise to the rule. The defining feature of that domain, of course, is *punishment*. Any account that explains the rule as something that arises uniquely out of punishment will be well situated to explain why the rule is exclusive to the criminal domain.[112]

This insight, however, is hardly sophisticated. Could it really be original to this chapter? Most likely, others have had the same thought, but concluded it leads nowhere because nothing that arises uniquely out of punishment is suited to explain the ER. Consider: Punishment is often severe and damaging, yes, but so are many forms of civil liability (deportation, for instance). Punishment often severely restricts liberty, but so does involuntary civil commitment. Punishment requires proof beyond a reasonable doubt, but the reason the ER suppresses evidence has nothing to do with its *reliability*, so the heightened standard of proof seems irrelevant in this context.

Of course, many theorists have claimed that punishment involves blame,[113] where blame is something like an attitude of resentment, indignation, or ill will directed toward a person because she is *blameworthy* (i.e., a culpable wrongdoer, though not necessarily a criminal wrongdoer).[114] But this too seems a nonstarter. Your culpability and the nature of your wrongdoing generally have nothing to do with how I acquired evidence against you, so the ER

---

[112] At least one commentator has attempted to explain the ER using resources from the legal and philosophical scholarship on punishment. *See* Brooks Holland, *The Exclusionary Rule as Punishment*, 36 RUTGERS L. REC. 38 (2009). On Holland's view, retribution requires punishing government officials who culpably violate others' constitutional rights, and the ER *is* that punishment. This account, though focused on punishment, has nothing to do with the fact that *the domain of criminal law is the domain of criminal punishment*; it thus cannot explain why *officers* should only be punished when the evidence they obtain would be used at *criminal* trial.

[113] *See, e.g.*, Duff, *supra* note 6, at 130 & n.16; Feinberg, *supra* note 6, at 400 *passim*; Husak, *supra* note 6, at 449-50; sources cited *infra* notes 140-141.

[114] For the argument that attitudes of resentment and indignation constitute blaming, *see* Peter Strawson, *Freedom and Resentment*, 48 PROCEEDINGS OF THE BRITISH ACADEMY 1 *passim* (1962). It may be that *withdrawal of good will*, or withdrawal of willingness to share a relationship, is another form that blame can take. I take this to be a key point of T.M. SCANLON, MORAL DIMENSIONS: PERMISSIBILITY, MEANING, BLAME, ch. 4 (2008).

generally has nothing to do with what makes you *worthy* of blame. Blame, too, it seems, is a dead end.

But not so fast! It's true that your *blameworthiness* generally has nothing to do with how I acquired evidence against you. My *right to blame you*, however, depends on more than just your blameworthiness. It depends also on my *standing to blame*.[115]

Think of the standing to blame (or "standing" for short[116]) as a condition a would-be *blamer* must meet to have the right to blame her target. Just as you have no right to blame the innocent or exculpated, so you have no right to blame without standing. To say you have no right to blame is not to make a point about the *consequences* of blaming. It may be true that blaming the blameless and blaming without standing are unlikely to have good results. But even when serendipity strikes and the blame has wonderful effects—their life improves; the friendship is repaired; happiness abounds—your blame is still *unjust* in an important respect. You went *beyond your rights*. As a shorthand for these flaws, I will say that your blame is *illegitimate*.[117]

Like the punishment of an innocent, illegitimate blame may not always be morally wrong, all things considered. If the fate of the world depends on it, then you should probably blame illegitimately.[118] But even in the outlandish cases where illegitimate blame is all-things-

---

[115] For arguments to this effect, *see* sources cited *supra* note 7.

[116] Even aside from the legal sense of 'standing' (that is, standing to bring suit), 'standing' might not be a phenomenon restricted to blame. Some have argued that that *praise* is also subject to standing conditions. *See* Daniel Telech, In Praise of Praise, ch. 4 (Aug. 2017) (unpublished Ph.D. dissertation, University of Chicago), https://knowledge.uchicago.edu/record/1611?ln=en. Some may also intuit that one can lack standing to offer criticism that falls short of blame, as in, "who are you to criticize Dostoyevsky's prose? You're a 9th grader!" I do not know whether the kind(s) of 'standing' that that may exist in these other contexts is the same kind of 'standing' at issue in blame, but I put 'standing' in these other contexts aside in what follows.

[117] Philosophers often use the term "inappropriate" to capture the phenomenon, but I think that term misleadingly connotes *etiquette* more than *justice* and so I avoid it here.

[118] That is, if you *can*. Chapters 2 and 3 of this dissertation shed light on why it is so difficult to sincerely blame a person for "forward-looking" reasons like these. For the phenomenon underlying the difficulty, *see* Mark Schroeder, *The Ubiquity of State-Given Reasons*, 122 ETHICS 457 (2012). For parallel arguments focused on attitudes other than blame, *see* Roger Crisp, *Review of Jun Kupperman*, 75

considered right, it is still illegitimate. In this particular respect, blame is like chess. When a demon credibly threatens to destroy the world lest you move your rook diagonally, you'd best move your rook diagonally. Your move will be praiseworthy, but you err if you think that makes your move *legal*. The rules of chess, like the conditions underlying the right to blame, are not so dependent on their consequences in individual and idiosyncratic cases.

When we analogize blame to chess, we risk suggesting the 'rules' of blame matter as much as the rules of a board game. However—and I apologize for making so trite a claim while also *rhyming*—blame is no game. Blame involves ill will or a withholding of good will toward its target.[119] Being the target of blame hurts.[120] The blamer adopts a stance of superiority, demanding the target acknowledge her wrongdoing, feel guilt, and apologize.[121] When the blamer has no right to take up that stance, whether because she lacks standing or her target is not blameworthy, the 'rule' she breaks is a moral rule. And those matter in a way chess cannot.

B. Standing to Blame

When we talk about the conditions for blameworthiness, we start with wrongdoing and culpability. We then debate what makes an act wrong and what makes an agent culpable for it. We might say that culpability requires control and knowledge. We then debate what 'control'

PHILOSOPHY 458 (2000) (focusing on the attitude of desire); Bernard Williams, *Deciding to Believe*, *in* PROBLEMS OF THE SELF 136 (1970) (focusing on belief).

[119] *See supra* note 114 and accompanying text.

[120] *See* Pamela Hieronymi, *The Force and Fairness of Blame*, 18 PHIL. PERSP. 115, 117-19 (2004) (describing the "characteristic force" of blame, including "negative repercussions for the wrongdoer's relations with herself and others" and the "harms, burdens, demands, and expectations contained in or resulting from . . . blaming actions and actions that go beyond the simple belief" that the target is blameworthy).

[121] *See* Zac Cogley, *Basic Desert of Reactive Emotions*, 16 PHIL. EXPLORATIONS 165, 168-71 (2013) (discussing the communicative function of blame and claiming that the "consensus view" is that the "communicative aim" of blame attitudes is the target's "emotional and behavioral recognition of wrongdoing, and – more specifically – the target feeling guilt or remorse and expressing it via apology, amends, and the like" (quoting Coleen Macnamara, *"Screw You!" & "Thank You,"* 165 PHIL. STUD. 893, 895 (2013))).

amounts to—e.g., free will of the sort incompatible with causal determinism?—and what 'knowledge' amounts to—e.g., knowledge that one's act is morally wrong? Because these issues are immensely complex, we've debated them for millennia.[122]

Standing should be understood against this background. The concept of standing is a toddler beside the concept of blameworthiness, and the literature on standing is dwarfed by the literature on blameworthiness.[123] If standing is remotely like blameworthiness, a definitive and complete analysis will not come soon, and it would be foolish to attempt one here. But we do not need completeness to understand why the standing to blame provides the best explanation for the exclusionary rule. For that, the following discussion will do.

Unlike blameworthiness, standing is more easily understood as the 'default' position for persons. Blameworthiness requires culpability and wrongdoing. And putting aside Eve's apple-eating misadventure, you and I aren't culpable wrongdoers by default. Standing, in contrast, is

---

[122] For an overview of the history and key issues within these debates, *see* Dana K. Nelkin, *Moral Luck*, *in* STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2018); Timothy O'Connor & Christopher Franklin, *Free Will*, *in* STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2018); Fernando Rudy-Hiller, *The Epistemic Condition for Moral Responsibility*, *in* STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2018); Matthew Talbert, *Moral Responsibility*, *in* STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2019); Neal Tognazzini & D. Justin Coates, *Blame*, *in* STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2018); Kadri Vihvelin, *Arguments for Incompatibilism*, *in* STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2017).

[123] I am unsure when the concept of the standing to blame entered the philosophical lexicon. I have heard some philosophers attribute the idea to Cohen, *supra* note 7 (2006). But it was present in Duff's work before that time. *See* Matt Matravers, *'Who's Still Standing?' A comment on Antony Duff's Preconditions of Criminal Liability*, 3 J. MORAL PHIL. 320, 321 n.2, 322 *passim* (2006) (suggesting the concept was present in some form in Duff's 1986 TRIALS AND PUNISHMENTS but that it evolved significantly between that work and Duff's 2001 PUNISHMENT, COMMUNICATION, AND COMMUNITY). Additionally, Justin Coates, *supra* note 8, at 457, suggests the concept is also present in Gary Watson, *Responsibility and the Limits of Evil: Variations on a Strawsonian Theme*, *in* PERSPECTIVES ON MORAL RESPONSIBILITY 119 (2004) (originally published 1987). Watson never uses the term 'standing' (or any other term for the phenomenon) in that work, but I agree the concept is present there. *See id.* at 139 ("I too am a potential sinner . . . . [This] thought about moral luck turns one's gaze inward. It makes one feel less in a position to cast blame. . . . [But this] does not, of course, mean that [the target of blame] is not actually a vile man.").

something normal human adults have *unless* one of several conditions obtain. The most widely accepted of these are *hypocrisy*, *complicity*, and *meddling*.

Hypocrisy is the most often discussed. Very roughly,

1. Agent A lacks standing to blame Agent B for B's culpable wrongdoing if A is a *hypocrite* with respect to B's culpable wrongdoing.

2. A is a hypocrite with respect to B's culpable wrongdoing if

    a. A is blameworthy for the same type of wrongdoing, and
    b. A is at least as blameworthy for her wrongdoing as B is for his.[124]

To illustrate, imagine that you are traveling down a congested sidewalk lined with "no biking" signs. A person riding their bike is coming your way and getting precariously close to some traveling schoolchildren. You find yourself blaming this biker: you feel a bit indignant and shoot him a disapproving look. "Jerk," you think, "the signs are everywhere." So far, this seems like a perfectly reasonable case of blame. The biker is committing a minor culpable wrong. (At least, you have no reason to think he's *not* culpable.) And your indignation and disapproving look are consistent with the extent of his blameworthiness.

But we left out a detail. Your head was so far in the clouds that you'd simply forgotten: *You are riding your own bike on this very same sidewalk right now.* Indeed, you often ride your bike down this sidewalk, and your only justification for doing so is that it's more convenient to bike than to walk. Had you been a pedestrian, you'd have had every right to blame the blameworthy biker. But as you look at him disapprovingly from atop the high seat of your bike, *your* blame is illegitimate. You lack standing because you are a hypocrite, in the sense defined above. Notice that the biker could ask, "Who are *you* to blame me?," and his question would

[124] Why the two-part structure? Because it allows us to define "hypocrite": this emphasizes that hypocrisy is a single, unified way to lose standing, and it makes "hypocrite" a semi-technical term, preserving the possibility some cases of 'hypocrisy' in the layperson's sense may not meet the definition given here.

have force. The force and relevance of this particular question is a clue that *standing* is the issue.[125] The point is not that your blame has bad *consequences* or that *he* does not deserve it. The problem is *you*.

Note, too, that when you realize your hypocrisy, and you have this realization clearly in mind, you'll have a hard time keeping up your sincere indignation. After all, who *are* you to blame him for something you yourself are doing? Realizing you have no answer is a bit like realizing that the target of your blame is innocent: such realizations, when clear and sincere, cause blame to *evaporate*, much like your desire for a chocolate cake evaporates when you realize it's made of dirt.[126] If you want to keep blaming the biker, you need to convince yourself you're not a hypocrite after all.

Finally, note that your standing would not be undermined if the other biker was on a *motor*bike. For then, even if you're committing the same 'type' of wrongdoing (biking illegally on the sidewalk), his version is much dangerous and much more blameworthy than yours.[127] Your standing would also not be undermined if, rather than riding a bike on the sidewalk, you were sending an unnecessarily rude e-mail to your secretary that will hurt his feelings. Even if your email is more blameworthy than your target's bike riding, these are not the same type of wrongdoing. If we lacked standing to blame others every time we were culpable for something—*anything*—equally or more blameworthy, practically none of us would have standing to blame one another for the sorts of mild to moderate culpable wrongdoings that pervade our lives and our relationships.[128] (If you blamed me for lying to you, for example, I could successfully retort

[125] Todd, *supra* note 7, at 349 ("[Whether] the 'who are *you* to blame me?' reply is appropriate is . . . what is distinctively at issue concerning the 'moral standing' to blame.").
[126] *See* chapter 2 for a more detailed discussion of this point.
[127] I am neglecting the question, what is it for one wrongdoing to be of the same *type* as another?
[128] This may seem too quick. It's possible to redeem ourselves for blameworthy actions we've taken through apology, recompense, and changing our ways. And the mere passage of time might reduce our

that you've wasted *far* too much water in your lifetime to blame anyone for anything less serious than theft of property worth at least $200.)

Consider next complicity:

1. Agent A lacks standing to blame Agent B for B's culpable wrongdoing X if A is *complicit* in X.

2. A is complicit in B's culpable wrongdoing X if

    a. A is herself blameworthy for X, and
    b. A is at least as blameworthy for X as B is.

To illustrate, say that you and your friend are high-school students and you both know that your teacher keeps a large amount of candy stored behind her desk. Outside the classroom after school, your friend says he is hungry, and you remind him of the stored candy. "It's not that big a deal," you say, "just go for it! She's already gone for the day." Only because of your encouragement, your friend steals the candy. The next day, you're called to the principal's office: the teacher noticed the missing candy and knows of your involvement. If you now blame your friend for stealing the candy, you go beyond your rights.[129] Your friend *is* blameworthy, indeed, but *you* can't blame him: you were involved in it yourself, and the role you played makes you as blameworthy as he is.

Consider finally meddling.[130] Meddling is hard to formulate in a way that's precise and informative. We might say, A lacks standing to blame B for B's blameworthy wrongdoing if B's

---

blameworthiness without such acts on our part. (Am I really still blameworthy for carelessly hurting a short-term boyfriend's feelings 10 years ago? 40 years ago?) Even still, most of us do blameworthy things often in our lives and do not redeem ourselves. This is not particularly unusual, and it does not undermine our blame indefinitely into the future. *See* chapter 4; *see also* Section V.A.3 (noting that future work must engage with the question whether and how the state can *recover* its standing to blame a defendant).

[129] You might blame him *for getting caught*, but he's probably not blameworthy for that. He's blameworthy for stealing from the teacher.

[130] This condition has gone under many names, including the "business" condition and the "community" condition. I choose meddling because it has the same moral valence as hypocrisy and complicity: "business" and "community" are what one *lacks* when one meddles.

blameworthy wrongdoing is *none of A's business*. But while that captures meddling's intuitive thrust, it does not answer the question, *when is B's wrongdoing none of A's business?* We might answer, it's when A and B are not "fellow members of an appropriate moral community."[131] But even if that were precisely right, it gives us no way to define the meddling condition with precision, or even clarify it, unless we have precise and clear accounts of the thick, contested notions of "appropriate moral community" and "fellow[ship]" within it.

Still, meddling is not so hard to understand. Say you've recently been failing to treat a dear friend as well as you should—failing to reciprocate her acts of friendship, failing to reach out to her during a difficult time, etc. It may well be that you are culpable for this wrong, and you might well grant that this is true. But now imagine learning that *I* blame you for this. I learned about the matter through gossip; I don't know you, and I don't know your friend: I am just very nosey, and I am now indignant toward you for wronging your friend. You might reply to me, "who are *you* to blame *me* for how I've been as a friend to *her*? You don't know me; you don't know her; you don't know *us*, or our friendship: This is *none of your business*!"[132]

Structurally, meddling is just like hypocrisy and complicity. Hypocritical and complicit blame each involve two key features. First, the blamer is also blameworthy for an equally or more blameworthy wrong. Second, the blamer's blame is *linked* to her comparably blameworthy wrong: In hypocrisy, the blamer's wrongdoing is the *same type* as her target's. In complicity, the

[131] Duff, *supra* note 6, at 126.

[132] My nosy character seems the sort to riposte: "Who am I? I am also someone who has friends, and I'm a person, like you, who has to manage the expectations of friendship along with all the obligations of life. It is not *only* your business how you treat your friends, and I'm allowed to blame you for treating yours badly." But is this reply convincing? I don't think so. It seems to deny that there exist any private obligations and private wrongdoings that are not the business of the broader community. In this, it seems very "middle school." If you think otherwise, you might be inclined to deny that a meddling condition exists at all. But that might be too fast: You might just think that all genuinely culpable wrongdoings are your business!

blamer's wrongdoing is the *same token* as her target's. In meddling, the blamer's wrongdoing is her violation her target's right to privacy. What links the blamer's *blame* to her blameworthy privacy violation? Easy: the blame is part of the privacy violation—the blamer *wrongs* her target, in part, *by blaming her.*

A fourth way to lose standing, discussed next, shares in this common structure.

C. Epistemic Illicitness

I have claimed that an agent can lose the standing to blame when the agent has culpably violated her target's rights, the agent knows her target is a culpable wrongdoer only because of that violation, and the agent is at least as blameworthy for her own violation as her target is for her wrongdoing. Call this *epistemic illicitness*. Formally,

1. A lacks standing to blame B for B's culpable wrongdoing X if A's knowledge that B committed X is *epistemically illicit*.
2. A's knowledge that B committed X is epistemically illicit if
    a. A culpably violated B,
    b. That culpable violation was the factual and proximate cause of A's possessing evidence on the basis of which A knows B committed X,
    c. Without the aforementioned evidence, A would not know that B committed X, and
    d. A's culpable wrongdoing is at least as blameworthy as X.

As we noted above, epistemic illicitness is structurally similar to hypocrisy, complicity, and meddling. In all four cases, the blamer loses standing because (i) she is blameworthy for a wrongdoing of her own, and (ii) something links her own blameworthy wrongdoing to the blame she would direct at another. The general nature of the requisite link is a deep question in the philosophy of standing that I cannot fully explore here. But in the context of epistemic illicitness, the link is easy enough to identify.

The first step is to recognize that my right to blame you depends on my *epistemically justified belief* that you are blameworthy.[133] My belief is epistemically justified only if I have evidence to support it.[134] This evidence might take many forms, and it needn't be admissible in a court of law to be sufficient for epistemically justified belief. But without such justification, I have no right to blame you. How much justification is required? I assume, following Justin Coates, *enough for knowledge* (Gettier cases aside).[135] My right to blame you therefore depends on my *knowledge* that you are blameworthy.[136] (I integrate this assumption into my account and

---

[133] *See, e.g.*, Dana Kay Nelkin, *Rational Belief and Statistical Evidence: Blame, Bias, and the Law*, *in* LOTTERIES, KNOWLEDGE, AND RATIONAL BELIEF: ESSAYS ON THE LOTTERY PARADOX 5, 25 (Igor Douven ed., 2021) ("[B]lame, to be justified, depend[s] on the rational belief that the agent in question is culpable."); Coates, *supra* note 8, at 457 *passim* ("[I]t is inappropriate for us to blame others if it is not reasonable for us to believe that they are morally responsible."); *id.* at 458 n.3 ("In saying that *A* blame is inappropriate if it is not reasonable for her to believe that *B* is morally responsible, I mean to pick out a notion of 'reasonability' that is closely allied with epistemic justification."); Todd, *supra* note 8, at 350-51 ("[I]n order appropriately to blame a given wrongdoer, one must be justified in believing that the given agent is indeed blameworthy."). Gideon Rosen relies on this proposition when he claims that "we are never entitled to any significant confidence" that an agent is blameworthy for a given act, so we should "presumably stop blaming" our friends and neighbors and "presumably also take steps to revise or reconceive any social practice whose justification depends on the assumption that people are morally responsible for their bad actions." *Skepticism About Moral Responsibility*, 18 PHIL. PERSP. 295, 309-11 (2004). Lara Buchak does the same when she gives the following case as proof that statistical evidence alone is not sufficient for epistemically rational belief:

> You leave the seminar room to get a drink, and you come back to find that your iPhone has been stolen. There were only two people in the room, Jake and Barbara. You have no evidence about who stole the phone, and you don't know either party very well, but you know (let's say) that men are 10 times more likely to steal iPhones than women. I contend that this isn't enough to make you *rationally believe* that Jake stole the phone. If you accused Jake, he could, it seems to me, rightly point out that you don't have evidence that he in particular stole the phone.

Belief, Credence, and Norms, 169 PHIL. STUD. 285, 292 (2014) (emphasis added). Because statistical evidence is not sufficient for *rational belief*, Buchak tells us, Jake can rebut your blame by pointing out that you only have statistical evidence. Underlying this argument is the proposition that legitimate blame requires evidence sufficient for rational belief.

[134] I don't claim this goes for all beliefs, but it goes for my belief that *you are blameworthy*—an empirical, *a posteriori* proposition.

[135] Coates, *supra* note 8, at 465 ("[W]e should refrain from blaming . . . if we do not *know* that the presumptive wrongdoer is, in fact, morally responsible for her action." (emphasis added)).

[136] I endorse this proposition in part because I think a related proposition, endorsed by Sarah Moss, is correct, namely, that legal proof sufficient for criminal conviction requires knowledge. *See* Moss,

in the discussion to follow, but this is largely for ease of exegesis; the account's success does not turn on speaking of knowledge rather than justified belief.[137, 138])

Further, as Section IV.C further elaborates, when I culpably violate your rights, and my culpable violation is the proximate and factual cause my acquisition of evidence against you, my acquisition of this evidence inherits the blameworthiness or "taint" of my culpable wrongdoing. It follows that my right to blame you depends on my knowledge that you are blameworthy, and my knowledge that you are blameworthy depends on evidence tainted by my culpable wrongdoing.[139] This is what links my blame to my blameworthiness: my *right to blame* you is grounded in evidence *tainted* by my *blameworthiness* for wronging you.

To see how the epistemic illicitness condition appears in the interpersonal context, imagine that you, a nosy parent, sneak into your 16-year-old daughter's room while she's away. You open her desk drawer, find her diary, and read it. You have no good reason for doing so. Your daughter cares very much about her privacy, and she'd be appalled if she knew what you

---

Knowledge and Legal Proof, forthcoming in 7 Oxford Studies in Epistemology (2023), http://www-personal.umich.edu/~ssmoss/Moss%20-%20Knowledge%20and%20Legal%20Proof.pdf.

[137] If we insisted that the justification required need *not* be sufficient for knowledge, we would need to modify the formulation of epistemic illicitness by replacing "knowledge" with "belief," and replacing "know" with "justifiably believe" in condition (2.c). This would be cumbersome, but it would not cause particular problems. We would need to make similar alterations elsewhere when discussing the 'link,' in epistemic illicitness, between the blamer's blame and her blameworthiness for wronging her target.

[138] Note, I do not think it matters for formulating epistemic illicitness that *knowledge* entails that the justified belief is also *true*. If one wished to argue that a blamer lacks standing only when the conditions for legitimate blame are otherwise in place, then it is essential to epistemic illicitness that the justified belief be true; but I do not endorse this argument myself. If one does, it would be important to speak of knowledge, or if one thinks the justification required falls short of that required for knowledge, of justified *true* belief.

[139] Though not essential for the account's success, the epistemic illicitness condition comports most easily with David Brink and Dana Nelkin' view that "[t]he core of blame . . . involves a belief that the target of blame is blameworthy." David O. Brink & Dana Kay Nelkin, manuscript, *The Nature and Significance of Blame*, at 13 (forthcoming in The Oxford Handbook of Moral Psychology (J. Doris & M. Vargas, eds.)), https://davidobrink.com/sites/default/files/work-in-progress/BrinkNelkinTheSignificanceOfBlamePost.pdf. On their view, it is easy to see why blame would be unreasonable if the belief were unreasonable—and why blame would be tainted if the evidence upon which the belief rests is tainted, too.

were doing. To your surprise, you learn that your daughter has been breaking curfew to make out with her boyfriend after his parents go to sleep. Angry that she would break your rules like that, you confront her angrily when she comes home. "You've been breaking the rules of this house," you say, indignant, "and breaking your curfew to make out with your boyfriend. That's unacceptable!"

She's briefly confused (*how did you know?*) but soon puts two and two together. "Did you read *my diary!?*" she asks, outraged. After a pause, she continues, "Are you kidding right now? You violate my privacy in the most serious way possible, find out I broke *curfew*, and now you want to stand there on your high horse and give me a lecture for something that's not half as bad as what you did? Something *you only know about because you read my diary*? Yeah, no, that's insane. How *dare* you read my diary, and how *dare* you stand here and use what you found to act all high and mighty?"

Notice, your daughter's insightful rebuke does not involve a denial of her own wrongdoing. Instead, she diminishes the gravity of her wrongdoing relative to yours. And notice you cannot reasonably say, "Yes, I was wrong to read your diary, but that's a different issue; right now, we're talking about what you did." For your daughter can rightly insist, "It is not a different issue! It is the only reason you're standing here right now blaming me!" This is not a blanket challenge to your moral authority; it is a challenge to your attempt to separate your *blaming* her from your *wronging* her. They are not two separate issues because they are not separate: Your right to blame depends on evidence tainted by your wrongdoing.

The formulation of epistemic illicitness provided in this Section includes components—e.g., factual and proximate causation in (2.b)—that the chapter has not yet explained. These threads are picked up in Part IV, where explaining those components serves also to explain key

features of the ER. At this point, I nevertheless hope to have said enough to motivate the existence of the epistemic illicitness condition, both by exposing its basic structure and illustrating how it comes up in our interpersonal lives. Even if we accept that epistemic illicitness undermines the standing to *blame*, however, this does not mean it undermines the standing to *punish*. We turn to this point next.

D. Standing to *Punish* Requires Standing to *Blame*

Many theorists have argued that the state cannot legitimately punish a defendant if it lacks the standing to blame them. Most of the arguments in this literature operate at the intersection of hypocrisy and complicity. Antony Duff, Victor Tadros, and Gary Watson have all argued that in some actual cases, the state's standing to punish criminals is undermined by its hypocrisy or complicity in causing the criminal behavior.[140] Many others have followed these

[140] *See, e.g.*, Duff, *supra* note 6 *passim*; Tadros, *supra* note 7 *passim* (2009); Watson, *supra* note 7, at 172 *passim*. Though I am unsure whether he has made an argument of the form discussed here, Douglas Husak is also associated with the proposition that legitimate punishment depends on legitimate blame. *See* Husak, *supra* note 6, at 449-50 ("The retributivist dream, as I will construe it, is to create a world in which impositions of criminal liability and punishment correspond to our considered judgments of blame and desert."). But I hasten to add that nothing in this chapter assumes that retributivism is true; that includes the claim that legitimate punishment depends on legitimate blame. Tadros, for one, is known in part for his trenchant arguments *against* retributivism, which he claims is "barbaric." *See* VICTOR TADROS, THE ENDS OF HARM: THE MORAL FOUNDATIONS OF CRIMINAL LAW, ch. 4 (2011). Duff, moreover, is perhaps the most well-known contemporary expressivist about punishment. *See* DUFF, PUNISHMENT, COMMUNICATION, AND COMMUNITY (2001); Joshua Glasgow, *The Expressivist Theory of Punishment Defended*, 34 LAW & PHIL. 601, 603 & n.3 *passim* (2015) (describing Duff as a "communicative" expressivist and distinguishing expressivism from other views, including retributivism). Expressivism is perhaps the view with which the claim that legitimate punishment depends on legitimate blame has been *most* closely associated, ever since Joel Feinberg famously argued that punishment is *defined* by its function as an "expression of attitudes of resentment and indignation, and of judgments of disapproval and reprobation." *See* Feinberg, *supra* note 6, at 400 *passim* (1965).

theorists.[141] And even accounts that criticize their arguments have often accepted the core premise that hypocrisy and complicity can undermine legitimate punishment.[142]

The form of argument these theorists have made is not restricted to hypocrisy and complicity, however. Gabriel Mendlow argues that meddling (or "lack of business") undermines the state's standing to punish defendants for certain crimes.[143] These are crimes that primarily victimize an individual person (e.g., sexual assault) rather than the state or the community at large (e.g., drug dealing), and for which the state refuses to allow the primary victim to be involved in the prosecution. In crimes like sexual assault, it is foremost the *victim*'s business to blame the wrongdoer: when the state unilaterally prosecutes her wrongdoer, it acts as if *her* sexual assault is foremost *its* business. The state may have the political authority necessary to punish in general, but when it usurps an individual victim's role as the one to whom accountability is foremost owed, its standing to blame and punish is undermined.[144]

The literature from Duff, Tadros, Watson, Mendlow, and others shows that this chapter fits soundly into a widely endorsed theoretical program predicated on the premise that the legitimacy of the state's punishment depends on its standing to blame. True, the chapter is the first to raise the epistemic illicitness condition (aside from its discussion later in this dissertation); it is the first to argue that the standing to blame explains the exclusionary rule; and it is certainly the first to develop a comprehensive theory of the rule using the epistemic

---

[141] See, e.g., Gustavo A. Beade, *Who Can Blame Whom? Moral Standing to Blame and Punish Deprived Citizens*, 13 CRIM. L. & PHIL. 271 (2019); Kyle G. Fritz, *Hypocrisy, Inconsistency, and the Moral Standing of the State*, 13 CRIM. L. & PHIL. 309 (2019); Matravers, *supra* note 123.

[142] *See, e.g.*, Peter Chau, *Duff on the Legitimacy of Punishment of Socially Deprived Offenders*, 6 CRIM. L. & PHIL. 247 (2012) (arguing that the state can be understood to act solely on behalf of "the just citizens" in criminal proceedings, who are by definition those who have committed no culpable hypocrisy or complicity that could undermine their standing). *But see also* TOMMIE SHELBY, DARK GHETTOS: INJUSTICE, DISSENT, AND REFORM 240-43 (2016) (replying to Watson, *supra* note 7).

[143] Mendlow, *supra* note 8, at 1153-56.

[144] *See id*. at 1166-71.

illicitness condition. But it is *far* from the first to suggest that the legitimacy of punishment depends on the legitimacy of blame; that the state lacks standing to punish when it lacks standing to blame; or that the standing to blame can have radical implications for the criminal law.[145]

I make these points because I do not have the space in this chapter to prove that legitimate punishment presupposes legitimate blame. If I were to try, I would have to either oversimplify the complex issues that underly the claim or give up any right to call this a "chapter."[146] So I opt to take the claim as an assumption here, and I devote this chapter to its novel contributions. Because the claim is widely endorsed, the chapter remains valuable even for

[145] Moreover, since hypocrisy, complicitly, and lack of business have all been leveraged to argue that the state cannot legitimately punish guilty criminals in wide classes of cases, it would be surprising if the epistemic illicitness condition was *uniquely* incapable of supporting a similar kind of argument. And while hypocrisy, complicity, and lack of business have primarily been used to undermine state punishment, not to explain major features of it, *blame* has long been used to explain major features of criminal punishment, so this chapter is not so unusual in this regard, either. *See also* Duff, *supra* note 6, 133-36 (investigating whether complicity might explain why entrapment undermines the legitimacy of a criminal conviction but concluding that the explanation is most likely unsuccessful).

[146] To illustrate the magnitude of the difficulty, consider that some theorists reject the position that legitimate punishment presupposes legitimate blame on the basis of idiosyncratic views about the nature of blame, punishment, and criminal law. For example, David Shoemaker endorses a version of Tim Scanlon's view of blame and blameworthiness, which Shoemaker interprets to claim that blameworthiness tracks only one's *reasons* for action. *See* David Shoemaker, *Blame and Punishment*, *in* BLAME: ITS NATURE AND NORMS 100, 101-03 (D. Justin Coates & Neal A Tognazzini, 2013); SCANLON, *supra* note 114, ch. 4. He juxtaposes this view with a conception of legitimate criminal liability that tracks only the *impermissibility* of one's action, where impermissibility is independent of one's reasons for action. Shoemaker, *supra*, at 103-05. Not even Tim Scanlon holds this *combination* of views, for Scanlon adopts the more orthodox position that distinctions between crimes are matters of the agent's *reasons* for action—this is what mens rea *is*, for Scanlon, which for him suggests "that criminal penalties are . . . expressions of blame [in] the sense I describe." T.M. Scanlon, Blog Comment, May 19, 2011, 12:51 PM, on David Shoemaker, *Scanlon on Blame, Part 3: Criminal Blame and Meaning*, PEA SOUP, May 19, 2011, https://peasoup.typepad.com/peasoup/2011/05/scanlon-on-blame-part-3-criminal-blame-and-meaning.html. Nevertheless, Shoemaker's combination of views is not obviously wrong—*see* Shoemaker, *supra*, at 105-113 (responding to the objection that criminal liability is a function of one's reasons for action)—and together, the views imply that the legitimacy of punishment does not depend on the legitimacy of blame: their legitimacies derive from different sources, neither of which is a part or presupposition of the other. To prove that legitimate punishment depends on legitimate blame, I would have to prove Shoemaker wrong. But I cannot do that *quickly* unless I make Shoemaker's position sound daft, which it is not. For an appropriately 20-page discussion of Shoemaker's position, *see* David Lefkowitz, *Blame and the Criminal Law*, 6 JURISPRUDENCE 451 (2015).

those who conclude it proves only a *conditional*, one whose antecedent reads, *If legitimate state punishment requires the standing to blame . . . .*

But I do want to block a potential misunderstanding. The claim is about actuality, not necessity. There are some who believe that blame both *can* and *should* be stripped from state punishment.[147] I reject both propositions, but they are compatible with my account of the ER because they are compatible with the claim that the legitimacy of state punishment as it *actually* is depends on the legitimacy of blame.[148] The possibility, feasibility, and wisdom of *changing* this fact by purging blame from our system of criminal punishment is only relevant to a separate question about the ER: whether we should *maintain* it. It might be that the ER is so costly that we should remove blame from punishment to get rid of the ER. Or it might be that, ER aside, we

---

[147] As I read them, Tommie Shelby and Erin Kelly endorse these propositions. *See* SHELBY, *supra* note 142, at 240-41 (2016); Erin Kelly, *The Ethics of Law's Authority: On Tommie Shelby's,* Dark Ghettos: Injustice, Dissent, and Reform, published online July 1, 2019, https://doi.org/10.1007/s11572-019-09498-5.

[148] "Compatible" is an understatement: if we *can* and *should* strip blame from punishment, then presumably blame must *be* a part of punishment now. Consider, with this in mind, Kelly's suggestion that we should "reject the blaming function of punishment" because

> [t]he state doesn't have a moral leg, or even a toenail, to stand on from which to proclaim the blameworthiness of people for committing illegal actions that turn out to be last-resort measures to rescue a basic sense of self-worth from the damage inflicted by the state's unjust policies and practices.

Kelly, *supra* note 147, at 1, 10. This is another version of the Duff-, Tadros-, and Watson-style argument: state punishment is often illegitimate because of the state's hypocrisy or complicity in impermissibly creating the criminogenic conditions that caused the crimes it now condemns. It follows that we would be mistaken to read "reject the blaming function of punishment" to mean "deny the *proposition* that punishment functions to blame": Kelly's view instead *takes the proposition as a premise*, adds that the state's hypocrisy and complicity *systematically* and *severely* undermine its standing to blame, and concludes we should remove blame from state punishment to restore and preserve the latter's legitimacy. *See also* Andy Engen, *Punishing the Oppressed and the Standing to Blame*, 97 RES PHILOSOPHICA 271-72, 278-81 (2020) (replying to Shelby's claim that "we needn't attach any *symbolic* significance to the sentence itself" by noting that Shelby's claim is made "not in terms of what we actually do, but in terms of what's possible" (quoting SHELBY, *supra* note 142, at 241)).

already have sufficient reason to strip blame from punishment. The ultimate justifiability of the ER turns on these possibilities. But this chapter's explanatory intervention does not.

E. Statement of the Standing to Blame + ("STB+") Account

We are ready, in this Section, to formally state this chapter's account of the ER.

Because the legitimacy of state punishment depends on the state's standing to blame, and epistemic illicitness undermines the standing to blame, the state cannot legitimately punish a defendant when it meets the conditions of epistemic illicitness with respect to that defendant. And here's the key connection: The ER functions to exclude evidence that, if essential to a guilty verdict, would satisfy the epistemic illicitness condition and therefore undermine the state's ability to legitimately punish a defendant on the verdict's basis. If the ER were maximally consistent with its normative foundation, it would exclude all and only evidence so described. The actual ER does not meet this ideal, but it does a better job than one might expect, and when we understand ER doctrine as the law's attempt to realize that ideal, the ER becomes explicable, and its "chaotic" doctrine acquires an order that has heretofore been elusive.

Call the set of claims made in the preceding paragraph the **Standing to Blame Theory** of the ER. The theory is compatible with a formulation of epistemic illicitness that deviates from my precise formulation. But in what follows, I will assume the formulation of epistemic illicitness stated in Section III.C and repeated here:

Epistemic Illicitness: *General Form*

1. Agent A lacks standing to blame Agent B for B's culpable wrongdoing X if A's knowledge that B committed X is *epistemically illicit*.
2. A's knowledge that B committed X is epistemically illicit if
    a. A culpably violated B,

b. That culpable violation was the factual and proximate cause of A's possessing evidence on the basis of which A knows B committed X,

c. Without the aforementioned evidence, A would not know that B committed X, and

d. A's culpable wrongdoing is at least as blameworthy as X.

For the most part, it's obvious how this applies to the state in context of criminal investigations and prosecutions, but for one twist. In the context of state punishment, when we talk about the state's "knowledge that B committed X," we aren't talking about the *state's* justified true belief that B is guilty (whatever that would amount to). Rather, we're talking about the *guilty verdict* that the state obtained against the defendant. Here, I assume that juries (or judges as factfinders in bench trials) determine whether the state's evidence is sufficient for knowledge of guilt. A guilty verdict on the basis of a set of admitted evidence plays the role of *belief* that the defendant is guilty, *justified* by the admitted evidence. If the belief is not true or the evidence epistemically insufficient, the guilty verdict is insufficient for knowledge. If the belief is true and the evidence sufficient, the state has the sort of knowledge required to subject a defendant legitimately to criminal blame and punishment, *unless* the knowledge is epistemically illicit.

Integrating these assumptions into the formulation of Epistemic Illicitness gives us the following:

Epistemic Illicitness: *Criminal Law Context*

1. The state lacks standing to blame defendant D for D's culpable wrongdoing X on the basis of an accurate and epistemically justified guilty verdict if that verdict is *epistemically illicit*.

2. An accurate and epistemically justified guilty verdict is epistemically illicit if

    a. The state culpably violated D,

b. That culpable violation was the factual and proximate cause of the state's possessing and submitting evidence on which the guilty verdict is based,

c. Without the aforementioned evidence, there would be no guilty verdict or the guilty verdict would rest on insufficient evidence, and

d. The state's culpable wrongdoing is at least as blameworthy as X.

In Part IV, when I refer back to, say, "condition 2.b of Epistemic Illicitness," context generally makes it obvious which "condition 2.b" I mean. If I am ever unclear, it because I am treating the two versions as effectively interchangeable in the ER context, in light of the assumptions stated in this Section.

To conclude and recapitulate, I've made three sets of claims in this Section. The first set is the Standing to Blame Theory, which states a theory of the exclusionary rule but does not take a stand on the precise formulation of epistemic illicitness. The second is a precise, *general* formulation of epistemic illicitness that does not take a stand on the connection between the state's knowledge of a defendant's guilt and the role of judges or juries as factfinders in the criminal context. And the third is a set of assumptions about that connection and the precise, *specific* formulation of epistemic illicitness they entail in the criminal context. The Standing to Blame Theory may be correct even if I've erred in articulating the second and third set of claims. I therefore call the conjunction of all three sets of claims the **Standing to Blame + account**, or just the **STB+** for convenience. The next Part aims to show that the STB+ provides the best available explanation for the ER.

IV. GIVING ORDER TO EXCLUSIONARY RULE DOCTRINE

This Part provides the chapter's "top down" argument for the Standing to Blame + account of the ER. Specifically, it illustrates that the account can explain the seemingly chaotic features of ER doctrine.

Section IV.A applies the STB+ to illustrate why the ER is restricted to a particular subset of the criminal domain, and Section IV.C explains the rule's causation requirements and its bar on vicarious assertions. In these Sections, the STB+ explains why the doctrine is correct *at least given* assumptions made elsewhere in the law (e.g., compensatory damages are not punishment). Note that if one of these assumptions is wrong, the law has erred, not the account: the STB+ merely shows how a false assumption that might have seemed unrelated to the ER in fact infects its doctrine. The account's explanatory success turns only on the fact that the law actually makes the assumption at issue.

In Section IV.B, the STB+ does prescribe change in ER doctrine. Specifically, it suggests that the good faith exception and the ER's application to all and only constitutional contexts are too simple. Where the law draws hard lines—between gross and mere negligence, between constitutional and non-constitutional violations—the ER requires gradations based on degrees of blameworthiness. The fact the STB+ diagnoses doctrinal error here counts against it as an explanation of the ER, but only a little. As Section I.B stressed, we should not expect perfect consistency between the ER's doctrine and its foundations. And we should be particularly open to the suggestion that ER doctrine is mistaken *here*, where America is alone among polities with an ER,[149] for if America were in the right, the STB+ would be inconsistent with every other country's exclusionary rule.

A. The Criminal Domain and Vicarious Assertions[150]

[149] *See* sources cited *infra* note 182 and accompanying text.
[150] *See supra* Sections I.A.1-3 for a review of the doctrines presently explained.

1. Punitive Damages

Roughly, the STB+ says the ER applies to the criminal domain because it is the domain of punishment and blame. But this oversimplifies, for punishment and blame appear outside the criminal domain in the context of punitive damages.[151] Punitive damages are awarded in *civil* law, yet as their name suggests, they are *punishment*. Why, then, does the ER not apply?

It's not because punitive damages are an unjustifiable mistake. Even if they are a mistake, our law *takes* them to be legitimate, so this cannot explain why our law does not apply the ER. Nor is it because introducing the ER in the punitive damages context would raise administrative difficulties. True, juries would have to consider some evidence for one purpose (compensatory damages) and then ignore it for another (punitive damages). But the ER poses similar problems of administration in the criminal context, and our law assumes juries can compartmentalize.[152]

Might it be that punitive damages *aren't*, in fact, punishment? Punitive damages are not subject to many of the protections we take a system of legal punishment to require, after all. Think of the difference between *by a preponderance of evidence* and *beyond a reasonable doubt*. Moreover, as Judge Posner argued in *Kemezy v. Peters*, punitive damages have a compensatory purpose. They fill the gap left by compensatory damages' systemic failure to fully compensate victims.[153]

But this reply fails too. As one scholar notes, "The very label 'punitive' manifests the penal nature of punitive damages," and the "vast majority of courts, including the Supreme

---

[151] Some readers might hold the view that blame underlies the whole domain of tort law. I set this aside because it is a minority view, and it is not an objection to the STB+. If the view is correct, then the ER's restriction to the criminal domain is easily explained, even if mistaken, by the fact that the view is largely rejected in the law.

[152] *See* Zafiro v. United States, 506 U.S. 534, 540 (1993) ("[J]uries are presumed to follow their instructions."); United States v. Hill, 643 F.3d 807, 829 (2011) (11th Cir. 2011) ("[W]e apply 'the strong presumption . . . that jurors are able to compartmentalize evidence by respecting limiting instructions.").

[153] 79 F.3d 33, 34-35 (7th Cir. 1996).

Court, recognize that punitive damages serve the function[] of retribution."[154] At least in the law's eyes, punitive damages are punishment. And punishment involves blame. Blame is subject to the standing to blame, so punitive damages are too. The STB+ cannot deny these facts to explain the ER's non-application in this context.

Nor need it do so. The reason the ER does not apply to punitive damages is found in something else Posner said in *Kemezy*: "Punitive damages relieve the pressures on the criminal justice system . . . by giving *private individuals*—the tort victims themselves—a monetary incentive to *shoulder the costs of enforcement*."[155] Indeed, in the punitive damages context, the "prosecutor" is the *plaintiff*—the victim. She is the one who argues that the defendant is a culpable wrongdoer, deserving of punishment in the form of punitive damages. If she is a private individual, then by definition she did not violate the defendant's constitutional rights to acquire evidence against him. That was the state. The plaintiff uses the fruits of the state's violation, true. But she does not risk epistemic illicitness, for she flunks its most basic requirement: She did not wrong the defendant. Indeed, as far as we are aware, *she is not blameworthy for anything*. This explains the ER's non-application in the vast majority of punitive damages contexts.

But say the plaintiff is the *state*. It seeks punitive damages on the basis of evidence that the ER would exclude in the criminal context. In such circumstances, the STB+ must say that the ER should apply. But note: it is not clear that it *doesn't* apply. I can find no record of any case in which a state or state agency *ever* sought punitive damages on the basis of evidence acquired unconstitutionally. Cases in which a state or state agency seeks punitive damages seem to be exceptionally rare to begin. The STB+ cannot be charged with explanatory inadequacy for its

[154] Andrew M. Kenefick, *The Constitutionality of Punitive Damages Under the Excessive Fines Clause of the Eighth Amendment*, 85 MICH. L. REV. 1699, 1721-22 (1987).
[155] *Kemezy*, 79 F.3d at 35 (emphases added).

implication that, *if this question is ever asked*, the answer should be to apply the ER. Indeed, the STB+'s implication that the ER should apply when the state seeks punitive damages is nothing like other accounts' implication that the ER should apply when the state seeks *compensatory* damages. Only the latter implication is inconsistent with the settled law.

What of punitive damages cases where a private plaintiff acquired evidence by culpably violating the defendant?[156] It's easy enough to understand why the law does not apply the ER: it restricts the ER to constitutional violations, and private individuals cannot independently violate one another's constitutional rights.[157] If ER doctrine on this point were to change, as Section IV.B recommends, this question would become live, and one lesson of the STB+ would be that punitive damages should not be awarded to a plaintiff on the basis of evidence the plaintiff acquired by culpably violating the defendant's rights. That is no bullet to bite.

2. Deportation Proceedings

The ER does not apply in deportation proceedings because deportation is not punishment—at least in the eyes of the law. It is no objection that deportation is cruel, inhumane, hateful, or unjust. Grant all of that; it doesn't follow that deportation is *punishment*. True, deportations are sometimes predicated on criminal convictions. But they need not be,[158] and many of the most unjust deportations are of persons who have lived unimpeachable lives—genuine innocents.

Even if we were to grant that deportation is a form of punishment whose legitimacy depends on the standing to blame, this would be no objection to the STB+. To issue a refrain I'll

---

[156] *See* WAYNE R. LAFAVE ET AL., 2 CRIMINAL PROCEDURE § 3.1(h) (4th ed. 2020) (describing "[t]he exclusionary rule and 'private' or nonpolice searches").
[157] *Cf.* Burdeau v. McDowell, 256 U.S. 465 (1921) (refusing to apply the ER in against a private plaintiff in tort because the Fourth Amendment "was intended as a restraint upon the activities of sovereign authority and was not intended to be a limitation upon other than governmental agencies").
[158] *See* 8 U.S.C. § 1227(a)(5).

repeat: this would instead be an objection to our law's mistaken assumption—here, that deportation is not punishment.[159]

3. *No Pre-Trial ER*

The pre-trial process is part of *criminal* procedure. Yet the ER does not apply in grand jury, bail, or probable cause proceedings. The reason is clear: the law denies that indictments, bail requirements (however demanding), and pre-trial detention are punishment.[160] If these are not punishment, nothing in the STB+ suggests they depend on the legitimacy of blame.

Again, one could argue that the law is wrong about this—that some or all of these things are punishment. But again, the *refrain*: this would be an objection to the law's contrary assumption—here, about the nature of pre-trial proceedings and their outcomes.

4. *Impeachment*

The ER allows the use of tainted evidence to impeach the defendant's testimony, so long as the jury is instructed not to consider the evidence in the context of determining guilt and the prosecution did not intentionally create the impeaching opportunity as a vehicle to admit the tainted evidence. The STB+ sheds no light on whether this policy reflects a reasonable conception of jurors' capacities. But assuming, with the law, that it does, the policy is justified by the fact that impeachment is not punishment and the evidence is not part of the jury's basis for its knowledge of guilt.

---

[159] To the extent that deportation proceedings might be the basis for an objection to my account, it's because some circuits *have* applied the ER at deportation proceedings in particularly egregious contexts. *See* Oliva-Ramos v. Attorney Gen. of U.S., 694 F.3d 259 (3d Cir. 2012). But these circuits are in the minority, and insofar as "particularly egregious contexts" are those that "undermine the probative value of the evidence," *Olivia-Ramos*, 694 F.3d at 272 (citing I.N.S. v. Lopez-Mendoza, 468 U.S. 1032, 1050-51 (1984)), these circuits' doctrine causes no problem for my account. The standing to blame does not explain why unreliable evidence is not admissible.

[160] For a statement to this effect in the context pre-trial detention, *see* United States v. Salerno, 481 U.S. 739, 748 *passim* (1987) (We conclude . . . that the pretrial detention contemplated by the Bail Reform Act is regulatory in nature, and does not constitute punishment before trial.").

The fact that impeachment is not punishment seems obvious and I say nothing more about it. The difficulty lies in the notion of the jury's "basis" for its knowledge of guilt, and the distinction between basis and but-for cause. For even granting the jury an unlimited power to compartmentalize, the tainted evidence will sometimes be a but-for cause of their knowledge of guilt. In fact, this is why the impeachment exception exists: out of concern that defendants will successfully bamboozle jurors into returning not-guilty verdicts *unless* the jury sees the tainted evidence.

Distinguishing evidence as a *basis* for knowledge from evidence as a *factual cause* of knowledge invites us into complicated epistemological territory. But we can decline this invitation, for the distinction is easy to see here. Consider, the jury would not have known a defendant to be guilty if the jury did not understand English. Yet its understanding of English is not part of the evidence upon which it concludes the defendant is guilty. That is to say, it is not part of the jury's basis for its knowledge of guilt. (Compare the jury's awareness of the video evidence showing the defendant at the scene of the crime, or of his brother's testimony that he'd confessed.) Similarly, the jury would not have known the defendant was guilty if it had been bamboozled by his lies, and it would have been bamboozled if not for its exposure to the tainted evidence. But that does not mean the tainted evidence was any more a basis for its knowledge of guilt than was its understanding of English.

To illustrate further, take a jury that concludes a defendant robbed a bank because multiple witnesses testified he did so. The defendant, for his part, testified that he was not in the state—but his twin brother was, and the witnesses have confused him for his brother. Had the jury thought the testimony plausible, the testimony would have been part of its evidence base. On the basis of all the evidence it was privy to, the jury would have reasonably doubted the

defendant's guilt. But when the prosecution cross-examined the defendant, it was allowed to submit a piece of tainted evidence that showed the defendant was in the state and has no siblings. The jury used this evidence only for the purpose of disregarding the defendant's testimony—of *not including* that testimony in its evidence base. The tainted evidence was a but-for cause of the jury's guilty verdict; but it was no more part of the basis upon which the jury concluded the defendant was guilty than was the defendant's testimony itself, which is to say, it was no part of the jury's basis for its knowledge of guilt.

You might object that in the example I've given, the tainted evidence does not affirmatively support the defendant's guilt; if it did, the case would be different. But that's a mistake, for we are assuming the jury's power to compartmentalize. Our concern was that, *even if* the jury is able to ignore the ways in which the evidence affirmatively supports the defendant's guilt, the evidence undermines standing because it is a but-for cause of the jury's guilty verdict. If you are concerned that a jury cannot possibly avoid drawing conclusions about the defendant's guilt on the basis of evidence that affirmatively supports her guilt, you are concerned about the law's view of juries' mental powers, not the relevant distinction between bases and but-for causes that underlies the impeachment exception.

The book version of this chapter would spend more time on impeachment.[161] But what's been said here is enough to show that the STB+ has a clear and viable explanation for the ER's impeachment exception.

5. Sentencing

[161] Specifically, it would grapple with the objection that both the defendant's testimony *and* the evidence that disproves it are necessarily part of the jury's basis for its ultimate conclusion. It would thus step firmly into philosophical territory we've tried to avoid. And it would attempt to explain why standing is undermined only when tainted evidence is part of the basis for knowledge of guilt—that is to say, why but-for causation is not enough.

The ER does not apply at sentencing. Yet of all the steps in the criminal process prior to the *infliction* of punishment (i.e., prison), sentencing may appear to be the most punitive and condemnatory. Could it be that judges are allowed to consider tainted evidence at sentencing for the same reason they're allowed to consider other inadmissible evidence at sentencing? No. Judges are allowed to consider other inadmissible evidence at sentencing "on the basis of 'epistemic exceptionalism'—the claim that judges' cognitive processes can be trusted to operate with greater competence and objectivity than those of laypersons."[162] Cognitive competence and objectivity have nothing to do with the ER, so the general rationale for allowing inadmissible evidence into the sentencing phase is simply inapplicable.

Before offering the STB+'s explanation for the ER's non-application at sentencing, it is worth stressing that no extant account can explain this feature of ER doctrine. If the courts are concerned about judicial integrity, for example, then it is nonsense to allow *judges*, i.e., the *judiciary in particular*, to use evidence too tainted to be put before the jury at trial.[163] And with respect to deterrence, allowing excluded evidence at sentencing gives officers a powerful incentive to commit unconstitutional searches. If they seek a conviction, they seek the punishment and consequences attendant to it. So if illegal searches produce evidence that can be used to enhance punishment, there is *always* a good reason for the conviction-driven officer to

[162] James R. Steiner-Dillon, *Epistemic Exceptionalism*, 52 IND. L. REV. 207, 207 (2019).
[163] For additional argument on this point, *see* WAYNE R. LAFAVE, SEARCH AND SEIZURE: A TREATISE ON THE FOURTH AMENDMENT § 1.6(f) (6th ed. 2020) ("Broader use of the exclusionary rule at sentencing is also supported by other recognized bases for the rule. If evidence obtained in violation of the defendant's Fourth Amendment rights is accepted by the sentencing judge and used by him to enhance the defendant's sentence, it is unquestionably true that the government has profited from its lawless behavior and that the judge bears the taint of partnership in official lawlessness."); *see also* Adam Triplett, *Iowa's Exclusionary Rule: Expanding the Application of the Exclusionary Rule During Sentencing Hearings Under the Iowa Constitution*, 63 DRAKE L. REV. 987 (2015) (making related arguments based on fairness).

commit unconstitutional searches to further his goals: "a more patent end-run around the exclusionary rule is hard to imagine."[164]

The STB+ offers what I submit is the best explanation available for the ER's non-application at sentencing. But it is not simple, nor is it uncontentious, and its success should be understood against the background of other accounts' failures to explain this doctrinal feature.

I begin with a shorter and slightly simplified version of the explanation. For purposes of punishment and blame, the jury's verdict plays the role of the state's knowledge of the defendant's blameworthiness.[165] Epistemic illicitness is thus a feature of the *verdict*, in the case of state punishment, and of the blamer's *knowledge of blameworthiness*, more generally. Yet the judge's sentence does not affect the epistemic licitness or illicitness of the *jury's verdict*. Her sentence is simply not part of the apparatus of epistemic illicitness in the state punishment context.

One might argue this is a failing of the apparatus, for the state's standing to blame will be undermined by the judge's sentence if the judge predicates her sentence on tainted evidence and the state proceeds to blame and punish the defendant on the basis of the additional wrongdoing or culpability the evidence brings to light. But I claim that we don't need to understand the state to do this, even when the judge chooses a high sentence within statutory limits on the explicit basis of such evidence. We can understand the state to be *constrained* by the sentence without understanding the state to adopt the sentence, or judge's rationale for the sentence, as its basis for punishing the defendant.

---

[164] Victor Jay Miller, *An End Run Around the Exclusionary Rule: The Use of Illegally Seized Evidence Under the Federal Sentencing Guidelines*, 34 WM. & MARY L. REV. 241, 241 (1992) (quoting United States v. McCrory, 930 F.2d 63, 70-71 (D.C. Cir. 1991) (Silberman, J., concurring)).
[165] *See* Section III.E.

Granting this, the state punishes the defendant on the basis of the *crime for which she was convicted*, and the judge's *sentence* cannot undermine the state's standing to blame and punish the defendant on that basis. So long as the sentence does not impose a magnitude of punishment the conviction cannot independently justify, the legitimacy of the state's punishment is secure even when the judge imposes a relatively harsh sentence on the basis of tainted evidence.

Now the more thorough version of the explanation. The first premise is that the judge is not the state. By "the state," I mean the executive branch of the government in its law-enforcement capacity. It is the state whose officers seek out lawbreaking activity, whose prosecutors present the state's case for condemnation in court, and whose agencies administer punishment after conviction. These groups are unified by the goal of enforcing the criminal law by ensuring that lawbreakers are punished. And they are unified in such a way that it makes sense to say that *the state* cannot prosecute a crime and punish a criminal on the basis of evidence that *the state* acquired by culpably violating the defendant's constitutional rights.

The judge, in contrast, has a different goal: to apply the law fairly. She may be paid by the state, true. But so are public defenders. If the judge shares the same goal as the prosecution, then she is not a neutral arbiter applying the law. She is a mole, and the proceeding is a sham. Insofar as the judge *is a judge*, she is, ipso facto, *not the state*.

The second premise is something we already know: standing is called into question only when the blamer is the one who violated her target. *The state* cannot blame a person based on evidence *the state* acquired by violating that person's rights. But the judge? The judge didn't violate the defendant's rights. *Her* blame is unimpeachable. And to the extent her sentence depends on the legitimacy of her blame, her sentence is legitimate even when predicated in part on evidence that would render *the state*'s blame illegitimate.

This premise may seem to prove too much. If the judge can rely on such evidence at sentencing, why can't the jury rely on it at the conviction phase? (The jury didn't illegally search the defendant's house, after all.) The answer is that if the jury returned a conviction on the basis of such evidence, the state could not legitimately punish the defendant on the conviction's basis. The problem, in other words, would not be with the jury's blame, e.g., with its foreperson's indignant pronunciation that the defendant is "guilty!" The problem would be that the verdict could not support the legitimacy of *the state's* infliction of punishment.

This explains why the judge can treat the defendant with indignation at sentencing, even when her blame is based on tainted evidence. But it does not explain why the judge can base her sentence on such evidence without undermining the state's standing to punish the defendant on the sentence's basis. Nor should it, for indeed, the state *cannot* legitimately punish the defendant on the sentence's basis to the extent the sentence is based on tainted evidence.

This is key to understanding the import of the third premise, that the state *need not* punish a defendant on the basis of his *sentence*. The state can punish a defendant on the basis of his *crime*—what he was *convicted* of. This is not a claim about what the state can *legally* do: sentencing is part of the required legal process. It is a claim about morally just punishment and legitimate blame. An accurate and justly obtained conviction for a legitimately criminalized wrong is a sufficient moral basis for the state to punish and blame a criminal up to some limit. If the statutory maximum sentence for the conviction is not higher than that limit, the state can legitimately punish the defendant on the basis of the conviction to any extent within the statutory range, *whatever* the judge's reasons for choosing a particular punishment in that range. In other words, if the judge's sentence falls within the statutory range, then whatever the judge's bases

for her sentence, the state may legitimately blame and punish the defendant, in accord with her sentence, *for the crime of which she's been convicted*.

What we've said thus far shows that there is a way to understand sentencing such that the ER need not apply. If we assume that statutory maxima are not higher than the limits on punishment that the corresponding criminal convictions can support, and that the state can be understood to punish defendants on the basis of their *convictions*, not their *sentences*, at least when the legitimacy of the punishment is at stake, then even when the judge employs tainted evidence and sentences harshly within the statutory range, the punishment remains legitimate.

Now, we can't be blind to the fact that statutory maxima are often much too high. But the fourth premise is that the law stipulates statutory maxima as the limits on punishment that given convictions can independently legitimate. We can criticize these stipulations from a moral perspective, and we will often be correct, but we will be disagreeing with the law on a substantive moral question that's only indirectly related to the ER and its foundations. Add to this a fifth and final premise, that the state is best understood to punish criminals on the basis of their convictions, not their sentences, at least where the punishment at issue would otherwise be illegitimate. If premises 1, 2, and 3 offered a way to understand sentencing such that the ER *need not* apply, premises 4 and 5 claim that understanding is enough of a mirror of reality to explain why the ER *does not* apply.

Notice how these premises coalesce with *Apprendi*, which held that any basis upon which a statutory maximum is increased must be proven at the conviction stage.[166] It follows from this holding that if evidence must be excluded at the conviction stage, it cannot be a basis for

[166] *See* Apprendi v. New Jersey, 530 U.S. 466, 490 (2000) ("It is unconstitutional for a legislature to remove from the jury the assessment of facts that increase the prescribed range of penalties to which a criminal defendant is exposed. It is equally clear that such facts must be established by proof beyond a reasonable doubt." (citations omitted)).

sentencing a defendant beyond the statutory maximum corresponding to her charged crime. This is precisely what the explanation given in this Section requires. If evidence barred by the ER can be considered at sentencing, that evidence cannot be used to punish the defendant beyond the statutory maximum, else the punishment that the state will inflict on the defendant will go beyond the blame it can legitimately subject her to.

To recapitulate: the judge is not the state, so the judge's standing to sentence and blame the defendant on the basis of tainted evidence is secure. Moreover, the state need not punish the defendant on the basis of the judge's sentence. The state can legitimately punish the defendant on the basis of her conviction—that is, up to a certain limit, which the law stipulates to be the statutory maximum penalty. Because the state is best understood to punish defendants on the basis of their *convictions*—at least when the state would lack the standing to punish them on the basis of their *sentences*—the state can legitimately punish a defendant within the statutory range even when the judge chooses a punishment at the top of the range based on tainted evidence.

Now, even if all this is correct, one might still argue that the ER should apply at sentencing because the law gets statutory maxima wrong far too often. When it sets them too low, issues of legitimacy aren't implicated. But when it sets them too high, it gives judges the power to sentence the defendant to a punishment that the conviction alone cannot support. When the judge does so on the basis of tainted evidence, a portion of the defendant's punishment will be illegitimate *even if* the defendant deserves it when the tainted evidence is taken into account. This is because the conviction alone will not be able to justify that portion of the punishment, and the state will lack standing to inflict it on the basis that *can* justify it. I suspect this state of affairs is not unusual, given how often statutory maxima are set far too high and how often tainted evidence may come in at sentencing. I thus suspect that the benefit we acquire by

allowing the judge to consider tainted evidence at sentencing is not worth the cost of illegitimate punishments. And yet to issue our refrain: this disagreement is with the law's choices of statutory maxima, not the STB+'s explanation for the ER's non-application at sentencing.

6. Probation and Parole Revocation

Probation and parole are odd animals. The specific procedures and entities responsible for administering them vary among jurisdictions, and some state courts have held that in probation- or parole-revocation proceedings (hereafter "revocation proceedings," for short), their constitutions' exclusionary rules *do* apply.[167] These variations complicate our discussion because they make it difficult to make fully general claims. But suffice to say that in general, the ER does not apply to revocation proceedings because revocation is not punishment in the eyes of the law. Instead, revocation is based on a solely forward-looking judgment that prison is a better *kind* of punishment than probation or parole, which are, themselves, punishment too. Because revocation is not punishment, it does not depend on the legitimacy of blame, and standing is not implicated.

An example may be helpful. You're being punished by your teacher: You must stand in the corner of the classroom. But now the janitor needs to clean the classroom, and your presence in the corner is a nuisance. So your teacher tells you to go to the cafeteria, where you're to serve out your punishment by eating quietly by yourself and thinking about what you did. You're happy about this: much better to sit and eat than stand in a corner with your teacher staring you down.

But when the teacher calls you back to the classroom to finish serving out your assigned punishment in the corner—the janitor now needs to clean the cafeteria—she is not adding additional punishment or punishing you anew. She's modifying the same punishment for

[167] *See, e.g.*, Commonwealth v. Arter, 637 Pa. 541 (2016) (holding that Pennsylvania's constitution requires application of the ER at probation-revocation proceedings).

forward-looking reasons. That's exactly what she did when she sent you to the cafeteria in the first place. She would not be adding punishment or punishing you anew *even if* she'd called you back because you'd been a nuisance to the lunch staff and chatted with your friends, so long as her reasons are forward-looking and no time is added to your sentence.

In the eyes of the law, probation and parole work just the same. "Rather than being an ad hoc exercise of clemency, parole is an established *variation* on imprisonment of convicted criminals."[168] Probation and parole are not granted because a criminal deserves better treatment. Their "purpose is to help individuals reintegrate into society as constructive individuals as soon as they are able, without being confined for the full term of the sentence imposed. [They] also serve[] to alleviate the costs to society of keeping an individual in prison."[169]

In the words of one scholar, probation and parole are systems of "social engineering."[170] In this respect, decisions whether to grant or revoke parole are much like decisions whether or not to involuntarily commit a person. All these deprivations of freedom *can* be based on criminal behavior, but none need be.[171] And even when parole is revoked or involuntary commitment is imposed because of criminal behavior, the aim is not to *hold responsible* or *condemn* the behavior. The aim is to ensure the safety and wellbeing of the community, to rehabilitate the subject, and to best distribute scarce resources. Punitive aims, like blame itself, are never supposed to come into play.

---

[168] Morrissey v. Brewer, 408 U.S. 471, 477 (1972) (emphasis added).
[169] *Id.*
[170] Brian L. Crowe, *The Exclusionary Rule in Probation Revocation Proceedings*, 13 LOY. U. CHI. L.J. 373, 375 (1982).
[171] *Morrissey*, 408 U.S. at 478 ("Typically, parolees are forbidden to use liquor or to have associations or correspondence with certain categories of undesirable persons. Typically, also they must seek permission from their parole officers before engaging in specified activities, such as changing employment or living quarters, marrying, acquiring or operating a motor vehicle, traveling outside the community, and incurring substantial indebtedness.").

That the law sees probation, parole, and their revocation this way is reflected in the many opinions denying that revocation is punishment,[172] and the many opinions denying petitioners basic protections of the criminal process on that basis.[173] For example, indigent parolees needn't be afforded an attorney, and the bases for revocation need only be proven by a preponderance.[174] Justice Thomas, in his opinion making clear that the ER is inapplicable in the revocation context, said that applying the rule would "transform those proceedings from a 'predictive and discretionary' effort to promote the best interests of both parolees and society into trial-like proceedings 'less attuned' to the interests of the parolee."[175]

It is worth stressing: the STB+ is not committed to the proposition that the law is *correct* to deny that revocation is punishment. It's enough that the law *does* deny it. The STB+ is fully consistent with the three-Justice dissent to Thomas's five-Justice majority opinion, which begins by stating that the majority's opinion rests upon a mistaken conception of the "actual function of revocation," which is, "[i]n reality . . . the same function as a criminal trial."[176] That is to say, *it is punishment*. The STB+ does not tell us if the dissent is correct—only that the dissent correctly identifies the crux of the issue.

* * *

[172] *See, e.g.*, Gagnon v. Scarpelli, 411 U.S. 778, 782-83 (1973) ("Probation revocation, like parole revocation, is not a stage of a criminal prosecution."); State v. Divan, N.W.2d 865 (N.D. 2006) (denying that probation revocation proceedings are criminal prosecutions); State v. Johnson, 26 A.3d 59 (Conn. 2011) (holding that probation-revocation proceedings are not criminal proceedings and are therefore not subject to the statute of limitations for criminal prosecutions).
[173] Alongside *Morrissey*, 408 U.S., *see, e.g.*, Pennsylvania Bd. of Prob. & Parole v. Scott, 524 U.S. 357, 367 (1998) (denying that the ER applies in revocation contexts because such proceedings are "predictive" rather than punitive); United States v. Brown, 899 F.2d 189 (1990) (denying petitioner the opportunity to present an insanity defense because probation revocation is not a criminal prosecution); cases cited *supra* note 172.
[174] *See Gagnon*, 411 U.S. at 787 (no attorney provided for the indigent); United States v. Bujak, 347 F.3d 607, 609 (2003) (preponderance is the standard of proof).
[175] *Scott*, 524 U.S. at 367.
[176] *Id.* at 370 (1998) (Souter, J., dissenting).

This Section, IV.A, has not attempted to explain every last facet and feature of the ER's domain of application.[177] But it has accounted for the major doctrinal complexities and illustrated STB+'s explanatory power in an area where extant accounts fall far short.

B. Constitutional Application and the Good Faith Exception[178]

This Section first discusses the ER's application to all and only constitutional contexts before turning to the good faith exception. These features of ER doctrine are well motivated but miss the mark for the same reason: they draw bright lines where gradation is required. To illustrate, I first retread a bit of ground from Part III.

Recall that epistemic illicitness requires both that the blamer culpably violated her target (condition 2.a), and that the blamer's culpable violation is at least as blameworthy as her target's own wrongdoing (condition 2.d). If I wrongfully read your diary and discover you've been breaking curfew to make out with your new boyfriend after his parents go to sleep, then my standing to blame you is in serious jeopardy. I seriously violated your privacy—in some ways,

---

[177] I have in mind the fact that Fourth Amendment ER errors are not reviewable on federal habeas if they were litigated fully and fairly in state court. *See supra* note 14. I do not attempt to explain this feature of the ER for several related reasons. First, the feature is inconsistent with the fact that Fifth and Sixth Amendment ER errors are reviewable at Federal Habeas. This asymmetry suggests that it is something about the Amendments in the light of Federal Habeas, not the ER, that explains this feature. Second, the feature is peripheral to the ER. It is, arguably, less a feature of the rule than a feature of ER error doctrine on collateral review, and it has never been expanded. And third, I suspect the feature is explained by contingencies of recent history, specifically a Court driven to limit Federal Habeas on grounds of federalism, comity, and efficiency that are not specific to the ER. The Court's concern in *Stone v. Powell* was that imaginable Fourth Amendment violations exist in so many cases that, in light of the expansion of Federal Habeas that had preceded the 1976 decision, the federal courts would be in the business of consistently undermining state courts. I therefore agree with Justice Brennan in dissent who claimed, "despite the veil of Fourth Amendment terminology employed by the Court, [this case] plainly do[es] not involve any question of the right of a defendant to have evidence excluded from use against him in his criminal trial when that evidence was seized in contravention of rights ostensibly secured by the Fourth and Fourteenth Amendments." Stone v. Powell, 428 U.S. 465, 502-03 (1976) (Brennan, J., dissenting). Rather, "the real ground of today's decision—a ground that is particularly troubling in light of its portent for habeas jurisdiction generally—is the Court's novel reinterpretation of the habeas statutes." *Id.* at 515. In sum, I do not think my account should explain this feature of the rule because I do not think anything about the rule that explains it.

[178] *See supra* Sections I.A.4-6 for discussion of the doctrines presently explained.

the most meaningful privacy you have as a teenager. You legitimately broke my rules, but if I think you committed a serious wrong, I'm a bad parent: I should be glad to learn that your secret confession involves a mere malum prohibitum committed to do something safe and innocent. Because I'm more blameworthy for my wrongdoing than you are for yours, I can legitimately reaffirm the importance of curfew, but I can't express the kind of indignation that would be legitimate if I had the standing to blame you.

Now change the case: what I learn from your diary that you've been putting small doses of rat poison in your sibling's food because you're embarrassed to have to take him to school every day and want him to be too sick to go. Put yourself in my shoes and consider how my own violation—reading your diary for no good reason—seems virtually irrelevant in the face of something so gravely serious. There *is* a question about whether I can legitimately blame you, and whether I should, but the question is whether you're so psychopathic that you don't understand right and wrong.[179] The question is not whether I have *standing* to blame you: I certainly do.

With this in mind, consider the ER's application to all and only constitutional violations. Even assuming that all constitutional violations are graver than all non-constitutional violations—a doubtful assumption—the gravity of the state's violation is only one element of the comparison we need to make. In order to determine the state's blameworthiness, we also need to know about its culpability or mens rea—the issue with which the good faith exception is concerned. And in order to know whether the state's culpable violation is at least as blameworthy as the defendant's, we need to know about the alleged crime.

---

[179] For a helpful discussion of the literature surrounding the culpability of psychopaths, *see* Neil Levy, *Psychopaths and Blame: The Argument from Content*, 27 PHIL. PSYCH. 271, 271 (2014) ("The recent debate over the moral responsibility of psychopaths has centered on whether, or in what sense, they understand moral requirements."). Perhaps the seminal work in this area is Watson, *supra* note 123.

Automatically applying the ER to constitutional violations means applying the rule even when the blameworthiness of the defendant's crime outweighs the blameworthiness of the state's violation. It thus means excluding evidence and releasing criminals even when the standing to blame does not require it—e.g., freeing rapists on technicalities. Inversely, applying the ER to *only* constitutional contexts means *not* applying it when the state's non-constitutional violation is more blameworthy than the defendant's crime. It thus means punishing defendants without the standing to do so—e.g., prosecuting a defendant for marijuana possession on the basis of evidence acquired through intentional violations of statutory and international law.

These doctrinal errors are explicable when we consider the American legal system's preference for bright-line rules over balancing tests.[180] The Constitution is exalted, so constitutional violations have a special gravity. The graver the violation, the more blameworthy it is, all else equal. Indeed, the ER's application to all and only constitutional violations would likely be *correct* if it were necessary to have a bright-line rule.

We might even query whether it *isn't* necessary. Perhaps judges should not be in the business of weighing and balancing different kinds of violations. Perhaps this business of 'comparing and contrasting' wrongdoings thrusts judges into a domain of unhindered discretion

[180] This preference is certainly not overriding, but it is at least a preference, all else equal, and it is traceable, arguably, as far back as Lord Coke. *See* C. La Rue Munson, *How Far Shall the Justice and Rights of the Particular Cause Prevail over a Strict Application of Established Rules of Law*, 13 YALE L.J. 124, 133 (1904) ("Lord Coke may well stand for an example of lawyers of his class; learned, profound, honest and wise, and who would measure what is right, just and fair between man and man by the rigid rules of law, rather than to measure the law by the rules of justice and right."). The preference is as clear today as it ever was thanks to Justice Scalia's influence. *See* Antonin Scalia, *The Rule of Law as a Law of Rules*, 56 U. CHI. L. REV. 1175 (1989) (arguing for rules over standards and balancing tests on the grounds of predictability and the judicial function); Jamal Greene, *The Age of Scalia*, 130 HARV. L. REV. 144, 165-66 (2016) ("Justice Scalia . . . sought to change the Court . . . . He wanted the Court to eschew balancing tests in favor of categorical and general rules."); *The Federalist Society 2016 National Lawyers Convention Showcase Panel II: Rules Versus Standards in Constitutional and Statutory Interpretation*, 53 TULSA L. REV. 539, 539 (2018) ("Justice Scalia revolutionized the caselaw he inherited from the Burger Court by eliminating as many balancing tests as possible and replacing them with rules.").

and goes beyond the judicial function.[181] In fact, even if you give no credence to these Scalia-esque worries, you might nonetheless worry about the feasibility and administrability of comparing the state's violation to a yet-unproven crime whose precise contours may not be clear at a pre-trial suppression-of-evidence hearing. Indeed, those least inclined to Scalia-esque worries may be the most concerned about the possibility that prosecutors will begin to pursue the gravest charges a piece of evidence might support: what would have been misdemeanor assault III will become felony, aggravated assault II, for anything less would result in successful suppression of evidence.

If these worries are sufficient to justify our blunt doctrinal rule, all the better for the STB+'s explanatory adequacy. But the worries are belied by the fact that many countries have exclusionary rules, and all of them involve comparing the gravity of the rights violation to the gravity of the crime the evidence supports.[182] America is not so exceptional that it, alone, is incapable of this kind of balancing. The worries are further belied by the recognition that U.S. judges *already* perform this kind of balancing: finding the demands of justice clear and overwhelming in a particular case, they cut back the ER itself to avoid dismissing the grave

---

[181] *See* ANTONIN SCALIA, A MATTER OF INTERPRETATION 41-47 (1997); Scalia, *supra* note 180.

[182] *See* Liptak, *supra* note 29 ("The United States is the only country to take the position that some police misconduct must automatically result in the suppression of physical evidence . . . . In every other country, it's up to the trial judge to decide whether police misconduct has risen to the level of requiring the exclusion of evidence."); Ma, *supra* note 29, at 313-17 (discussing England, Wales, Canada, France, and Germany, and noting that none has adopted a mandatory rule); Stribopoulos, *supra* note 29, at 90, 93 (noting that Scotland, and Australia have rejected an automatic exclusionary rule for a discretionary one); Bradley, *supra* note 29, at 380 (discussing Argentina's rejection of an automatic ER for a discretionary one); *id.* at 376-78 (discussing the European Court of Human Rights' adoption of a discretionary ER in *Khan v. United Kingdom*, No. 35394/97 (Eur. Ct. H.R. May 12, 2000)); S.E. Van Der Merwe, *The Good Faith of the Police and the Exclusion of Unconstitutionally Obtained Evidence,* 11(3) S. AFR. J. CRIM. JUST. 462, 462 (1998) (discussing South Africa's discretionary ER, according to which evidence must be excluded if admission "would render the trial unfair or otherwise be detrimental to the administration of justice").

crime before them on account of a constitutional violation that is barely culpable and barely a constitutional violation.[183]

Thus, though it is beyond the scope of this chapter to respond fully to the worries articulated above or to articulate a more specific balancing regime that goes toward addressing them, I do not pretend to think they might succeed. The rules produce too much injustice. This is underlined by the sense of injustice felt when murderers or rapists go free on "crummy technicalities"— indeed, these are the kinds of cases that give the ER a bad name.[184] In recognizing that these cases *are errors*, we legitimate that sense of injustice, and we legitimate the ER, too, by diagnosing, and beginning the task of excising, a feature that is unjust by the lights of its own foundations.

---

[183] *See* Giudo Calabresi, *The Exclusionary Rule*, 26 HARV. J. L. & PUB. POL'Y. 111, 111-13 (2003) (arguing that in close cases, judges and Justices distort doctrine in legally ad hoc ways, producing problematic precedents to avoid releasing "a clearly guilty murderer or rapist" on "technicalities"); *see also* Amar, *supra* note 2, at 799 ("Judges do not like excluding bloody knives, so they distort doctrine, claiming the Fourth Amendment was not really violated."). Randy Barnett makes a related point:

> [A] criminal court judge may be seen to have at least three responsibilities: to do justice, to protect the public from those persons perceived to be dangerous, and to protect the fourth amendment and other constitutional rights of defendants and the public. A judge who takes all three responsibilities seriously is faced with a conflict of these interests if she believes that the police have in fact acted illegally. When deciding whether to suppress evidence, whatever ruling the judge makes will be inconsistent with one or more of these responsibilities. What balance is to be struck will depend upon the preferences of the judge.

Resolving the Dilemma of the Exclusionary Rule: An Application of Restitutive Principles of Justice, 32 EMORY L.J. 937, 963 (1983).

[184] *See* Amar, *supra* note 2, at 799 ("The exclusionary rule renders the Fourth Amendment contemptible in the eyes of judges and citizens. . . . In the popular mind, the Amendment has lost its luster and become associated with grinning criminals getting off on crummy technicalities. When rapists are freed, the people are *less* secure in their houses and persons—and they lose respect for the Fourth Amendment."); Caldwell & Chase, *supra* note 78, at 52 ("The public perceives the Rule as a device used by cunning defense lawyers to suppress evidence no matter how slight the police error and no matter how crucial the evidence is to the conviction of a criminal the public thinks is guilty of heinous crimes."); Perrone, *supra* note 2, at 67 ("[I]t is difficult to dispute that many instances where evidence is excluded for extremely technical or unforeseeable Fourth Amendment violations seem perverse to our own sense of justice.").

We've been focusing on the ER's application to all and only constitutional violations, but much the same analysis applies to the good faith exception. To be sure, that exception is precisely correct that unless the rights violation is culpable, the ER should not apply. (That is a bright-line rule we can keep: no culpability means no blameworthiness.) But the doctrine treats genuine negligence that does not meet the standard of gross negligence as if it was *never* sufficient for *any* culpability. More likely, there are crimes petty enough to be outweighed by merely negligent but sufficiently grave violations. Requiring at least gross negligence will mean punishing some crimes without the standing to do so.

At least as important, the law must begin to account for the different degrees of culpability involved in mental states that meet the culpability threshold. Holding fixed the violation, there will be times when the ER should apply if the violation was intentional or knowing, for instance, but not if it was reckless or grossly negligent. Indeed, the law currently treats gross negligence the same way it treats intentionality in this context, and we know that is a mistake. As in other polities, ER doctrine must evaluate the state's *degree* of culpability as one part of a broader evaluation of the magnitude of the state's blameworthiness for its rights violation.[185]

The fact the STB+ diagnoses these errors and offers these prescriptions does not weaken its explanatory adequacy by much. For the STB+ makes these errors eminently explicable; the

[185] Discretionary exclusionary rules generally allow judges to consider the magnitude of an officer's culpability as one factor. *See*, *e.g.*, Robert A. Harvie, *The Exclusionary Rule and the Good Faith Doctrine in the United States and Canada: A Comparison*, 14 LOY. L.A. INT'L & COMP. L. REV. 779, 797-98 (1992) (explaining that officers' mens rea is one factor considered in Canada's discretionary ER); Van Der Merwe, *supra* note 182 *passim* (making clear the same for South Africa); Stephen C. Thaman, *Balancing Truth Against Human Rights: A Theory of Modern Exclusionary Rules*, *in* EXCLUSIONARY RULES IN COMPARATIVE LAW 403, 427-28 (Stephen C. Thaman ed., 2013) (listing "good faith of the officers" as one of the "plethora of factors" in the "broad and unstructured" discretionary balancing test applied in England and Wales). There may be exceptions to this general rule that I have not found.

errors are unique to America's ER; and the errors are also responsible for much of the felt outrage toward the rule. Some inconsistency is to be expected, and *this* inconsistency is a good one to find. Indeed, the STB+ is a more valuable contribution to law and legal scholarship for the clear path forward it prescribes on these issues.

C. Causation & Vicarious Assertions[186]

There are three causal features of the ER to explain. First, the rule only applies if the state's violation proximately caused its acquisition of evidence. Second, the rule only applies if its violation factually caused its acquisition of evidence. And third, when a court *erroneously fails* to apply the ER and the state obtains a conviction, that conviction will only be overturned if the error was a but-for cause of the conviction (we'll call this "remedy doctrine"). We discuss these in order before turning finally to the ER's bar on vicarious assertions.

1. Proximate Causation

Proximate causation is a general requirement for the transfer of blameworthiness from actions to outcomes.[187] For instance,

> Frank chooses to get too drunk despite knowing he might try to drive when blacked out. This is precisely what happens, and he causes an accident that injures others. Because Frank is blameworthy for getting so drunk, and because his drunkenness proximately caused the accident, Frank is blameworthy for the accident even if he was too drunk to be acting voluntarily when the accident occurred.

Cases like Frank's aim to illustrate the transfer of blameworthiness generally, so they focus on *bad* outcomes where it seems essential that blameworthiness obtains. But is it possible for blameworthiness to transfer to outcomes that aren't bad? Possible definitional complications aside,[188] I suggest the answer is yes.

---

[186] *See supra* Sections I.A for a review of the doctrines presently explained.
[187] *See* Michael Moore, *Causation in the Law*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY § 2.1 (2019).
[188] You might define blameworthiness in such a way that it simply *cannot* transfer to something that isn't bad. If so, fine; then we are dealing with blameworthiness*, and for convenience, I drop the asterisk in

Recent upheavals surrounding Heidegger provide an illustrative case.[189] Abstracting and simplifying: When a Nazi produces a valuable piece of philosophical work, we worry decades after his death that the work is tainted by his Nazism. It matters to us how remote or proximate the connection is between the Nazism and the work. If we conclude that Nazi ideology motivated key parts of the work, then even if we think the work is fully compatible with the rejection of Nazism, and even if we think it is philosophically brilliant despite its terrible background and author, we nevertheless see reason to reject and disregard it. What we see, I suggest, is transferred blameworthiness. "Taint" is our word for it.

I hesitate to use Heidegger's case because nothing is comparable to Nazism, and "taint" of this sort is everywhere. Say I give you a piece of jewelry which, years later, you learn I stole. The statute of limitations is passed, and the victim can't be found: I picked her pocket on a random street in New York City. You give it to the police, trying to rectify my wrong, but they return it. It's been too long; it's yours now. I say to you, "it's a good thing you have this beautiful piece of jewelry! Yes, what I did was wrong, but that's irrelevant now." Here, I miss the fact that my blameworthy action *is* relevant: that blameworthiness taints what is otherwise a good thing.

And notice, it would be different if there were no proximate causation, say because voluntary interveners broke the chain.[190] If Heidegger's work had been voluntarily reworked and edited by someone without Nazi sympathies, and if what we had now was that reworked piece,

---

what follows. I describe the phenomenon at issue in terms of the transfer of blameworthiness, but we generally call it "taint," and that may be the better term in this context.

[189] *See* Joshua Rothman, *Is Heidegger Contaminated by Nazism?*, NEW YORKER, April 28, 2014, https://www.newyorker.com/books/page-turner/is-heidegger-contaminated-by-nazism.

[190] Moore, *supra* note 187, § 3 ("Under the intervening human actor branch of the common law's superseding cause doctrine, there is no liability if a subsequent human actor (rather than a natural event) intervenes to 'break the causal chain' otherwise existing (because of counterfactual dependence) between the harm and the defendant's earlier act.").

it's plausible that that voluntary intervention would have broken the chain of proximate causation between Heidegger's Nazism and the piece. To the extent it would have broken that chain, the taint would have been limited. For this same reason, when victims use the wrongdoings committed against them to them to learn and grow and become stronger, that learning and growth and strength is not tainted by the blameworthiness of their victimization, even if it was a but-for cause of their newfound strength. Finally, if you decide to keep the jewelry I stole because the police, despite your honesty and offer to give it up to them, voluntarily refused to take it—*it's yours now*, they said—I think you're on to something. I don't know if you're right, because I don't know whether the police's intervening acts broke the proximate causal chain; but to the extent they did, the taint is lifted.

The uncertainty I've just expressed about the specifics of proximate causation reflects the fact that I do not have an account of it. Without such an account, I cannot say whether the specifics of the ER's attenuation doctrine are correct. But this is not a problem for this chapter's purposes. If the specifics of the ER's attenuation doctrine are incorrect, it will be because the Court's specific understanding of proximate causation is incorrect. The STB+ needs only to explain why there is a proximate causation requirement. What I've said so far is the first part of that explanation: Blameworthiness can transfer to outcomes that are not bad, but proximate causation is necessary. Thus, blameworthiness for a rights violation can transfer to the acquisition of evidence, but only when the violation and the acquisition are connected by proximate causation.

The second step is to explain why the state's standing is only undermined when this transfer occurs. We know the answer already: a blamer loses standing just when (i) she is blameworthy for a wrongdoing of her own, and (ii) something links her own blameworthy

wrongdoing to the blame she would direct at another. Proximate causation is an element of (ii). It is part of the *link* between the state's culpable violation and its blame, in a particularly clear sense: if blameworthiness doesn't transfer from the state's violation to its holding of the evidence, then whatever the relationship between the state's blame and its holding of that evidence,[191] the blame will not be tainted because the evidence will not be tainted.

In sum, epistemic illicitness requires a proximate causal connection between the state's blameworthy violation and its acquisition of evidence because without that connection, the state's blameworthiness does not transfer to her blame. And since epistemic illicitness requires proximate causation, the ER does too.

2. Factual Causation

Much the same goes for the ER's requirement that the state's violation be the factual cause of its acquisition of evidence. In the law, criminal or tort liability for an outcome depends on both a proximate- *and* factual-causal connection to a harm or wrongdoing.[192] We can thus infer that in the law's eyes, the transfer of blameworthiness from act to outcome requires factual causation. As the previous Section suggested, epistemic illicitness requires that blameworthiness transfer from the state's violation to its acquisition of evidence. Putting these pieces together, because the transfer of blameworthiness requires factual causation, at least in the law's eyes, and epistemic illicitness requires the transfer of blameworthiness, the ER requires factual causation.

---

[191] Recall: your right to blame depends on your knowledge of my blameworthiness; that knowledge is constituted in part by your epistemic justification for your belief, i.e., the evidence you hold; and that evidence carries the taint of your blameworthy violation. I hesitate to attempt greater specificity about the relationship between your right to blame and your holding of that evidence. But it's not obviously wrong to say that your right to blame is partially constituted by your holding that evidence—and that your very right to blame is therefore tainted by your own blameworthiness. *See* nn.133-139 and accompanying text.

[192] *See* Moore, *supra* note 187, § 2.1 ("The conventional wisdom about the causation requirement in both criminal law and torts is that it in reality consists of two very different requirements for liability. The first requirement is that of 'cause in fact.' . . . [T]he second requirement[is] that of 'proximate' or 'legal' cause.").

This is all the STB+ needs to say to explain the ER's factual causation requirement. If you deny the proposition that factual causation is required for the transfer of blameworthiness, then you deny the law's presumption to the contrary, not the STB+'s explanation. Additionally, if you disagree with the ER's specific factual-causation doctrine, including the stringent requirement of but-for causation embedded in the inevitable discovery doctrine, then you disagree with the Court's understanding of factual causation in the ER context. (There is an important objection to this move, predicated on the prima facie inconsistency between the stringent but-for causation requirement in the ER context and the more defeasible, 'swiss-cheese' version of but-for causation required in contexts like homicide. But because I think this objection relies on a mistaken understanding of factual-causation doctrine generally, and because responding to it requires substantial discussion of that doctrine, I state and answer the objection only in a footnote.[193])

[193] The objection is as follows.

> **Inconsistency Objection**: On the explanation I've provided, factual causation is a general requirement for the transfer of blameworthiness. The STB+ cannot explain why the factual causation requirement in the ER context would be inconsistent with the factual causation requirement for criminal and civil liability generally. Yet this is precisely the doctrinal situation.
>
> Elsewhere, if a member of the Bloods kills Victim Vince, it will not fly to prove that a member of the Crips would have killed Vince had the Crip not known the Blood would do it for him. Nor could I have murdered one of the recently executed federal death row inmates, even on the day of their execution, and successfully defended my murder charge on the ground that *inevitably*, the state would have killed that same inmate legally (that very same day, even!). The inevitable discovery doctrine, in contrast, permits the use of evidence acquired through a constitutional violation on the ground that the state would inevitably have acquired that same evidence legally. It even permits this when the legal acquisition would have come much later. In sum, outside the ER context, but-for causation is not *really* a necessary condition for factual causation. At most, but-for causation is *generally but defeasibly* required. In the ER context, however, but-for causation is a genuinely necessary.
>
> It is no problem for the STB+ if the Court's understanding of factual causation is *wrong*. It is a problem, however, that the Court's understanding of "factual causation" in the ER context is *different* from its understanding of the factual causation necessary for the transfer of blameworthiness in other contexts, for that suggests the inevitable discovery

doctrine is *not* in fact explained by the fact that factual causation is required for the transfer of blameworthiness.

To this, the STB+ has several interconnected replies.

First, the objection overstates the unity of factual causation doctrine outside the ER context. There is no unified factual-causation doctrine applied "elsewhere": factual causation doctrine is disunified and theoretically unmoored across doctrinal contexts. The doctrine is so sundry that canvassing it is beyond the scope of this chapter, but two recent cases illustrate the point: *Burrage v. United States*, 571 U.S. 204 (2014), and *State v. Brelo*, No. CR13580457, 2015 WL 3514818, at *1 (Ohio Ct. Com. Pl. May 23, 2015).

In *Burrage*, the Court claimed that factual (or "actual") causation requires but-for causation "at least where" the act in question "is not an independently sufficient cause" of the outcome. *Burrage*, 571 U.S. at 218. That is to say, but-for causation is required at least unless the act in question was sufficient by itself to cause the outcome. We might think that in the criminal domain generally, the independently sufficient cause exception has long been well accepted, and the inevitable discovery doctrine is an outlier because it flouts that exception. But this is not so. In *Brelo*, a court was confronted with one of a dozen men who'd fired 137 rounds into a car with two people in it. One victim was shot 23 times, the other 24. *See* Ben Gifford, State v. Brelo *and the Problem of Actual Causation*, 44 AM. J. CRIM. L. 157, 158-59 (2017) (detailing these particulars). The court found that Brelo, who fired over a third of the 137 rounds, inflicted at least one of each victim's four or five independently fatal gunshot wounds. *Brelo*, No. CR13580457 at *10. Yet the court also found the evidence insufficient to show that Brelo inflicted *all* the independently fatal wounds on each victim. *Id.* So the court acquitted him of both killings. Speaking of the victim Russell, the court said, "Brelo's deadly shot would have caused the cessation of life if none of the other three were fired, but they were and that fact precludes finding beyond a reasonable doubt that Russell would have lived 'but for' Brelo's single lethal shot." *Id.* at *11. The court is not saying that Brelo's fatal shots might have only come once the victims were already deceased; it is rejecting the independently sufficient cause exception to the but-for causation requirement.

Why would the court reject that exception to the case at bar? Two reasons seem clear. One is that it is false that the but-for causation requirement has a categorical exception for independently sufficient causes, even in the criminal context. The other is that Brelo and his compatriots were the *police*. And what both these reasons underline is that insisting on an exceptionally stringent but-for causation requirement in the ER context, where police misconduct is generally at issue, is not nearly so remarkable as the objection suggests.

The second reply on the STB+'s behalf is this: Just as the STB+ is compatible with the Court's holding a mistaken view of factual causation, so it is compatible with the Court holding *mistakenly inconsistent* views on factual causation. The inconsistency objection is worrisome because it suggests that that the Court isn't *really* talking about factual causation in the inevitable discovery doctrine. But *Burrage* is again instructive. There, the Court aimed to defend its view that but-for causation is a requirement for factual ("actual") causation and that "substantial" and "contributing" factors are not factual causes unless they are but-for causes. *Burrage*, 571 U.S. at 215-16. To do so, the Court argued that exceptions to the but-for causation requirement for factual causation are outliers by offering a laundry list of doctrines where factual causation has been interpreted to require strict but-for causation, including Title VII, the ADEA, hate-crime statutes, and "suppression of evidence obtained as a result of police misconduct[, which] requires, at a minimum, a but-for causal relationship between the misconduct and collection of the evidence." *Id.* at 212-14. This shows that, far from taking the ER's factual causation requirements to be an outlier, the Court takes them to be part of factual-causation doctrine's *core*.

Indeed, and third: The Court sees deviations from the stringent but-for causation requirements to be doctrinally deviant. In *Burrage*, the Court refuses to endorse *any* deviation. And state level criminal courts generally make their deviations extremely localized. In my example of Vince, who'd have been killed by a Crip had he not first been killed by a Blood, state courts often appeal to an "acceleration rule,"

Before moving on, we should note a final point about both factual and proximate causation. The STB+ treats factual and proximate causation as if each were binary: the state's violation is the factual and proximate cause of its acquisition of evidence, or it is not. But what if factual and proximate causation are matters of degree, just like culpability, gravity of a wrongdoing, and blameworthiness? If that's so, then ER doctrine is mistaken insofar as it, too, treats factual and proximate causation as binary—and so is the STB+ as currently formulated

Fortunately, the STB+ can easily be modified to treat either or both factual and proximate causation as matters of degree. The specifics would depend on *how*, precisely, partial causation affects the transfer of blameworthiness—something beyond the purview of this chapter. But if we assume isometric proportionality, for illustration, then 50% proximate causation and full factual causation would mean discounting the state's blameworthiness by half when comparing

---

according to which the Blood is guilty because he accelerated Vince's death. But putting aside the obvious point that this rule is extensionally inadequate, courts that explain its justification "usually offer some variation" on the point that "since human beings are mortal, *all* homicide merely hastens the victim's death." *See* Eric A. Johnson, *Cause-in-Fact After* Burrage v. United States, 68 FLA. L. REV. 1727, 1749; *id.* at 1746-50 (detailing the acceleration rule). This is key: the acceleration rule is predicated on *death* as the outcome—and given that, it certainly should not apply to the illicit acquisition of inculpating evidence.

These first three points of reply substantially weaken the objection. It may well be that the ER's factual causation doctrine differs from factual-causation doctrine in some other contexts; but this is because *any* factual causation doctrine differs from factual causation doctrine in some other contexts. Moreover, the ER's factual-causation doctrine, in the Court's eyes if not in fact, typifies the doctrinal norm: but-for causation is a necessary condition, not the piece of swiss cheese it sometimes appears to be when, e.g., criminal homicide is at issue.

With this in mind, the fourth reply is that the inconsistency we find is not surprising. On the STB+, the ER's good faith exception and its automatic application in constitutional contexts each operate to wrongly punish some defendants who've committed relatively petty crimes; but they also operate to wrongly free some defendants who've committed relatively outrageous crimes. The latter injustice is the more outrageous one. And so the upshot is that errors in ER doctrine outside the factual-causation context currently stack the deck *against* the state. Insofar as factual causation is such a theoretical quagmire, it offers a locus for the Court to correct the imbalance. Whether the ER's factual-causation doctrine is right *or* wrong, it would be inconsistent with factual-causation doctrine elsewhere *no matter what*, so we should expect the Court to choose a version of the doctrine that benefits the state; this 'unstacks the deck.' The fact the law has long preferred to condone rather than condemn police misconduct when given the choice only underlines this point.

its blameworthiness to the gravity of the defendant's alleged crime. I say no more about this because the specifics are superfluous to the explanatory adequacy and power of the STB+. The key is that the account can integrate the best theories of proximate and factual causation, even if they are not binary.

3. Remedy Doctrine

Aside from *Stone*'s ban on reviewing Fourth Amendment ER claims at Federal Habeas,[194] ER errors are treated the same as other non-structural errors. To merit relief, the error must potentially have been a but-for cause of the conviction. "Potentially" specifies a degree of probability determined by the current stage of review and the first stage at which the defendant objected to the error.[195]

One might think that given STB+, ER errors should merit automatic reversal because they entail that the state lacks standing to punish. But putting aside the fact that ER errors are quite unlike the so-called "structural errors" that sometimes merit automatic reversal,[196] ER errors do not entail that the state lacks standing to punish. The ER is a prophylactic that bars tainted evidence whether or not that evidence turns out to be essential to a conviction. If a court

[194] *See supra* note 14.
[195] *See* sources cited *supra* note 47.
[196] *See* Weaver v. Massachusetts, 137 S. Ct. 1899, 1907 (2017) ("[T]he defining feature of a structural error is that it 'affects the framework within which the trial proceeds,' rather than being simply an error in the trial process itself.' For the same reason, a structural error 'defies analysis by harmless error standards.'" (citations omitted) (quoting Arizona v. Fulminante, 499 U.S. 279, 309-310 (1991))). I cannot find a case where the Court explicitly deals with the question whether ER errors are structural. But there are myriad cases treating such errors as harmless, so the Court has *answered* the question even if it has never bothered to *ask* it. *See, e.g.*, Chambers v. Maroney, 399 U.S. 42, 52-53 (1970). Errors affecting the admission of a single piece of evidence are paradigmatic non-structural errors because, the Court presumes, the reviewing court can evaluate the counterfactual, *what if that piece of evidence hadn't been before the jury?* Compare, when a court erroneously deprives a defendant of paid counsel of their choice, this affects everything about the trial, so the counterfactual—*what if the defendant had had a different lawyer all along?*—is much harder to evaluate. *See* United States v. Gonzalez-Lopez, 548 U.S. 140 (2006) (holding that erroneous denial of a defendant's right to paid counsel of choice is a structural error for those reasons).

erroneously fails to apply the ER and the evidence turns out to be essential to the conviction, then the state lacks standing to punish, and the conviction must be overturned. But if the evidence turns out not to be essential, then though the court erred, the state's standing to punish will be secure.

If this is so, why does the ER apply to non-essential evidence? Because (1) it is often impossible to make a confident judgement to this effect *ex ante*; (2) the state loses nothing when the ER bars evidence that turns out to be superfluous to conviction; and (3) not applying the ER to such cases would force reviewing courts to make post-hoc speculative judgments that, if mistaken, will either permit the state to punish illegitimately or overturn a conviction unnecessarily. There is little question that the ER should apply to all tainted evidence, but the state's standing is not undermined when a court erroneously fails to apply the ER to tainted evidence that is ultimately superfluous to the conviction.

One might nonetheless worry: If standing is so important, why take the risk that a reviewing court will *wrongly* conclude that tainted evidence is superfluous to a conviction? I think this is a perfectly legitimate worry. But it is a worry about criminal procedure generally. Consider: If not punishing the innocent is so important, why require a defendant who presents newly discovered evidence on collateral review to prove *by a preponderance* that in light of the new evidence, "no reasonable juror" would have found her guilty?[197] Or, if the prosecution intentionally lies to the defense and claims it has no exculpatory evidence to turn over, and the defense does not object because it believes the lie, why require the defendant to later show that this lie caused *Strickland*-level prejudice—one level *higher* than a "substantial and injurious

[197] *See* House v. Bell, 547 U.S. 518, 537-38 (2006) ("[P]risoners asserting innocence as a gateway to defaulted claims must establish that, in light of new evidence, 'it is more likely than not that no reasonable juror would have found petitioner guilty beyond a reasonable doubt.'" (quoting Schlup v. Delo, 513 U.S. 298, 327 (1995))).

effect on the verdict?"[198] Punishing the guilty without standing is no less legitimate than punishing the innocent, after all. Yet such is American criminal procedure, justified, to the extent it is, by the need for finality and our system's lack of resources.

Another possible objection is that the STB+'s explanation of the ER's remedy doctrine is *ipse dixit*. It may follow from the account's formulation of epistemic illicitness that the state has standing so long as the tainted evidence is not essential to the verdict. That's just condition (2.c). But the problem is that we have no independent reason to believe that condition (2.c) is a feature of epistemic licitness, rather than something tacked on to the account just to explain the remedy doctrine. This is an important objection that calls out for a deeper explanation of epistemic illicitness and the standing to blame generally. Because I cannot offer that here, I provide an example to show that, intuitively, (2.c) is a real feature of epistemic illicitness.

Recall the case in which you violate your daughter's privacy by reading her diary and discover that she has been breaking curfew to make out with her boyfriend. But now modify the case: what you read is not a big surprise, for you'd already been monitoring when she returned at night and knew she'd been breaking curfew. Her excuse, the times you'd confronted her, was that she'd been studying late with a friend, but her testimony did not inspire confidence. What's more, one of her teachers had recently mentioned overhearing her making a plan to meet her boyfriend late in the evening—a night she'd again come home very late. And now, the night after you read her diary, you stay up to see that she's broken curfew again. You confront her and discover to your chagrin (and hers) that she has a hickey.

---

[198] *See* United States v. Dominguez Benitez, 542 U.S. 74, 82 (2004) (Scalia, J., concurring in judgment) (laying out the standards of prejudice in terms of their demandingness); United States v. Bagley, 473 U.S. 667, 681-82 (1985) (applying the prejudice standard of *Strickland v. Washington*, 466 U.S. 668 (1984), to prosecutorial errors to turn over exculpatory evidence); Kotteakos v. United States, 328 U.S. 750, 776 (1946) (applying the "substantial and injurious effect on the verdict" standard to non-constitutional errors preserved below and brought on direct appeal).

You violated her privacy when you read her diary, and you owe her an apology. But I don't think you have to bring that up now. You can blame her on the basis of what you know to be true, because what you know doesn't depend on the diary. If she says, "You're wrong! You don't know I've been doing that," you can reply that you weren't born yesterday and lay out the evidence you *legitimately* gathered. If there's a question about the legitimacy of your blame, it's whether the evidence you legitimately gathered is sufficient for knowledge—not whether the diary undermines your standing to blame despite being unnecessary to your case.

To be sure, if your daughter goes to her room and finds evidence that you read her diary, she can legitimately confront you, blame you, and demand an apology. She might reasonably accuse you of knowing what you know *only* because you read her diary, and then say that you can't blame and punish her for what you read in her private diary. But you have a successful reply. You're *not* doing what she accuses you of. Really, *you did not need the diary* to know what's going on. And yes, you do owe her an apology, and yes, you are sorry: she has every right to blame you for it. But no, it is not the reason you know what she's done: what you did and what she did are *two separate matters*.

Your daughter is still going to be upset with you. But just as she can legitimately blame you for reading her diary *despite* the fact that she has been breaking curfew and lying to you about it, so you can legitimately blame her for her wrongs *despite* the fact that you committed your own. If you aren't moved by this example, either because your intuitions are muddled by unclarity about whether your legitimately obtained evidence is sufficient for knowledge, or because you want an example in which you *first* base your case on illicitly obtained evidence and then try to *take it back* because it wasn't necessary, I offer another example in a footnote.[199] But

[199] In this version, your daughter's boyfriend, a weirdo, writes you an email confessing that he and your daughter have breaking curfew to make out in the park after it closes. Your daughter's been lying that

the key point is that the STB+ can explain the ER's remedy doctrine, and that's so even granting we need a fuller account of standing to explain the STB+.

4. Bar on Vicarious Assertions

practice for the school play has been running late, he adds. And in fact, he and your daughter are going to do the same thing tonight!

Somewhat enraged, you search your daughter's room for her diary and read it to see what *else* she's been lying about. There, too, you read confirmation of the things her boyfriend wrote (but nothing else incriminating). You stay up waiting for her that night. When she walks through the door, you confront her, and she has a hickey. You accuse her of breaking curfew again and again to make out with her boyfriend and lying to you about it. She denies it—*you don't know that!* In reply, you list your evidence: the hickey, the diary, and the boyfriend's confession. Your daughter is outraged (and only partly because her boyfriend inexplicably betrayed her trust). She's outraged that you would seriously violate her privacy and outraged you would use what you found to blame her, particularly when her wrongdoing is relatively petty, all things considered.

In fact, when you go to think about it, what your daughter's been doing is blameworthy, yes, but was it a justification for reading her diary, given how much she values her privacy? Everything else you found in there is what you should have expected from a good kid like her. You were angry, and that clouded your mind to the fact that you had no right to read her diary in a situation like this one. You really did screw up.

You owe your daughter an apology for what you did. But you can say, "you're right, I'm sorry. You have every right to blame me for that. Really. But *I did not need to read your diary* to know what you've been doing. Put it completely aside; I was wrong to ever mention it in this context. It's a different matter. The fact is, your boyfriend told me everything on a night you came home after curfew with a hickey: you have been lying again and again in order to break the rules of this house. *You cannot do that*, *and you are grounded for it.*" To be sure, you handled this badly. First you wrongly violated your daughter's privacy, then *you yourself* raised that violation in the context of your blame. Even still, your right to blame your daughter does not depend on the diary evidence, for your *knowledge* does not depend on the diary evidence. For that reason, your standing to blame is not undermined.

Because I've not given a full account of the standing to blame, I can give no deeper explanation for why your standing to blame survives in this situation. Nor would I think it remarkably strange to intuit, about the case I've just given, that your standing to blame is *at least* weakened by your mistakes. But I do take myself, in this Section, to have illustrated a coherent and plausible theoretical perspective from which ER errors, if they are not necessary for knowledge of blameworthiness or a conviction of guilt, do not undermine the standing to blame.

But if you remain unconvinced, I would only note that the worst case scenario is that the STB+ fails to explain ER remedy doctrine, and that would be a bullet the account can afford to bite. Because prejudice is a general requirement for reversal in criminal procedure, and ER errors are distinctly unlike the kinds of errors that merit automatic reversal, it would not be so unsurprising if law errs here: treating ER errors the way it treats similar errors rather than the way the ER itself requires. That my own intuitions align with the law would then underline the fact that, even granting that the standing to blame explains the ER, some doctrinal questions remain difficult.

The ER may apply when the state illegally searches Ann's home and finds evidence against her, but it never applies when the state instead finds evidence against her brother, Bob, who does not live or stay there. This is the bar on vicarious assertions.

The Supreme Court once described the bar as a matter of 'standing' in the legal-procedural sense[200]: just as Texas lacks 'standing' to sue Pennsylvania for purported voter fraud,[201] so Bob lacks 'standing' to raise the state's violation of Ann when he moves to suppress the evidence in his own criminal trial. If this were correct, then the bar on vicarious assertions would not reflect the *merits* of Bob's case for exclusion. It would stop the courts from ever reaching them. But the Court has rejected this conception of the bar on vicarious assertions and declared that it is a matter of "substantive" ER doctrine.[202] The STB+ comes to the same conclusion, which follows from its formulation of epistemic illicitness in terms of just two agents. The victim of the state's violation must be the one against whom the state would use its tainted evidence. Otherwise, the stated conditions for epistemic illicitness are unsatisfied, and the STB+ provides no reason to think the state lacks standing to blame. Bob's case for exclusion thus fails *on the merits,* as far as the STB+ is concerned, just as it would if, say, the state's search was not wrongful.

This shows that the STB+ can explain the bar on vicarious assertions. Yet the explanation provided is not fully satisfying, for it immediately raises the question why epistemic illicitness requires the victim of the state's violation to be the one against whom the state would use the tainted evidence. In other words, quite simply, why does epistemic illicitness require what it

---

[200] *See, e.g.*, Jones v. United States, 362 U.S. 257, 261-65 (1960) (treating the bar on vicarious liability as a matter of legal 'standing'); Rakas v. Illinois, 439 U.S. 128, 134-37 (1978) (criticizing *Jones*'s treatment of the bar in terms of legal 'standing' and holding it was "dictum . . . which we now expressly reject").
[201] Texas v. Pennsylvania, No. 155, ORIG., 2020 WL 7296814 (Dec. 11, 2020).
[202] *Rakas*, 439 U.S. at 135, 139.

does? Why *doesn't* the state lose the standing to blame Bob? After all, the state committed a culpable wrongdoing, and its knowledge of Bob's blameworthiness depends on the evidence it acquired by doing so.[203] But I must conclude this Section, and this Part, with a promissory note to provide this answer in other work. The STB+ as formulated clearly accounts for the bar on vicarious assertions, and my considered view is that there is no way to explain *this* feature of epistemic illicitness without offering a deeper analysis of the standing to blame than the constraints of this chapter allow.

V. CONCLUSION AND FUTURE WORK

This chapter's central argument is now complete.

Part III argued that the standing to blame is a real phenomenon and that an agent loses the standing to blame when it meets the conditions of epistemic illicitness. Because the standing to *punish* depends on the standing to blame, the state lacks standing to punish a defendant whom it lacks the standing to blame. The exclusionary rule functions to obviate cases in which the state lacks the standing to punish a convicted defendant because it meets the conditions of epistemic illicitness. This was the basis for the novel STB+ account of the ER. Part IV used the STB+ to explain the rule's doctrinal contours and to show that the account is uniquely well positioned to explain the actual ER. In effect, the STB+ succeeds where Part II's alternative accounts fail. If the chapter's central argument succeeds, then the STB+ is the best available explanation for the ER.

---

[203] This question may seem especially pressing when we compare hypocrisy. My bike riding undermines my standing to blame the other biker *whether or not* my bike riding wrongs *him* in particular. Or so it seems, given this chapter's formulation of hypocrisy. Recall, A is a hypocrite with respect to B's culpable wrongdoing if

a. A is blameworthy for the same type of wrongdoing, and
b. A is at least as blameworthy for her wrongdoing as B is for his.

But outstanding questions remain. I conclude by raising these questions in Section V.A, both to highlight what this chapter has *not* done and to outline key avenues for future work.

A. Outstanding Questions

1. What Explains Standing?

The deepest outstanding question is why the standing to blame has the conditions and structure that it does. This question was especially important in the context of the ER's remedy doctrine and its bar on vicarious assertions, but the question can be raised about any condition of epistemic licitness—and of hypocrisy, complicity, and meddling, too. The question demands a robust explanation that philosophers of blame and standing have not yet provided in full. This is the most important project for future work.

2. Justification v. Explanation

Does this chapter's central argument imply the ER is *justified*? In one sense, yes. We have reason not to blame or punish in the absence of standing, and the ER facilitates a state of affairs in which the state does not do so. (Compare a hypothetical explanation that the rule is predicated on a misreading of the Constitution; even if conforming to the Constitution is valuable, we have no reason to conform to a misreading.)

But in another sense, no. This chapter has not argued that the standing to blame is so *important* that it requires the state to release rapists and murderers to avoid punishing them without standing. On my own view, punishing a person without standing is unjust in the same way that punishing an innocent is unjust. But even if that's so, it doesn't follow that it's always wrong, *all things considered*, to punish without standing. You might think that the sense of "justice" I refer to is not so important, or that, even though "justice" is important, it is not always

overriding. So deep questions remain: In what sense is punishing a person without standing wrong? And is it *always so wrong* that it merits paying the costs of exclusion?

3. Apology, Amends, and Time

Many believe that standing can be restored when a person sincerely apologizes and tries to make amends. Moreover, when it comes to my standing to blame you for riding your bike down the sidewalk, the fact I rode my bike down the sidewalk years ago, but stopped because I realized it was wrong, is different from the fact I'm riding my bike down the sidewalk right now. In light of this chapter's account, should we maintain the ER, or should we implement a mechanism by which the state apologizes and makes amends for its constitutional violations in order to restore its standing to punish?

Whether the latter alternative is realistic depends on the specifics of how standing can be restored. It might be that the state cannot restore its standing *in order to punish* the person it violated: compare the characters in *The Good Place* who try to get into heaven by doing good deeds but fail because their motive is self-interest. And even if it's *possible*, it may not be feasible for the same reasons that replacing the ER with civil damages is unfeasible. Qualified immunity and related doctrines make it exceptionally difficult for the law to ensure that officers who violate a defendant's constitutional rights apologize or make amends to her for that violation.

But the underlying idea certainly deserves further consideration, especially if you believe, like I do, that apologies, amends, and time can generally mitigate one's *blameworthiness*. For there will be cases where, if the state can mitigate its blameworthiness even a bit, the scales will tip away from exclusion. And in the absence of careful analysis, there is also a possibility that faux apology will be thought to tip the scales when it does not.

4. Is the STB+ 'Majestic'? Is it 'Constitutional'?

Among the views criticized in Part II, the STB+ is most like the "majestic" accounts. Philosophically, how we categorize the STB+ matters little. But legally, it matters some, for there have always been members of the Court who have insisted that the ER has a majestic purpose beyond mere deterrence. The fact the STB+ is best understood as a majestic account thus gives it a bit of a doctrinal toehold. Perhaps the STB+ is never *stated* in ER doctrine, but it fits that doctrine better than the alternatives, and the idea of a hard-to-articulate but nevertheless fundamental, majestic basis for the rule has stayed alive in concurrences and dissents for decades.

A separate question is whether the STB+ has anything to do with the *Constitution*. The answer to this question matters because the ER has been incorporated against the states. Even if the ER is a requirement of morality or justice, the legal legitimacy of incorporation depends on whether the ER has a constitutional basis. If it does not, then while the states may still have overriding reason to implement an ER—and most do require an ER in their own constitutions[204]—it should not be imposed on them as a matter of federal constitutional law.

Yet the answer to this question is not clear. Richard Re's account suggested the ER is a requirement of procedural due process. But Re struggled to offer an account of due process that explained the ER's doctrinal features, ultimately gesturing toward an inchoate conception of due process in terms of the state's *authority*.[205] If the standing to blame can play the role of *authority* in Re's account, might it be possible that the STB+ can be understood as a constitutional

[204] *See* Thomas K. Clancy, *The Fourth Amendment's Exclusionary Rule as a Constitutional Right*, 10 OHIO ST. J. CRIM. L. 357, 383-88 (2013) (listing trends in the states).
[205] *See* notes 70-73 and accompanying text.

requirement of due process, thereby vindicating the ER's incorporation? This possibility, along with other ways the STB+ might vindicate incorporation, deserve further consideration.

5. Why not Involuntary Commitment?

The standing to blame is not at issue in cases of involuntary civil commitment because, in the eyes of the law, commitment is not punishment and—unless things are awry—does not involve blame. So if the *need* for exclusion stems from the nature of punishment, and much of the *cost* of exclusion stems from the costs of releasing a dangerous criminal back into the community, why doesn't the STB+ imply that we should involuntarily commit those whom we cannot legitimately punish?

I doubt the justice of this suggestion. And it is not doctrinally viable at present: mental illness is a necessary condition for involuntary commitment, even in the context of sex offenders. That said, many criminal acts will be evidence of personality disorders which, though not properly mental illnesses, are in the DSM, so this doctrinal point is not necessarily intractable. And many believe that the involuntary commitment of those who are dangerous to themselves or others is often permissible and necessary. So the STB+ does raise the possibility that involuntary commitment, not release, is the best response when a dangerous, guilty criminal cannot be punished in light of the standing to blame. Whether this is a possibility we should embrace or—as I think—deny merits further discussion.

6. Inference to the Best Explanation

Say the STB+ explains the ER better than any extant alternative. Does it follow that we should *believe* it? This is a question about inference to the best explanation, and when it permits

or requires inferences to the best *available* explanation.[206] How good must the best *available* explanation be to merit belief? Does the STB+ meet that standard? These philosophical questions go beyond this chapter. But we should feel secure, whatever their answers, that if this chapter's novel account is the best available explanation of the rule, then it urgently deserves serious analysis and scholarly discussion, and its prescriptive implications urgently deserve serious consideration, before the exclusionary rule is interpreted away.

[206] To be sure, the best available explanation does not always merit belief. There must be situations in which the epistemically best attitude toward a set of inconsistent explanations is to suspend belief, not believe the best of a bad lot.

CHAPTER 2

BLAME'S APPROPRIATENESS AND THE EVAPORATION HEURISTIC

Abstract: It is commonly assumed that blame is appropriate just in case its target (the one being blamed) is blameworthy. This chapter puts forward a tool to discern whether a given consideration is relevant to blame's appropriateness—*the evaporation heuristic*—and uses it to argue that *contra* the commonly assumed view, many facts relevant to blame's appropriateness have nothing to do with the target's blameworthiness. They include the standing to blame. The chapter thus highlights an underexplored question: what explains why each of these facts is relevant to blame's appropriateness?

A is blameworthy just when *A* is morally responsible for a wrong.[207] Whether the wrong in question must be *moral* is open to debate.[208] But the notion of a wrong belongs, broadly, to first-order ethics. Thus, the theory of blameworthiness is largely a theory of moral responsibility, which has at least two dimensions. The most often discussed is the dimension of freedom or control over one's actions or attitudes. But the epistemic dimension of moral responsibility—think, '*what did he know and when did he know it?*'—has also received ample attention in recent years.[209]

---

[207] I offer this as a stipulation, but one I take to be very widely assumed. *See, e.g.*, Lloyd Fields, *Coercion and Moral Blameworthiness*, 15 INT. J. APPLIED PHIL. 135, 135 (2001) ("A person is morally blameworthy if and only if she performs an action which is, on balance, morally wrong, and for which she is morally responsible."). Some writers define moral responsibility *as* blameworthiness or praiseworthiness. *See, e.g.*, Gideon Rosen, *The Alethic Conception of Moral Responsibility*, *in* THE NATURE OF MORAL RESPONSIBILITY 65, 66 & n.2 ("[T]o say that *X* is morally responsible for *A* is to say that *X* is either morally praiseworthy or blameworthy for *A*," noting this "entails that we are not morally responsible for neutral acts."). My stipulation is consistent with this definitional choice.

[208] *See* Andrew C. Khoury, 45 J. VALUE INQUIRY 135, 135 (2011) ("It is commonly held that agents can be blameworthy only for acts that are morally wrong. But the claim . . . should be rejected. Agents can be blameworthy for acts that are not morally wrong.").

[209] *See* RESPONSIBILITY: THE EPISTEMIC CONDITION (Philip Robichaud & Jan Willem Wieland eds., 2017) (collecting 16 new papers on the topic).

If *A* is blameworthy, then in some sense, *A* is *worthy* of *blame*. Blame is an attitude; resentment and indignation, forms of righteous anger, are its paradigms.[210] Blameworthiness is a high-stakes topic, in part, because blame is often a high-stakes enterprise. Knowing that someone blames you, i.e., that you are the *target* of blame, tends to hurt.[211] Blame motivates hostility, moreover, which often breeds further hostility. And it is often thought that just punishment, particularly at the state level, presupposes the blameworthiness of its target, perhaps because punishment involves or expresses blame.[212]

To say that *A* is blameworthy is not to say that blaming *A* will have good consequences. Blaming the murderous tyrant may well cause his behavior to worsen. Instead, it is to say, roughly, that *A* is an *appropriate* target of blame. This notion of appropriateness is closely analogous to the notions of *fittingness* and *correctness*. To illustrate, a desire is *fitting* just in case its object (the thing desired) is *desirable*. And a belief is *correct* just in case its object (a proposition) is *true*. Even if you'll be paid a million dollars to desire a cup of mud or to believe that 1+1 is 3, your desire will be unfitting and your belief incorrect. Just the same, blame for a target who is not blameworthy is inappropriate regardless of its effects.

These points sit comfortably with a widely assumed view:

> **APPROPRIATE IF BLAMEWORTHY** ("**AIB**"): It is appropriate to blame *A* for *X* if *A* is blameworthy for *X*.[213]

---

[210] *See* P.F. Strawson, *Freedom and Resentment*, *reprinted in* FREEDOM, RESENTMENT, AND THE METAPHYSICS OF MORALS 107 (Pamela Hieronymi ed., 2020) (construing blame in terms of resentment and indignation). I do not assume that guilt is the first-personal analogue to resentment and indignation, though this is a common view attributed to Strawson.

[211] *Cf.* Pamela Hieronymi, *The Force and Fairness of Blame*, 18 PHIL. PERSPECTIVES 115, 117 (2004) (describing blame's "characteristic depth, force, or sting").

[212] *See* Joel Feinberg, *The Expressive Function of Punishment*, 49 THE MONIST 397, 400 *passim* (1965).

[213] Gideon Rosen seems to endorse AIB when he first stipulates that "*X* is morally responsible for *A* iff *X* merits moral praise of blame for *A*," and then assumes that this is equivalent to "*X* is morally responsible for *A* iff it would be *appropriate* to praise or blame *X* for *A*." *See* Rosen, *supra* note 207, at 66; *see also*

On this view, the theory of blame's appropriateness is just the theory of blameworthiness, which is largely exhausted by the theory of moral responsibility. This explains why philosophers interested in blame have tended to focus on moral responsibility, rather than on blame's appropriateness *per se*.

---

infra note 219 and attached text (describing Rosen's rejection of the claim that hypocrisy can undermine blame's appropriateness). R. Jay Wallace seems to endorse the thesis in his early work when he says,

> [I]f you know that a moral obligation you accept has been breached, and there are no exonerating circumstances that you are aware of, and you *still* do not believe that the moral response of indignation or resentment would at least be appropriate on your part, then it seems doubtful that you really do hold the [target] morally responsible for his actions . . . .
>
> Blame is construed essentially in terms of emotions, but the emotions in question . . . have propositional contents that are fixed by their connection to moral obligations that we accept. Moreover the nature of this connection with moral obligations is such that the reactive emotions are made appropriate by certain kinds of beliefs, about the violation of the moral obligations we hold people to.

WALLACE, RESPONSIBILITY AND THE MORAL SENTIMENTS 76-77 (1994). Adam Piovarchy notes, as part of an argument against the view that "standingless blame" is "a kind of inappropriate blame," that "[i]n investigating what it means for blame to be appropriate, philosophers have generally focused on the conditions under which someone is an appropriate target of blame *in virtue of their wrongdoing*. . . . For example, Pereboom (2014) and Nelkin (2016) interpret the claim that S is an appropriate target of blame to mean that S is a *deserving* target of blame. This does not seem to generate any obvious explanation for how hypocrites could lack the standing to blame, because hypocritical blame still gives wrongdoers what they deserve." Piovarchy, *Hypocrisy, Standing to Blame and Second-Personal Authority*, 101 PAC. PHIL. Q. 603, 607 (2020). Dana Nelkin, *Accountability and Desert*, 20 J. ETHICS 173, 183, 185 (2016), claims that "X is blameworthy for action A in the accountability sense if and only if it is appropriate to respond to X's performance of A with the negative reactive attitudes such as resentment and indignation," and argues that "the same conditions are plausibly necessary and sufficient for both blameworthiness in the accountability sense and basic desert of sanction." She recognizes that it is an objection to her view that "the aptness of demands" requires "that the person making the demand must have the standing to make such a demand," which she associates with a "contractualist conception of accountability" that she rejects. *Id.* at 187. Michael Zimmerman claims that "[m]any philosophers have embraced" the thesis that "[s]omeone's being responsible for *x* consists in its being fitting to adopt some reactive attitude toward him in respect of *x*," and explains that "[t]he idea here is that, whether or not the person in question is in fact the object of some reactive attitude, if he is responsible for *x* then . . . it would be appropriate so to respond to him." Zimmerman, *Varieties of Moral Responsibility*, *in* THE NATURE OF MORAL RESPONSIBILITY, *supra*, at 45, 48 & n.6. Zimmerman rejects this thesis but endorses the nearby claim, "Necessarily, someone is responsible for *x* if and only if it is fitting to adopt some reactive attitude toward him in respect of *x,*" which is to say, a target's "responsibility *grounds* the fittingness of the reaction." *Id.* at 49. To his credit, Zimmerman recognizes that the standing to blame complicates his favored view. *Id.* at 56.

This chapter argues that AIB is false. Blame's appropriateness turns on far more than its target's blameworthiness. The analysis of moral responsibility, it follows, is only one part of the analysis of blame's appropriateness. Philosophers' focus on moral responsibility has thus left much to be explained.

Here's the plan. Section I further explains why AIB is such an appealing thesis. Section II articulates the competing views on AIB. Section III articulates a method—central to which is what I'll call *the evaporation heuristic*—for discerning whether a given consideration is relevant to blame's appropriateness. Section IV applies the heuristic to argue that the so-called *standing to blame* is a necessary condition on blame's appropriateness. Section V uses the heuristic to argue that there is more to blame's appropriateness than blameworthiness and the standing to blame alone. And Section VI concludes by articulating the explanatory challenge raised by this chapter's arguments.

I. The Plausibility of AIB

AIB is plausible because it is a simple and clear solution to a difficult problem. The problem is how to distinguish considerations relevant to blame's appropriateness from considerations that are only relevant to distinct issues like whether blame is good or whether you ought to try to blame. Consider an example.

> **DEMON:** A demon will destroy the world unless you blame your mother for the fact it is her birthday. You are certain she is not blameworthy for this.[214]

[214] *Cf.* Roger Crisp, Review, *Value . . . And What Follows. By Joel Kupperman*, 75 PHILOSOPHY 458, 459 (2000) (objecting to definitions of value that identify value with preferability because "there can be reasons for preferring things that have nothing to do with their value . . . . Imagine that an evil demon will inflict a severe pain on me unless I prefer this saucer of mud; that makes the saucer well worth preferring. But it would not be plausible to claim that the saucer of mud's existence is, in itself, valuable.").

We need to explain why it is not appropriate to blame your mother in DEMON. Closely related, we need to explain why you will have such a hard time blaming her, despite the fact it would be very good to do so and the fact you certainly ought to try if there is *any* chance you might be successful.

D'arms and Jacobson address this problem by arguing that an attitude is fitting or appropriate iff it represents its object veridically.[215] For example, the object of admiration is a person, and admiration represents its object as admirable. Admiration is therefore fitting or appropriate iff its object is admirable. Their view is buoyed by the semantic point that many of our attitude terms have corresponding evaluative predicates which appear to give their fittingness conditions. For example:

ADMIRATION: ADMIRABLE
CONTEMPT: CONTEMPTIBLE
DESIRE: DESIRABLE
DISGUST: DISGUSTING
ENVY: ENVIABLE
JOY: JOYOUS
PITY: PITIFUL/PITIABLE
REGRET: REGRETTABLE
SHAME: SHAMEFUL

And, particularly relevant for our purposes:

BLAME: BLAMEWORTHY

The D'arms and Jacobson view thus proffers an elegant argument for AIB.

1) Attitudes are fitting or appropriate iff they represent their objects veridically.

2) Blame of *A* for *X* represents *A* as blameworthy for *X* (and this is the extent of its representational content).

[215] *See* Justin D'arms and Daniel Jacobson, *The Moralistic Fallacy: On the 'Appropriateness' of Emotions*, 61 PHIL. & PHENOMENOLOGICAL RESEARCH 65, 65 *passim* (2000) ("[T]o call an emotion appropriate is to say that the emotion is *fitting*: it accurately presents its object as having certain evaluative features.").

3) So blame of *A* for *X* is appropriate iff *A* is blameworthy for *X*. (from 1 and 2)

4) So AIB is true: it is appropriate to blame *A* for *X* if *A* is blameworthy for *X*. (from 3)

II. The Debate and Piovarchy's Challenge

But however plausible the AIB might appear, it has not gone unchallenged. On the view I endorse, attributable to Angela Smith and recently advanced by Patrick Todd, a blamer's *hypocrisy* can undermine her standing to blame a blameworthy target; moreover, if a blamer lacks the standing to blame, this means her blame is *inappropriate.*[216] It follows from these premises that AIB is false: The appropriateness of blame requires more than the blameworthiness of its target. It requires, also, that the blamer satisfy a condition of non-hypocrisy. (Whether the standing to blame requires *more* than non-hypocrisy is a matter of dispute between Todd and me—a matter we'll reach in section IV.)

But many have denied that this is the correct way to understand the standing to blame. One common view is that the standing to blame is a condition on the propriety of *expressed* blame, not the appropriateness of *attitudinal* blame.[217] Macalester Bell argues that the standing to

---

[216] *See* Angela M. Smith, *On Being Responsible and Holding Responsible*, 11 J. ETHICS 465, 478-80 (2007) (stating that hypocrisy can "affect our standing to take up certain kinds of blaming attitudes, such as righteous indignation"); *id.* at 477 (making clear that "whether it would be appropriate for any particular person to have and express any particular [blaming] attitude" turns on her standing to blame); Patrick Todd, *A Unified Account of the Moral Standing to Blame*, 53 NOÛS 347, 348 (2019) (claiming "the fact that one's blame would be 'hypocritical' is a condition on even (so to speak) *feeling* blame," and explaining this in terms of the "inappropriateness" of standingless blame); *see also* G.A. Cohen, *Casting the First Stone: Who Can, and Who Can't, Condemn the Terrorists?*, 58 ROYAL INST. PHIL. SUPPLEMENT 113 *passim* (2006) (making similar claims but without speaking of appropriateness *per se*).

[217] *See, e.g.*, Marilyn Friedman, *How to Blame People Responsibly*, 47 J. VALUE INQUIRY 271, 271 (2013) (treating the standing to blame as a condition on "expressed blame," which has distinct appropriateness conditions from private blame because expressed blame "may be painful to the recipient or bring painful consequences in its wake and be undeserved by the recipient, . . . [It] is something that can be done irresponsibly."); Ori J. Herstein, *Understanding Standing: Permission to Deflect Reasons*, 174 PHIL. STUDIES 3109, 3118 (2017) (arguing that challenges to a blamer's standing are appropriate just when the blamer has expressed a directive—a kind of speech act that purports to give reasons—and that it

blame is nothing more than a set of conditions under which blaming tends to be *good* and that it therefore has nothing to do with blame's appropriateness.[218] And some have treated the debate as something akin to a verbal dispute, to be settled by stipulation. This is how I read Gideon Rosen, for example, who imagines blaming a person who robbed your apartment while you were out robbing someone else's. He says,

> You lack what is sometimes called "standing to blame" . . . . [But is] your resentment "inappropriate" in the intended sense? We could say "yes," noting instantly that even if it would be inappropriate for *you* to blame her, she might still be blame*worthy*, since that only requires that *someone* . . . be in a position appropriately to blame the agent. My own preference, however, is to say that even if it would be *odd* and *unreasonable* for you to blame her given your lack of standing, your resentment would nevertheless be appropriate in the intended sense.[219]

Perhaps the greatest difficulty for those who hold my view is to *prove it*. Adam Piovarchy, who denies that standing is necessary for appropriate blame, puts the point well:

> In investigating what it means for blame to be appropriate, philosophers have generally focused on the conditions under which someone is an appropriate target of blame *in virtue of their wrongdoing*. Here, then, is a challenge to the proposal that questions about standing to blame concern whether it is appropriate for an agent to blame: what reason do we have to think that blaming when one lacks standing is inappropriate *in the same way* that blaming the innocent is inappropriate? . . . The fact that an agent lacks standing to blame does not seem like the general kinds of [forward-looking] consequentialist considerations . . . that can make blame unjustified. But we cannot conclude from this fact alone that blaming when one

---

is a mistake to "nefariously deploy" challenges to standing to deflect not only the directive but also "the reason found in the attitude or emotion that the expressive merely conveys"). Cristina Roadevin treats standing similarly when she "argues that hypocritical blame renders blame inappropriate" because "[s]omeone should not *express* her blame if she is guilty of the same thing for which she is blaming others." *Hypocritical Blame, Fairness, and Standing*, 49 METAPHILOSOPHY 137, 137 (2018) (emphasis added). And Todd, *supra* note 216, at 348-50 argues that some but not all of the commonly discussed standing conditions are in fact merely conditions on appropriate expressions of blame.

[218] *See* Macalester Bell, *The Standing to Blame: A Critique*, *in* BLAME: ITS NATURE AND NORMS 263 (D. Justin Coates & Neal A. Tognazzini eds., 2013).

[219] Rosen, *supra* note 207, at 69.

> lacks standing is therefore inappropriate in the way that blaming the innocent is inappropriate.[220]

In the remainder of this chapter, I attempt to answer Piovarchy's challenge directly.

III. The Evaporation Heuristic

Imagine that you are hungry and find that someone has left you a delicious-looking chocolate cake. You love chocolate cake. Moreover, you're not on a diet, and you need the calories. You see the cake as *desirable*, and rationally enough, you *desire* it.

Are you imagining? Good. Now imagine getting close to the cake. You start to smell the cake, expecting a chocolatey smell—and then you realize something is off. This is not a *chocolate* cake. This is a *manure* cake. Do you still want it? Of course not! The instant you recognize that this 'cake' is anything but desirable, your desire *evaporates*.

This unfortunate example highlights a general fact. Attitudes that can be correct, fitting, or appropriate are responsive, in a distinctive way, to the attitude holder's recognition of considerations that show the attitude to be (in)correct, (un)fitting, or (in)appropriate.[221] It is easy to desire a manure cake, for example, so long as you *think* it is chocolate. But it's virtually impossible once you become aware that it is manure. That goes even if I promise you one million dollars to desire it. Much the same, it is easy to blame your mother for culpably invading your

---

[220] Piovarchy, *supra* note 213, at 607.

[221] *Cf.* Mark Schroeder, *The Ubiquity of State-Given Reasons*, 122 ETHICS 457, 459 (2012) (describing an "asymmetry of motivation" as a chief earmark of the difference between the "right" and "wrong" kinds of reasons for attitudes). I long framed this chapter's discussion in terms of the "right" and "wrong" kinds of reasons for blame before deciding that lens complicated more than it clarified. But I still believe that this chapter's points could correctly be articulated in those terms: considerations relevant to blame's appropriateness are "right kinds of reasons" for the attitude. For a seminal discussion of the right and wrong kinds of reasons for attitudes, *see* Wlodek Rabnowicz & Toni Rønnow-Rasmussen, *The Strike of the Demon: on Fitting Pro-Attitudes and Value*, 114 ETHICS 391 (2004).

privacy, but not for blamelessly having a birthday, even if the former blame promises to be counterproductive and the latter promises to save the world.

We could repeat these lessons for each of the attitudes discussed in Section 1, but let's keep our attention on blame and consider a case.

> **DEAN TILL**: You are a graduate student with minimal resources, studying at an exceptionally wealthy university. One day, you learn that you're slated to lose your student health insurance due to a technicality for which you are not at fault. Panicked, you reach out to various administrators asking for help, all of whom direct you to email Dean Frank Till. You write to Dean Till explaining the urgency of your situation and pleading for help. One day passes, then another, as you become increasingly anxious. It is only when you CC him on an email confirming your imminent loss of health coverage that he replies, a single line: "Please do not CC me on emails that are not relevant to my job."
>
> You're shocked and confused. Then, quickly, you're *outraged*. You knew that it was possible you'd be stonewalled. But Dean Till is the one you were told to contact. Talking with you about your situation *is his job*. For him to ignore you and then dismiss you like this is beyond the pale. "How *dare* he?", you think, fuming. You forward Dean Till's reply to one of the administrators who directed you to contact him, and you express your anger that he would treat you like this.

Imagine yourself in this position, experiencing the kind of blame attitude that seems manifestly appropriate. And try to hold onto it as we complete the story:

> The administrator responds, "Oh gosh, so sorry: you've been emailing Mr. Franklin Till, who works in Food Services. You should be emailing Dean Frank Till," whose email she proceeds to give you for a second time, seeing as you failed to copy and paste it correctly the first time. By the next day, Dean Till has resolved the situation in your favor.

I presume that these new details cause your blame to evaporate, as they do mine. This is no surprise: Blame is responsive to your recognition that Dean Till is innocent. His innocence entails his lack of blameworthiness, and blameworthiness is necessary for blame to be appropriate. But if we hadn't been sure whether Dean Till's innocence was relevant to your

blame's appropriateness, we could afford to be more confident now. The way your blame evaporates upon learning of his innocence is a reason to believe that, indeed, innocence *is* relevant to blame's appropriateness.

This, roughly, is what I'll call the "evaporation heuristic." If your blame *evaporates*, in the distinctive way described above, upon learning a new detail *X*, this is a reason to believe that *X* is relevant to (and undermines) blame's appropriateness. Now try applying it to another case.

> **TYRANT**: The ruler of the kingdom in which you live is a murderous tyrant. He's also highly insightful and can perceive whenever someone blames him. One day, the tyrant has your friend killed on a whim. Naturally, you blame him in the strongest sense possible. He has you brought before him, and as you stand, surrounded by his guards, you experience powerful resentment and indignation.

Now the twist:

> The tyrant says, credibly and truthfully: "I hate being blamed. I want you to stop blaming me. So how about this? If you stop blaming me right now, I'll give you a hundred gold pieces, enough to set you up for life. But if you don't stop blaming me right now, I'm going to have your neighbor killed. It's up to you! And remember, I can tell whether you blame me or not."[222]

The tyrant has given you powerful reasons to stop blaming him. Your economic security and your neighbor's life are on the line. But I submit that learning these details does not cause your blame attitude to evaporate. Not even a little. In fact, I suggest, no matter how hard you tried, you could not make your blame attitude disappear on the basis of such considerations.

Now consider a third case.

> **BIKER**: You are on autopilot, head in the clouds, as you travel home down a crowded sidewalk. Just ahead is a group of children and a bicyclist precariously traveling around them. You glance at the "No Biking" signs that line the sidewalk,

[222] This is an embellished version of a case from Rosen, *supra* note 207, at 69.

> then attend to the congestion and danger the biker is creating with his selfish behavior. Annoyed, then angry and indignant, you think, *This is why there's no biking on the sidewalk, jerk!*

Grant, as a feature of the case, that this bicyclist is just as blameworthy as you think he is, and your indignation is not disproportionate to his blameworthiness. Imagine yourself blaming the biker, angry and indignant as described, when,

> You suddenly realize something you'd forgotten with your head so far in the clouds: *You are riding your own bike on this very same sidewalk right now.* You have no excuse or special justification. Just like him, you're illegally riding down a crowded sidewalk past pedestrians and children, congesting traffic and creating risk. You don't doubt your counterpart is blameworthy, but nor do you doubt that *you* are blameworthy—and for the precise same thing you'd just been blaming him for!

If you're like me, then recognizing these additional details causes your blame to evaporate. The evaporation is immediate and effortless, and it does not involve a set of inferences, e.g., "if I'm doing this, then it must not be wrong." Any such inferences are dubious, and it's a feature of the case that the other biker is blameworthy *and* that your confidence in this fact is unshaken.

The evaporation heuristic thus comes to a clear result: the kind of hypocrisy that BIKER involves is relevant to blame's appropriateness. Smith was right: a blamer's hypocrisy can undermine the appropriateness of her blame. And the reason we should favor Smith's view is that recognition of one's own hypocrisy causes blame to evaporate in the same way that recognition of one's target's innocence causes blame to evaporate—the way distinctive to the recognition of considerations relevant to the attitude's *appropriateness*.

I do not claim the evaporation heuristic is infallible. If a blamer is irrational, we should not expect her blame to evaporate upon recognizing considerations in virtue of which it is

inappropriate. That is part of what it *is* to be irrational, and irrationality is commonplace. Moreover, the heuristic may only track what we pre-theoretically *take* blame's appropriateness to turn on, in much the way that the Trolley Problem and other cases in normative ethics may only show us what we pre-theoretically *take* morality to require. We could be wrong.

But I do suggest that the evaporation heuristic is the best methodological tool we have in this area. If it merely shows us what we take blame's appropriateness to turn on, so be it: it shows us that we pre-theoretically take at least some kinds of hypocrisy to undermine blame's appropriateness, and we pre-theoretically take AIB to be false. This is an important count against AIB—and one to which we will add more in the pages that follow.

*IV. **The Standing to Blame***

It is often said that there are several distinct ways to lose the standing to blame.[223] Hypocrisy is just one of them. Distinct from hypocrisy is *complicity*, in which, roughly, a blamer is involved in, and culpable for, the very same action for which she's blaming another. It is also said that one can also lose standing if one's blame is *meddling*, which means, roughly, that the wrongdoing at issue is 'none of the blamer's business.' And I have argued, in chapter 1, for an under-recognized fourth way, which I call *epistemic illicitness*.

1. *B* lacks standing to blame *A* for *A*'s culpable wrongdoing *X* if *B*'s knowledge that *A* committed X is *epistemically illicit*.
2. *B*'s knowledge that *A* committed *X* is epistemically illicit if
    a. *B* culpably violated *A*,

[223] *See* Todd, *supra* note 216, at 247-48; Bell, *supra* note 218, at 264.

b. That culpable violation was the factual and proximate cause of *B*'s possessing evidence on the basis of which *B* knows *A* committed *X*,

c. Without the aforementioned evidence, *B* would not know that *A* committed *X*, and

d. *B*'s culpable wrongdoing is at least as blameworthy as *X*.

The aim of this section is not to define any of these conditions precisely or to defend what definitions I've offered. Instead, the aim is to use the evaporation heuristic to suggest that blame can be rendered inappropriate by a blamer losing standing through complicity, meddling, or epistemic illicitness. One lesson, I suggest, is that part of *what it is* for *B* to lack standing to blame *A* for *X* is for *B*'s blaming *A* for *X* to be inappropriate.

A caveat: DEAN TILL and BIKER are each extreme cases. Realizing that one's target is wholly innocent, for example, may cause one's blame to evaporate more immediately and completely than would realizing, e.g., that one's target did wrong only because she is suffering from an episode of serious mental illness. Even still, the latter realization does cause blame to evaporate, albeit perhaps less immediately and completely, because serious mental illness does undermine blame's appropriateness. Similarly, I take it that the following cases may not cause one's blame to evaporate in quite so striking a manner as did BIKER. The question is whether the details in these cases affect one's blame in the same *kind of way* that recognition of considerations related to blame's appropriateness characteristically do, or, in contrast, if they affect one's blame in a way more like the ultimatum in TYRANT.

Consider, to start,

> **CORRUPTION:** Mark is a long-term employee of a morally scrupulous non-profit that helps low-income people find housing. You're a part owner of the company and Mark's direct supervisor. Recently, large sums of money have inexplicably gone missing, putting the company's future in jeopardy. Evidence comes to light

> that clearly shows Mark has been embezzling this money and using it to fund the purchase of a small yacht. There is no doubt that Mark is blameworthy for this.

Put yourself in the position of Mark's supervisor and consider the kind of blame attitude that would seem to be appropriate. This non-profit does morally good work; Mark has betrayed it and put the company in jeopardy for a maximally selfish end. Imagine feeling resentment, indignation, and anger at Mark about this. But now consider,

> You found out about Mark's actions before the evidence came to light. When you confronted him, he offered you a deal: "Don't tell anyone what I've been doing, and I'll have a sexual relationship with you." You accepted and have looked the other way since then.

I take it that these details affect your blame attitudes in the way characteristic of considerations that affect the attitudes' appropriateness. You chose to be complicit in Mark's wrongdoings to attain something you desired; to angrily blame him for those wrongdoings despite your complicity seems inappropriate, and at the very least, it seems substantially *less* appropriate than blame from others in the company who were not complicit.

You might wonder why the added detail wasn't something more straightforward, e.g.,

> You are the one who ordered Mark to engage in this embezzlement, and you've been demanding kickbacks ever since.

To answer: the line between hypocrisy and complicity is unclear and debated, and some have argued that the latter reduces to the former.[224] If you are as responsible as Mark for the very same culpable wrongdoing (embezzlement) committed for the very same reasons (economic greed), then the case seems relevantly identical to cases of hypocrisy. I constructed CORRUPTION

[224] *See* Todd, *supra* note 216, at 351-57.

so that you have not ordered or partaken in the embezzlement and your benefit is not economic, like it is for Mark, but something more lurid. It is still possible, of course, to insist that CORRUPTION is just an instance of hypocrisy broadly understood. You and Mark are both *corrupt*, to be sure. But if that's so, then CORRUPTION illustrates that the kind of hypocrisy that can undermine blame's appropriateness is broader than we might have thought.

Consider a related case:

> **MANAGER**: You are a manager at a restaurant. You receive a call from one of your employees informing you that Jeff, a new employee, caused a fire by using a dangerous chemical to clean the grill. You hurry to the restaurant and find Jeff sitting in the regional manager's office discussing what happened. Jeff explains that he asked another manager for directions by text, and produces the text he received in reply:
>
> > **NO MATTER WHAT, YOU MUST USE CHEMICAL X TO CLEAN THE GRILL.**
>
> The regional manager berates Jeff nonetheless, pointing out that the label on the chemical explicitly says never to use it for that purpose. "You were negligent here, and you should not have followed these directions." You find yourself agreeing with the regional manager and blaming Jeff for his role.

Put yourself in the position of blaming Jeff. His blameworthiness is substantially mitigated by the fact another manager told him to use the chemical. But he is still somewhat blameworthy for the fire given his negligence, and some blame is appropriate. Now add,

> You pipe up: "Jeff, who were you texting that told you to do this?" Looking uncomfortable and a bit bewildered, he slowly replies: "Well… *you* did. Last night. Don't you remember?" You drank heavily last night, and while you don't remember clearly, the text came from your number. In a stupor, you must have forgotten the word "NOT."

It seems clear that these additional details should affect your blame attitude. Jeff is still blameworthy for the fire. But it would be bizarre if *you* continue continued to blame him, given your role in the events.

Now turn to another kind of case.

> **SARAH**: You're in high school, and your friend Sarah has been having problems with her boyfriend. She treats him with kindness and respect, but he's cold and thoughtless toward her. Like any good friend would, you feel anger and blame toward Sarah's boyfriend for treating her badly. And indeed, he is blameworthy for treating her this way.

I take it that this situation is common enough. Insofar as you're Sarah's friend, your emotional investment in her relationship is reasonable, perhaps even virtuous. Your blame seems to be perfectly appropriate. However,

> When you say you're Sarah's "friend," you more mean "friend-to-be." Sarah doesn't know who you *are*, per se, and nor does her boyfriend. You haven't learned about her life through any illicit means—just the unavoidable gossip that goes around school. Moreover, you would *never* impinge on her privacy or "*meddle*" in her affairs. That's why you keep your blame attitude all to yourself.

This case functions a bit differently from the others. It invites you to recognize that if you were, in fact, Sarah's friend, your blame attitude would be appropriate. But because you are not her friend—and because her affairs are none of your business—you *at least* should not blame *to the same extent* a friend would. In other words, if you were rational, then recognizing the fact that her affairs are none of your business would cause your blame to evaporate in part or in full.

It is sometimes thought that "meddling" is not a condition on the appropriateness of attitudinal blame, but only on its expressions.[225] SARAH is constructed to avoid this possibility.

[225] This is another theme of Todd, *supra* note 216, at 348-50.

The fact you would not "meddle" in her life *through action* is compatible with the suggestion that your attitude is inappropriate, or at least less appropriate, because the attitude *itself* is meddling in the relevant sense.

Consider, finally,

> **DIARY:** You learn that your 16-year-old daughter has been breaking curfew to make out with her boyfriend. You rightly judge this to be a culpable violation of your rules, and you blame her for it, experiencing a sense of indignation that she'd do this.

Take it as granted that your indignation is not disproportionate to your daughter's wrong. But now consider how your attitude responds to the fact,

> You only learned this information because you snuck into your daughter's room, opened her desk drawer, found her diary, and read it. You had no good reason for this and knew you were violating your daughter's privacy when you did so.

This case illustrates epistemic illicitness, the focus of chapter 1. The way you learned about your daughter's actions makes her no less blameworthy. But the fact you learned about her actions by culpably violating her privacy, and the fact that your wrongdoing is arguably as blameworthy as hers, should cause your indignant attitude to evaporate in whole or in part. Your partner, if he discovered by licit means that your daughter had broken curfew, could appropriately be indignant. Moreover, you might rationally *punish* your daughter, or manifest *blame-like* attitudes and dispositions. But the anger and self-righteousness of sincere indignation are misplaced, for *you*.

At this point, I've offered five cases that use the evaporation heuristic to put pressure on AIB. BIKER illustrated hypocrisy, CORRUPTION and MANAGER illustrated complicity, SARAH illustrated meddling, and DIARY illustrated epistemic illicitness. It is possible to resist my

suggestion that in *every* case, the evaporation heuristic produces the result that considerations irrelevant to the target's blameworthiness are relevant to blame's appropriateness. But I submit that it would be implausible to insist that the evaporation heuristic produces that result in *none* of them. It follows that if the evaporation heuristic is a reliable guide in this context, AIB is false.

V. Being Blameworthy vs. Deserving Blame

Our argument against AIB has focused on features of *blamers* that are relevant to blame's appropriateness. This section suggests that AIB is false for another reason. Recall, this chapter began by defining blameworthiness as moral responsibility for a wrong. It also described moral responsibility in terms of a control condition and an epistemic condition. This much is orthodoxy. And it is a remarkable feature of orthodoxy that if *A* is blameworthy for *X* in this sense, then *A* is blameworthy for *X forever*.[226]

But consider:

> **MAX**: Max knows he shouldn't text and drive, but he does so anyway, and in the process, he loses control and drives his car into your lawn. Max isn't hurt, but the lawn was well-manicured, filled with flowers and ornaments—things you cared for—and much of it is destroyed in the accident. There is no question that Max is blameworthy.

Imagine being the homeowner and first discovering all this. Imagine angrily blaming Max for destroying something you care about because he chose to text and drive. At the time, your blame is clearly appropriate. But here's the thing:

> That was 10 years ago. You were angry and you let Max have it; the police did too, and he faced hefty fines and points on his license. But Max was sincerely sorry for

[226] This chapter is not the first to notice this surprising implication. *See* Andrew C. Khoury & Benjamin Matheson, *Is Blameworthiness Forever?*, 2018 J. AM. PHIL. ASS'N 204, 204 (2018) ("Many of those working on moral responsibility assume that once blameworthy, always blameworthy.").

> what he did. He pledged to never again text while driving, and he's kept his pledge. He also apologized profusely to you and immediately paid for the damages. With your permission, he even repaired your lawn himself and contributed more than he was obligated to. In the process, you got to know him and grew to care about him. One day, he asked if you'd be willing to forgive him for the mistake he made that day. You say, "Yes, I would. I do. That case is closed." He's been a good friend to you ever since.

It's not easy to pinpoint when or to say precisely why, but at some point in the past 10 years, your blame became inappropriate. Yet Max's blameworthiness—his moral responsibility for the wrongdoing—never changed. What features of the case might be involved in explaining why your blame became inappropriate? A number come to mind:

1) **TIME**: An abundance of time has passed.
2) **PAST BLAME/PUNISHMENT**: Max has already been blamed (by you and the police) and punished (with fines and points on his license) in substantial ways.
3) **SINCERE PENITENCE**: He was sincerely sorry for what he did.
4) **EXPLICIT APOLOGY**: He explicitly apologized to you.
5) **CHANGE IN CHARACTER**: He pledged to become, and did become, the sort of person who would not commit the same wrongdoing again.
6) **COMPENSATIONS**: He fully compensated you, and even gave more than he was obligated to.
7) **REPAIR OF RELATIONSHIP**: He took steps to repair his relationship with you and created a friendship.
8) **FORGIVENESS**: He requested and received your explicit forgiveness.

We needn't adjudicate whether each of these plays its own role in explaining why your blame is no longer appropriate. Nor need we worry too much about classifying them. You might think that (8) is best understood to affect your standing to blame, for instance, rather than whether Max is a deserving target of blame. But even if that is so, (1) through (7) seem to be

about whether *Max* deserves blame. And what this illustrates is something we already know too well. *Blameworthiness*, as we've defined it, is forever. *Deserving blame* is not.

VI. Conclusion

According to AIB, it is appropriate to blame *A* for *X* if *A* is blameworthy for *X*. We've seen that AIB misses at least two things: (i) that standing to blame is a condition on blame's appropriateness, and (ii) that being an appropriate target of blame requires more than being blameworthy. By leading us to equate blame's appropriateness with its target's blameworthiness, AIB paints a simple picture over a more complicated reality and masks the questions it raises. What makes it the case that a given consideration is relevant to blame's appropriateness? Why are both features of the blamer, like hypocrisy, and features of the target, like blameworthiness, relevant to blame's appropriateness? And what, at bottom, explains why a given instance of blame is appropriate? The remaining chapters of this dissertation articulate a view designed to provide answers.

CHAPTER 3

CONTRACTUALISM ABOUT BLAME

Abstract: This chapter begins by offering a view of attitudinal blame according to which blame constitutively desires punishment and thinks its target deserves that punishment from the blamer. Blame is appropriate just when its constitutive thought is true. The chapter then offers a contractualist analysis of deserved punishment in terms of unrejectable principles of the practice of punishment that permit those with standing to punish the target. The resulting view, *Contractualism about Blame*, holds that considerations are relevant to blame's appropriateness just when they are relevant to the deontic outputs of those principles.

Chapter 2 presented the evaporation heuristic and used it to argue that blame's appropriateness depends on much more than its target's blameworthiness. That chapter left us with a question: what explains why a given consideration is relevant to blame's appropriateness? To answer that question, this chapter provides the framework for a unified, contractualist account of blame's appropriateness conditions, which I call *Contractualism about Blame* (or "CB" for short).[227] On CB, questions about blame's appropriateness are ultimately answered by explaining why certain contractualist principles cannot be reasonably rejected.

Here's the plan. Section I offers a full account of the kind of angry, punitive blame at issue in this dissertation. Section II describes the structure of contractualism, drawing especially from T.M. Scanlon's most recent work.[228] Section III describes how contractualist principles that

[227] Note that Contractualism about Blame has no connection to T.M. Scanlon's view of blame. *See* SCANLON, MORAL DIMENSIONS: PERMISSIBILITY, MEANING, BLAME ch. 4 (2008) (explaining blame in terms of social distancing and the revision of attitudes constitutive of relationships like friendship). Scanlon does not meaningfully employ contractualism in developing his view of blame in *Moral Dimensions*, and we should resist using *contractualist* as a synonym for *Scanlonian*.

[228] T.M. Scanlon, *Contractualism and Justification*, *in* REASON, JUSTIFICATION, AND CONTRACTUALISM: THEMES FROM SCANLON 17 (Markus Stepanians & Michael Frauchiger eds., 2021) (hereafter Scanlon 2021).

permit punishment explain the kind of desert involved in blame. And Section IV situates those principles within the *practice* of punishment, which, like the practice of law or medicine, has its own body of norms distinct from the demands of all-things-considered morality. Later, chapters 4 and 5 will further illustrate CB by employing it to explain why hypocrisy can undermine the standing to blame and why appropriate blame requires that the blamer meet an epistemic condition.

I. A Theory of Blame Attitudes

This section lays out a view of attitudinal blame, specifically resentment and indignation, of the harsh and angry kind discussed by P.F. Strawson, R. Jay Wallace, and Gideon Rosen.[229] I say nothing about the softer kinds of blame described by, e.g., Michael Smith and T.M. Scanlon.[230] (By "blame," I always mean the harsh kind.)

On the view I put forward below, blame essentially involves a *desire* for a kind of punishment and a *thought* that the object of that desire is *deserved*. Because my view is well understood as a modified version of Rosen's "alethic view,"[231] I proceed by describing the alethic view, raising problems for it, and arguing that my view overcomes those problems without giving up what makes the view attractive.

---

[229] *See* P.F. Strawson, *Freedom and Resentment*, *reprinted in* FREEDOM, RESENTMENT, AND THE METAPHYSICS OF MORALS 107 (Pamela Hieronymi ed., 2020) (construing blame in terms of resentment and indignation); R. JAY WALLACE, RESPONSIBILITY AND THE MORAL SENTIMENTS (1994) (understanding blame to involve moral opprobrium); Gideon Rosen, *The Alethic Conception of Moral Responsibility*, *in* THE NATURE OF MORAL RESPONSIBILITY 65, 82-85 (Randolph Clarke, Michael McKenna & Angela M. Smith eds., 2015) (describing blame as an attitude that constitutively desires punitive sanctions).

[230] *See* Michael Smith, *Romance and Responsibility in Woody Allen's "Manhattan,"* 20 J. ETHICS 317 (2016) (describing blame in terms of loss of trust); *supra* note 227 (describing Scanlon's view of blame).

[231] Rosen, *supra* note 229; *see also* Rosen, *The Problem of Pure Negligence*, *in* AGENCY, NEGLIGENCE AND RESPONSIBILITY 15 (Veronica Rodriguez-Blanco & George Pavlakos eds., 2021).

1.1. What Does Blame Desire?

The alethic view claims that part of what it is for *B* to blame *A* for *X* is for *B* to desire that *A* be *punished* for *X*.[232] Punishment is the intentional infliction of suffering for an ostensibly culpable wrong, so this claim captures the sense in which blame is a hostile attitude. My view endorses the claim once clarified, so let's unpack it further.

To start, it is potentially misleading to say that the punishment blame desires involves *suffering*. This is for two reasons. First, *suffering* has a connotation of intensity. On my view, the inflicted suffering can be extremely minor (e.g., a small pang of guilt) and still constitute punishment. Second, the primary sense of *suffering* involves pain or unpleasantness: negative affective states. But punishment might involve the deprivation of a positive affective state rather than the imposition of a negative one. Moreover, punishment need not involve the target's affective states at all. As Nozick's pleasure machine teaches us,[233] affective states are only one of the things that can make a person's life go better or worse. Even if I know it will make no difference to your affective state, I might punish you by hindering your projects, by causing another person to see you in a negative light, or perhaps simply by celebrating your misfortune, insofar as your life is made worse by others' treating your misfortune as an occasion to celebrate.

[232] Rosen, *supra* note 229, at 82-85; *see also* László Bernáth, *Blame and Fault: Toward a New Conative Theory of Blame*, 12 DISPUTATIO 371, 387 (2020) (claiming that *x*'s blaming *y* for *E* constitutively requires that "[i]t matters to *x* for its own sake whether *y* gets the punishment that *y* deserves for *E*.").
[233] For Nozick's clearest discussion of how this thought experiment should cast light on the nature of goodness in life, *see* ROBERT NOZICK, THE EXAMINED LIFE 104-06 (1989). For a rebuttal of common objections to Nozick's thought experiment, *see* Ben Bramble, *The Experience Machine*, 11 PHIL. COMPASS 136 (2016).

There is a broader sense of *suffering* that is consistent with these points. To suffer, in this sense, is to sustain injury, disadvantage, or loss.[234] To this broader definition, I would add that the injury, disadvantage, or loss might be very minor and still constitute suffering. So long as suffering is understood this way, I endorse the view that blame essentially desires the infliction of suffering on its target.

A caveat: say Beth believes she inflicts a kind of injury upon Al by celebrating his misfortune, but she is wrong about this. Or say Beth believes that yelling at Al will cause him psychic pain, but it will bring him nothing but pleasure. If Beth only desires to do things that she *mistakenly* thinks will inflict suffering, could her attitude nevertheless be blame? Yes. Blame desires to take actions the blamer *takes* to inflict suffering; the possibility Beth is mistaken is not relevant to the question whether her attitude constitutes blame.

Turning to a different point, the idea that blame essentially involves a desire for punishment would be false if punishment were understood to imply legal or formal sanctions. Blame usually does not desire sanctions of that sort in response to the minor culpable wrongs that pervade everyday life. For example, if someone cuts you in line at the store and you blame them for this, you might desire that they be made to feel guilty and embarrassed for what they've done; you need not desire they be charged with a misdemeanor or officially barred from shopping there again. The thesis that blame desires punishment must therefore be understood to pick out a kind of punishment divorced from connotations of legality or formality.

---

[234] Dictionary.com lists this as the second definition of the intransitive form of 'suffer' and the first definition of the transitive form. *See* SUFFER, DICTIONARY.COM, https://www.dictionary.com/browse/suffer. Merriam-Webster offers, as definitions of the intransitive form, 1. To endure death, pain or distress; 2. To sustain loss or damage; 3. To be subject to disability or handicap. *See* SUFFER, MERRIAM-WEBSTER, https://www.merriam-webster.com/dictionary/suffer.

Of course, the fact blame *desires* punishment does not mean that blamers necessarily *act* on that desire. But if you were to act on your desire to punish the line cutter, you might do little more than shoot him the stink eye. The stink eye and the silent treatment (for friends and family) are among the most common punishments we inflict on one another. If you thought the offense merited harsher punishment, you might say, bitingly, "Get back to the end of the line!" You might even add, "jerk!" If you had simply wanted him to go to the end of the line, being *mean* like this would be a questionable strategy.[235] But you didn't simply want him to go to the end of the line. You wanted him to feel guilt, embarrassment, or a degree of psychic pain for what he did. You wanted to *punish* him. And given that hostile words like yours tend to cause psychic pain, you'll likely have succeeded.

A terminological question: should we say that you *punished* your target if you did not succeed? Here, I prefer to put conceptual analysis aside and stipulate that punishment is a success term. So if your biting words cause your target nothing but joy, and they do not cause him to suffer in any other way (e.g., by attracting his boss's attention, who now judges him poorly), then you did not punish him; you tried and failed to do so. I'll use the term *punitive hostility* for attempts to inflict punishment. Thus, in the example above, you inflicted punitive hostility on your target; whether you punished him depends on whether you succeeded in causing him to suffer for what he did. We can add, as a general point, that because blame desires punishment, it tends to dispose the blamer toward punitive hostility.

I hope that by clarifying the claim that blame desires punishment, I've shown it is not implausible. The most important point is that the punishment blame desires can be informal and extremely mild. But the fact that blame does desire *punishment* nonetheless captures the

[235] As the saying goes, you catch more flies with honey than with vinegar.

harshness and hostility of blame attitudes and highlights the reason so many have thought blame unpalatable. I suggest these are reasons to think the claim true. The most important reason to favor the claim, however, is its explanatory power, which will become clearer as this chapter proceeds and the content of blame's desire is clarified further.[236]

1.2. How Many Thoughts in Blame?

According to the alethic view, blame of *A* for *X* is partly constituted by "thoughts," which have representational, propositional content.[237] Thoughts are belief-like because they are correct if and only if their contents are true, and because, insofar as a thinker is rational, her thought that

---

[236] One important objection to the claim relates to the idea that punishment involves the *intentional* infliction of suffering. Return to our example at the store and say that you don't act on your desire. You don't need to. Because the man cut you in line, he missed the sign pointing out a wet spot on the floor, so he slips and falls in a slapstick manner. He's fine, but deeply embarrassed, at which time the cashier, who hadn't noticed that he cut you in line, says "sir, oh no! Didn't you see the sign? It's impossible to miss when standing in line!" By the time the man has gotten up, the cashier has rung up your purchase.
In this situation, no one punished the man. No one intentionally caused him to suffer for what he did. But wasn't your blame's desire satisfied? If so, this suggests that what blame essentially desires can be satisfied by something very close to, but less than, punishment. So blame must not essentially desire punishment. We might call this the problem of *poetic justice*.
Responding to this problem in full is beyond this chapter's scope. But here is how I am inclined to think about it. First, to some extent, our intuitions about this matter may reflect a lingering sense—or, for some, a full-blown belief—that God, karma, or the universe has acted here. But I will put that possibility aside and assume we are clear that what happened to the man was simply bad luck. Second, to the extent we are clear that what happened was simply bad luck, I think there's a clear sense that our blame was *not* fully satisfied. What we wanted was not for *luck* to intervene, but for the man to face consequences *for* what he did, which means that what he did was an agent's *reason* for inflicting the consequences. Finally—and this is my main reply to the problem—it is common, when we desire *X*, to be largely satisfied by something that approximates *X*. If your child has gone missing and you desire strongly for him to walk in the door safe and sound, your desire will be largely satisfied if he walks through the door with a twisted ankle. You wanted him home *safe and sound*. What you got was not quite that. But it was close enough to satisfy your desire in large part. And of course, the fact you experience this satisfaction does not mean you wanted all along for him to walk through the door with an injury.

[237] *See* Rosen, *supra* note 229, at 71-72; Jada Strabbing, *Accountability and the Thoughts in Reactive Attitudes*, 176 PHIL. STUDIES 3121, 3124 (2018) ("I take thoughts to be belief-like propositional attitudes, representing how things seem to the agent."); ROBERT ROBERTS, EMOTIONS: AN ESSAY IN AID OF MORAL PSYCHOLOGY (2003).

p will disappear when she comes to know that ~*p*.[238] However, thoughts are also perception-like. They can be mere "seemings" or "strikings" that fall short of full-blown belief, as when one *sees* the Müller-Lyer lines as unequal despite having no doubt that they are equal. For the alethic view, the thoughts that constitute blame include (1) *A* did *X*; (2) it was wrong for *A* to do *X*; (3) in doing *X*, *A* showed an objectionable pattern of concern; and (4) *A* deserves punishment for doing *X*.[239]

On the one hand, the alethic view hopes to explain blame's appropriateness in terms of these thoughts. For example, why is blame of *A* for *X* inappropriate if *A* did not do *X*? Because (1) is false, so a thought constitutive of blame is false. To the extent this strategy succeeds, the alethic view is distinct from what Rosen calls the "moral view" of blame, which explains blame's appropriateness in terms of moral fairness or, as I prefer, contractualism.[240]

On the other hand, the alethic view recognizes that the question *when does A deserve punishment for doing X?* belongs to first-order moral theory.[241] It must therefore explain *deserved punishment* in terms of moral fairness, contractualism, or something else. The more explanatory work done by the view's fourth thought, the closer it comes to a version of the moral view.

To illustrate, consider a theoretical possibility: To think that *A* deserves punishment for doing *X* is, *already*, to think (1) *A* did *X*; (2) it was wrong for *A* to do *X*; and (3) in doing *X*, *A* showed an objectionable pattern of concern. The alethic view's fourth thought would then be

---

[238] In this, thoughts are unlike Tamar Gendler's "aliefs" which are *a*rational. *See* Tamar Gendler, *Alief and Belief*, 105 J. PHILOSOPHY 634 (2008). Rosen also notes that aliefs are paradigmatically affective, whereas thoughts need not be affectively charged. Rosen, *supra* note 229, at 72 n.14.

[239] Rosen officially states the fourth thought as *A* deserves *to suffer* for doing *X*, but he also states the thought in terms of punishment and punitive sanctions. *Id.* at 83-84. My understanding, from conversation with Rosen, is that the fourth thought is better understood in terms of punishment.

[240] *See id.* at 70 (distinguishing the moral view from the alethic view).

[241] *See id.* at 84.

doing all the work to explain blame's appropriateness conditions.[242] It would be the case that (a) blame of *A* for *X* is appropriate iff and because *A* deserves punishment for *X*; (b) *A* deserves punishment for *X* iff and because it's morally fair or contractually just for *A* to be punished for *X*; and so by the transitivity of explanation, (c) blame of *A* for *X* is appropriate iff and because it's morally fair or contractually just for *A* to be punished for *X*.

This illustration, in fact, describes the structure of my view. I do not think the alethic view's first three thoughts do any explanatory work. On my view, there is one thought constitutive of blame.[243] It is close to the alethic view's fourth thought, but importantly different, and we'll identify its content in section 1.4.

1.3. What Does Blame Think?

Chapter 2 argued that blame's appropriateness depends in part on the blamer's standing. Take that as true. How can the alethic view account for it?

I suggest it has only one option, pending other revisions. It must identify one or more *additional* thoughts constitutive of blame that are false just when the blamer lacks standing. I suggested above that the view's first three thoughts may be explanatorily otiose: collapsible into the fourth. But the view's fourth thought cannot explain why blame's appropriateness depends on the blamer's standing, for the fourth thought is exclusively about blame's target: *A deserves punishment for X*. Distinctions among blamers have no place in this thought. Rosen has floated

[242] This theoretical possibility is not, by itself, an objection to the alethic view, except in the minimal sense that it is a more parsimonious alternative.
[243] For those who dislike "thoughts," my view is compatible with any cognitivism picture on which to blame is, in essential part, to take, perceive, believe, or recognize something as true.

the possibility in conversation, following Jada Strabbing, that blame contains the thought "I wouldn't do *X*."[244]

I have three concerns with the strategy of adding one or more thoughts. First, I worry that blame is too *simple* for the strategy to succeed. I doubt that every time we resent a person, we hold multiple thoughts with distinct content (or even a single thought with multiple distinct conjuncts).[245] Second, while Rosen's suggested thought might account for the loss of standing in cases of hypocrisy and complicity, it cannot account for the loss of standing in cases of epistemic illicitness.[246] In the DIARY case from chapters 1 and 2, it may well be true that the parent wouldn't break curfew like his daughter did. But if you agree that his standing is undermined by the fact he learned about her wrongdoing by culpably reading her diary, you'll need to add another thought to explain this.

Third, and most importantly, I do not see why blame's *desire* should tend to evaporate upon realizing that an additional thought, like the one Rosen floats, is false. To understand this concern, notice how the alethic view's fourth thought relates to blame's desire. The view's fourth thought is that *A* deserves punishment for *X*, and blame desires *A* to be punished for *X*. One tends

---

[244] *See* JADA STRABBING, MORAL RESPONSIBILITY: ATTRIBUTABILITY, ACCOUNTABILITY AND CAPACITIES 70 (2011) (Princeton University Dissertation).

[245] One might object: *The fact that several distinct bits of new information can cause the attitude to evaporate suggests that it has at least this kind of complexity*. But I'm inclined to disagree. Consider how many distinct bits of new information can cause the belief that *Bob deserves the tenure-track job* to evaporate. They include, (a) Bob plagiarized part of his writing sample; (b) Bob successfully threatened the lives of his recommenders to give him excellent letters; (c) Bob regularly has sexual relations with his students; etc. The fact that (a) and (b) and (c) etc. can all cause the belief to evaporate does not entail that ~(a) or ~(b) or ~(c) are part of the content of the belief, or that the content is any more complex than *Bob deserves the tenure-track job*. (We could make this same point without desert. Consider how many distinct bits of new information can cause a belief like *Bob is wise* to evaporate.)

If the only way to account for blame's full set of appropriateness conditions were to posit multiple thoughts (or one thought with multiple conjuncts), then doing so would be necessary. A key implication of my view, however, is that it is not necessary.

[246] It could not account for the loss of standing in cases of meddling, either, but I do not focus here because I am more confident that epistemic illicitness undermines standing than I am that meddling does, too.

to desire what one thinks desirable, and when one thinks that *A* deserves punishment, one thinks there is something desirable about *A*'s being punished.[247] Granting this, one will tend to desire that *A* be punished if and because one thinks *A* deserves to be punished. Moreover, if *A*'s punishment appears desirable in one's eyes only because one thinks it deserved, then one will tend to cease to desire that *A* be punished when one ceases to think *A* deserves it.

I believe this is the alethic view's best explanation for the fact that blame's desire tends to arise when its fourth thought arises and to evaporate when its fourth thought evaporates. *But blame's desire tends to evaporate in the same way when one recognizes one's own hypocrisy*. And there's the rub: the connection between whether *A* deserves punishment for *X* and the desirability of *A*'s being punished, on the one hand, is unlike the connection between whether the blamer would do *X* and the desirability of *A*'s being punished, on the other. The desirability of *A*'s punishment for *X* is fixed by *A*'s deserving punishment for *X*, but a blamer's hypocrisy cannot undermine the fact that *A* deserves punishment for *X*, so it cannot undermine the desirability of *A*'s punishment for *X*.[248] It follows that the strategy of adding one or more thoughts to blame might explain why hypocrisy undermines blame's appropriateness, but it cannot explain why blame's desire evaporates upon recognizing one's own hypocrisy.

It is not obvious how we should explain why blame's desire tends to evaporate when the blamer recognizes she lacks standing, though we know we should not add additional thoughts to blame. The next section evaluates several proposals, and the view I endorse emerges from that discussion.

[247] I further discuss this claim, and qualify it slightly, in section 3.
[248] At most, a blamer's hypocrisy could lead them to *reconsider* whether *A* deserves to suffer for *X*; but I have argued that recognition of one's hypocrisy leads one's blame to evaporate in a very different way.

1.4. How Does Blame Involve a Thought about Standing?

Consider a first proposal: Blame thinks that punishment *from me*, i.e., punishment inflicted by the blamer herself, is deserved. The blamer desires to inflict punishment because she sees punishment from her as deserved, but when the blamer lacks standing, the target does not deserve punishment *from her*, so the thought is false. The thought thus evaporates when the blamer recognizes she lacks standing, and her desire evaporates because she no longer sees punishment from her as deserved. A first-order moral theory of desert would need to explain why the target does not deserve to suffer at the blamer's hand when the blamer is a hypocrite, etc. But success is now possible.

Note, moreover, that this proposal responds to my first two worries about the alethic view. First, it avoids positing multiple distinct thoughts in blame. And second, it makes the view compatible, in principle, with any number of ways that a blamer might lose standing, including epistemic illicitness.

The proposal faces a serious problem, however: blame's desire can be fully satisfied without the blamer herself doing anything. For example, I severely blamed the men responsible for Ahmaud Arbery's death. I was powerfully indignant. When I watched the men's guilty verdicts on live television, I was celebratory and overjoyed. My blame's desire was fully satisfied; there was no sense in which I wished that *I* had been the one to punish the men.

A caveat: my act of celebrating the men's guilty verdicts may have been an act of punitive hostility. In my view, it is a kind of harm, captured in the broad sense of "suffering," for others to *celebrate* your life sentence in prison. If that view is correct, then my celebration may have been a (very minor) punishment. Moreover, because *I believed* my celebration was a kind

of harm, I conceived of it as punitive, and this may explain why I was able to celebrate their grave misfortune with a sense of self-righteousness. But even still, it was not my celebration that fully satisfied my blame's desire. It was their guilty verdicts. And I had nothing to do with those.

The defender of the proposal has two options. First, she can argue that my blame's desire was satisfied because I identified myself with those who brought the verdict about: the jurors or the state. Many prosecutions, after all, are in the name of "The People [of State X]" and states always undertake prosecutions under the doctrine of *parens patriae*: they are acting *for* their citizens. But this argument is a stretch. I do not identify myself with the state of Georgia, which prosecuted Arbery's killers, or its citizens. I have never even been to Georgia. And no part of me is under an illusion that *I* can take any kind of credit for the prosecution.[249]

Alternatively, the proposal's defender might claim that blame's desire often comes in train with another desire. Blame desires that the blamer herself punish the target; the other desire is, perhaps, for the target to get the punishment they deserve. It is not *blame's* desire, then, being satisfied in these cases; it is the other desire. But this option is a last resort. It is *ad hoc*, it would multiply our explananda, and it seems to misdescribe reality. It seems to me that my *blame's* desire is what's satisfied in the cases I've described, and that this is not surprising because blame does not essentially fixate on *personal* punishment.

Let's turn to a second proposal. On this proposal, blame thinks that the target deserves punishment *from someone with standing* and blame desires that the target get what he deserves. Importantly, the proposal is not that every blamer, even a young child, has an articulate thought

[249] Moreover, the argument does not cover all the cases it would need to. If the men were found not guilty, I might hope that God would punish them appropriately in the afterlife. If I somehow learned that God did just that, my desire would be satisfied. But I do not identify myself with God. Similarly, if someone cuts me in line at a store and I blame them for it, my desire would be satisfied if the cashier rebuked her. But I need not identify myself with the cashier or imagine that *I'd* been hostile to the customer to feel satisfied.

about "punishment from someone with standing." The proposal, rather, is that every blamer, including any child capable of blame, has at least a rudimentary, *in*articulate understanding of what wrongdoers deserve for bad acts; that blame thinks its target deserves that inarticulate thing; and that, as a matter of fact, that thing is *punishment from someone with standing.*

The second proposal overcomes the problem we raised for the first. When I blame Arbery's killers, I think they deserve punishment from someone with standing, and I desire that they get what they deserve. Insofar as I have standing, my desire could be satisfied by my own acts of punishment toward the killers. But it could also be satisfied, as it was, by a guilty verdict in a fair trial that I had nothing to do with.

Unfortunately, this proposal faces a new problem. It is not clear how anyone's lack of standing—even *everyone's* lack of standing—could make the proposed thought false. Indeed, the fact that *no one* has standing is perfectly compatible with the thought that the target deserves punishment from someone with it. This is especially damning because it entails that the blamer's *own* standing is not a condition for her blame's appropriateness. In other words, the fact the blamer lacks standing is irrelevant to the truth of blame's constitutive thought because the wrongdoer *does* deserve punishment from someone with standing. And indeed, that's all her blame essentially desires.

The third proposal we'll consider is my view. I claim that what blame desires, and what it thinks is deserved, is punishment *from the blamer or anyone <u>else</u> with standing*. This is identical to the second proposal—that blamers think their target deserves punishment from those with standing—but it adds that *blame essentially presupposes the blamer's own membership in the privileged group*. Blamers essentially desire to move against the target themselves; but their desire can be satisfied if someone else with standing does so in the blamer's stead.

This may seem a surprising proposal. But we should endorse it because it best accounts for the phenomenology of blame and solves the problems we've raised for alternative proposals. Consider: blame is a kind of anger because it presents the blamer's *own* acts of punishment as desirable, and so disposes the blamer toward punitive hostility. It is also self-righteous, not just because it presents the punishment as deserved, but because it presents the blamer as part of the morally privileged group with standing to inflict it. The second proposal failed to explain why blame's constitutive thought is false when the blamer lacks standing, but on this proposal, the content of the thought includes the proposition that the blamer has standing to inflict the desired punishment.

At the same time, the proposal explains the Arbery case that caused problems for the first proposal. Blame desires punishment *from the blamer or anyone else with standing*, so its desire can be satisfied by punishment from any member of this privileged group. The proposal also sheds light on the tendency, in such cases, to say things like "*we* won," and to experience oneself as part of a winning team (those with standing, or perhaps, the "good guys"), despite the fact we may not identify ourselves with, say, the prosecutors in Georgia who obtained the guilty verdicts.[250]

Finally, the proposal claims that blame involves a single thought, but one that is capacious enough to be consistent with any number of ways that one might lose standing. It therefore overcomes the problems I raised for the alethic view without abandoning that view's explanatory power. The reason that blame's desire evaporates upon the blamer's recognition that

[250] An objection: I did not have standing to punish the men by sending them to prison. So I did not have standing to inflict the kind of punishment I desired the men to receive. The reply: I take it that the state has a special authority *among* those with standing to inflict certain kinds of punishments—those that involve coercion, in particular—but that this authority is separate from standing *per se*, and that the state has this authority because when it punishes, it does so *on behalf* of those with standing to punish.

she lacks standing is that what blame desires, and what blame thinks is deserved, includes the blamer's own punitive acts toward the target. When the blamer realizes she lacks standing, she realizes blame's constitutive thought is false, and thus that what her blame desires is not what her target deserves.

In sum, I endorse the following view: blame essentially desires punishment from the blamer or anyone else with standing to inflict it; blame essentially thinks that the target deserves punishment from the blamer or anyone else with standing to inflict it; and blame is appropriate just when its constitutive thought is true.

II. Contractualism in Outline

The question for my view, then, is how to explain the key notion of *desert* involved in blame's constitutive thought. Because the strategy is to use contractualism, this section provides a high-level overview of that theory. The version of contractualism I find most plausible follows T.M. Scanlon's most recent work,[251] which makes more room for aggregative and agent-neutral reasons than his earlier iterations of contractualism.[252] But I begin with more general observations about all versions of the view.

Contractualism is a theory of moral principles. Moral principles can be true or false, and contractualism tells us that the true moral principles are those that *no individual can reasonably reject*. In identifying those principles, contractualism purports to provide "a general account of (at least a central portion of) non-consequentialist morality."[253] That means contractualism does

[251] Scanlon 2021, *supra* note 228.
[252] *See* T.M. SCANLON, WHAT WE OWE TO EACH OTHER ch. 5 (1998) (hereafter SCANLON 1998); T.M. Scanlon, *Contractualism and Utilitarianism*, *in* CONTRACTUALISM AND UTILITARIANISM 104 (Amartya Sen & Bernard Williams eds., 1982) (hereafter Scanlon 1982).
[253] Scanlon 2021, *supra* note 228, at 17-18.

not purport to exhaust morality: morality may have a consequentialist dimension, and non-consequentialist morality may have multiple dimensions, only one of which contractualism captures.

What *role* do we imagine a principle playing in our lives, when we consider whether anyone can reasonably reject it? Scanlon speaks of "general agreement"[254] and "acceptance in general."[255] Thus, when we evaluate a candidate principle, we imagine a world in which the principle is generally, but not universally, accepted. As I understand Scanlon, "acceptance" is akin to H.L.A. Hart's conception of *internalization* of a legal rule. A judge who internalizes a rule has conscious access to it and generally follows it. But she often follows it unconsciously. Moreover, the principle often structures her deliberation, which is to say, if an option would violate the principle, the judge does not consider it in the first place. Even still, whether due to akrasia, irrationality, or the belief that the principle is overridden by conflicting moral imperatives, she may sometimes act contrary to what the principle demands.

Importantly, reasonable rejectability is a "comparative" notion.[256] The strength of any given agent's case for rejecting a principle is always relative to the strength of other agents' cases for accepting it. A nearly unobjectionable principle might be reasonably rejectable if there's an alternative that's wholly unobjectionable, and an exceptionally objectionable principle might win the day if every alternative is worse.

If each individual's reasons to favor or disfavor a principle were all and only the relevant *agent-neutral* reasons, then every individual's reasons would be the same. Contractualism would

---

[254] *See* Scanlon 1982, *supra* note 252, at 110; SCANLON 1998, *supra* note 252, at 153 ("[A]n act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behavior that no one could reasonably reject as a basis for informed, unforced general agreement.").
[255] *Id.* at 204.
[256] *Id.* at 195.

then collapse into a form of rule consequentialism. Contractualism avoids this result by placing restrictions and qualifications on the kinds of reasons that play a role in reasonable rejection.

The early Scanlon endorsed two restrictions on the reasons for rejecting a principle.[257] These were the *impersonalist restriction*, which excluded agent-neutral reasons but allowed agent-relative reasons, and the *individualist restriction*, which excluded the aggregation of individuals' reasons.[258] To endorse both restrictions is to say that the only reasons relevant to rejecting a principle are the *personal reasons of each individual*.

The later Scanlon (2021), moved by arguments from Parfit, has loosened these restrictions. (Rather than restrictions, we might now call them "qualifications.") To explain, the later Scanlon claims that he endorsed the restrictions because of "the distinctive kind of remorse" we feel when we take actions that cannot be justified to each person in light of the reasons they have.[259] He describes this as "the remorse test," i.e., "the idea that the particular way in which it is appropriate to feel bad about doing something that it is wrong indicates . . . the kind of factors that make an action wrong in that particular way."[260] But now, he writes, "I erred in concluding, from the fact that actions are wrong in this way only if individuals have sufficient personal reasons for objecting to principles that would permit them, that the reasonableness of rejecting a principle depends only on the comparison of individual personal reasons."[261]

Put differently:

> The remorse test suggests that when actions are wrong in the way that contractualism aims to describe they are wrong because of reasonable objections that some individuals would have to principles permitting them, objections based on personal reasons having to do with the way they would be affected by living

[257] As he describes in Scanlon 2021, *supra* note 228, at 25.
[258] *Id.* at 25-26.
[259] *Id.* at 27.
[260] *Id.* at 26.
[261] *Id.* at 41-42.

> under those principles. I concluded . . . that whether such an objection is reasonable depends *solely* on the relative strengths of the personal reasons supporting that objection as compared with the strongest personal reasons that any individual has for objecting to alternative principles. The latter step led me to the Individualist and Impersonalist restrictions. But this step was too quick.[262]

It was too quick, he explains, because "[e]ven if impersonal reasons, or aggregative reasons, cannot be grounds for rejecting a principle[,] . . . they could be considerations that can make such rejection *unreasonable*."[263] Scanlon therefore offers a "revised version of the contractualist idea of reasonable rejection" on which aggregative and agent-neutral reasons can play some role, albeit still qualified.[264]

Unfortunately, Scanlon offers only minimal guidance about the precise role these reasons play. Regarding what I'll now call the *impersonalist "qualification*," Scanlon continues to hold that if "impersonal reasons . . . are the only reasons for rejecting a principle," they "do not suffice" as a basis for reasonable rejection.[265] But they can make a difference when alongside personal reasons.

Say, for example, we must choose "between rescuing a small number of miners who are already trapped in a mine and investing the same resources in improved mine safety, which would save the lives of miners in the future."[266] Scanlon says, "the benefits of improved safety should be assessed *ex ante*, as a reduced likelihood, for current miners, of being trapped in the mine in the future." And "as the number of miners becomes larger it becomes extremely likely that a larger number of miners will be saved" by investing the resources in mine safety.[267] Unlike

---

[262] *Id.* at 31.
[263] *Id.*
[264] *Id.* at 41.
[265] *Id.* at 36.
[266] *Id.*
[267] *Id.* at 36-37.

the situation in which impersonal reasons are the only reasons at play, here, "many miners have personal reasons to want their risk of being trapped to be reduced (even slightly)." Thus, "the impersonal value of the 20 'statistical' lives that would be saved when the number of future miners is larger is relevant because it makes it unreasonable for the trapped miner or miners to insist on being saved."[268]

Regarding what I'll now call the *individualist "qualification*," Scanlon compares two cases. The first is the famous "Transmitter Room," in which Jones will experience immense pain unless we interrupt a very large group's enjoyment of a sporting event. Here, he says, the number of individuals whose enjoyment will be interrupted is not relevant because "the reasons of those in the larger group are very different" from Jones's reasons.[269] The second is a case where we can either save 5 people or a different group of 20 people. Here, he says, "the reasonableness of rejecting the principle [permitting us to save the 5] must take into account the reasons that all the individuals have to want to be saved, and the fact that there are more of these reasons counting in favor of the other action."[270]

We might wish that Scanlon provided more guidance on how to evaluate agent-neutral and aggregative reasons alongside each individual's personal reasons. But Scanlon doubts that more guidance can be given. His revised view, he says, comes "at the price of putting a great deal of weight on judgments about the sufficiency of reasons, in a way that weakens the explanatory character" of contractualism.[271] He acknowledges he cannot answer why it is, "in some cases, such as Transmitter Room, that certain conflicting reasons are, in Frances Kamm's

[268] *Id.* at 37. Scanlon notes that it is "not *obvious* to [him] that this is so (or obvious what the right thing to do is in this case)," but that it is precisely the kind of reasoning "opened up by the interpretation of reasonable rejection that [he] is discussing." *Id.*
[269] *Id.* at 33.
[270] *Id.*
[271] *Id.* at 34.

phrase 'not relevant' to a certain decision."[272] But while the revised view "seems to offer less clear guidance" than his earlier view, Scanlon thinks that "is a price that must be paid."[273]

For our purposes, I will understand Scanlon's revised view as follows, sticking close to the lessons of his examples. Each individual's personal reasons still have pride of place in contractualism and remain necessary to any successful case for or against a principle. However, agent-neutral and aggregative reasons are no longer automatically excluded.

Agent-neutral reasons are relevant so long as personal reasons are also in play. But agent-neutral reasons are always discounted relative to personal reasons. The extent to which they are discounted depends on the importance of the considerations that underlie them relative to the importance of the considerations that underlie the personal reasons. For example, the agent-neutral value of a statistical life is discounted little when placed alongside a personal reason based on mild enjoyment. The former will therefore be a stronger basis for rejecting a principle because (i) putting aside discounting, life matters much more than mild enjoyment, and (ii) the agent-neutral character of the life discounts it only a little. In contrast, the agent-neutral value of one agent's mild enjoyment will be fully or almost fully discounted relative to an agent's personal reason based on loss of life. And because mild enjoyment was already a far weaker basis for rejecting a principle, the mild enjoyment will be effectively irrelevant to the principle's rejectability.

Aggregative reasons are treated similarly. They are always discounted relative to individual reasons, but the extent to which they are discounted depends on the relative strength of the individual reasons that underlie them. When the underlying considerations are identical in importance (e.g., five lives vs. twenty lives) aggregative reasons are extremely relevant. When

[272] *Id.*
[273] *Id.* at 35.

the considerations are far apart in importance (e.g., one million agents' enjoyment of an event vs. one agent's extreme suffering), aggregative reasons are wholly irrelevant. And when the considerations fall between those two poles (e.g., many agents' limbs vs. one agent's life, or many agents' enjoyment of an event vs. one agent's great enjoyment of a different event), aggregative reasons are discounted. The extent to which they're discounted depends on where they fall between the poles. If they fall close enough to Transmitter Room, they may be entirely discounted.[274] That said, there does come a point, in my view, when enough agents object to a principle on the ground it will cause them to lose a limb, say, that their objection wins the day against a single agent who objects on the ground it will cause her to lose her life.

Finally, while these rules should guide our thinking, we must rely on *phronesis* to adjudicate hard cases, and for the same reason, we should not expect clearer and firmer rules than the ones I've outlined. Scanlon's revised contractualism "does not provide a clear standard for assessing" the reasons individuals have to reject principles,[275] but practical reasoners remain capable of assessing those reasons—and of confidently reaching clear conclusions across a wide range of cases. Where this claim does not hold true, contractualism still makes a clear, valuable *prediction*. That is, it predicts deep-rooted disagreements about what morality demands.

III. Contractualism and Desert

If *B* blames *A* for *X*, then necessarily, *B* desires that *A* be punished for *X* by someone with standing, a group that includes the blamer herself, and *B* thinks that what she desires is something *A* deserves. (For ease of discussion in what follows, I'll sometimes say, as shorthand,

[274] Transmitter Room is an *example* of irrelevant aggregative reasons. We cannot assume that if the aggregative reasons in the case were a bit more important, or Jones's personal reason a bit less important, then the aggregative reasons would pop into relevance (however substantially discounted).
[275] *Id.* at 35.

that *B* desires punishment and thinks punishment is deserved.) Because I wed this view of blame to a contractualist analysis of *desert*, I call the package *Contractualism about Blame* ("CB"). The task for this chapter is to explain how desert is analyzed in contractualist terms.

On CB, the fact that punishment is deserved in a particular case is explained by the fact that there exists an unrejectable principle. But what is the content of that principle? More precisely, if the fact that principle *p* is unrejectable explains the fact that a person *deserves* punishment, what precisely does *p* say?

The answer is not obvious, for orthodox contractualism only yields principles that obligate, permit, or prohibit some conduct. We can understand principles as propositions that make this explicit. For example,

P1: It is **obligatory** for patrons to leave tips of more than 20% at restaurants.

P2: It is **permissible** for patrons to leave tips of more than 20% at restaurants.

P3: It is **prohibited** for patrons to leave tips of more than 20% at restaurants.

Orthodox contractualism does not yield principles like,

P4: Servers **deserve** for patrons to leave tips of more than 20% at restaurants.

One could attempt to modify contractualism to yield principles like *P4*. But I am pessimistic about the prospects for this major modification to the contractualist apparatus. Hence the difficulty of our question. Insofar as *P4* is off the table, the contractualist principles that explain why hostility is deserved do not do so by saying anything about "desert."

A first attempt, consistent with orthodox contractualism, is to analyze desert in terms of permissibility. Wallace, for instance, tells us that "wrongdoers who possess the powers of reflective self-control deserve the responses of blame and moral sanction, in the sense that it

would not be unfair to respond to them in these ways."[276] Insofar as 'not unfair' entails 'permissible,' Wallace is offering a merely permissive conception of desert. Call this *weak desert*. If weak desert is the kind that needs to be explained, the relevant principles are like *P2*. They state that the punishment blame desires is permissible.

But there is a problem: the fact that something is permissible does not mean that it is desirable or that there is a reason to do it.[277] The thought that punishment is deserved in this weak sense, therefore, does not present it as desirable. Without something more, the thought cannot explain why blame's constitutive desire tends to arise with its constitutive thought. Nor can it explain why blame's desire tends to evaporate with the thought, for the thought was never what presented punishment as desirable in the first place.

This problem for our first attempt suggests a second: analyze desert in terms of obligation. If *A* deserves punishment from those with standing, those with standing face an obligation to punish *A*. Call this "strong desert." The fact that *A* deserves punishment in this strong sense entails that those with standing have a reason to punish *A* and that doing so is desirable. Blame's constitutive thought thus presents the blamer's punishment as desirable (because obligatory) and can serve to bring about blame's desire. On this view, the relevant contractualist principles are like *P1*: obligating principles.

I am inclined to grant that punishment is sometimes obligatory. For example, say you are the parent of an adult child who was abused by an authority figure when she was younger. Your daughter still struggles with the abuse and the abuser was never held accountable. One day, you encounter the abuser at a wedding with your daughter looking on. For no reason but your happy mood, you're polite and treat him like you would anyone else. I believe your daughter has every

[276] WALLACE, *supra* note 229, at 227.
[277] If this requires argument, consider that it's permissible for you to do a chicken dance right now.

right to feel betrayed and outraged. Your failure to act with any hostility toward her abuser was wrong. Given the severity of his wrongs and your position—as the parent of a victim who is looking on—you were *required* to treat him with hostility for what he did to her.

But even still, it is implausible that punishment is morally obligatory whenever blame is appropriate. Say someone cuts in front of you in a long car line because they do not want to wait 10 minutes (like you did) to exit the freeway. In case 1, you appropriately blame the driver and, acting on your desire, honk your horn and give him a disapproving look. In case 2, you just say "whatever." If the desert involved in blame was always strong desert, then the fact you blame appropriately in case 1 would entail that you violate an obligation in case 2. But this is doubly implausible. It's implausible that blame's appropriateness entails that punitive hostility is obligatory, and it's implausible that you violate an obligation in case 2.

We might consider returning to the Wallace-style view to explain cases like the car line. But recall that by itself, a thought about weak desert does not present punishment as desirable and cannot rationally give rise to blame's desire. More generally, weak desert is insufficient to explain paradigmatic cases of blame where the target's culpable wrongdoing strikes us as *warranting* or *justifying* punitive reactions without *obligating* or *requiring* them. These cases are incompatible with strong desert, yet inexplicable in terms of weak desert alone.[278]

---

[278] One possible solution would be to modify contractualism so that it yields principles about what's supererogatory, i.e., permissible and good to do but not required. These principles would have the form

P5: It is **supererogatory** for patrons to leave tips of more than 20% at restaurants.

I think this solution is costly. The modification to contractualism would be substantial, it would complicate the view, and it would require ample development. If CB *requires* such a modification to contractualism, that's to the view's detriment. To raise just one of many potential problems, goodness comes in infinitely many degrees, so principles like *P5* seem to be insufficiently specific. Is it *a little tiny bit* good to leave tips of more than 20%? Is it *enormously* good? Are these questions off the table for some reason? Otherwise, how does the contractualist framework, already unclear about many particulars,

I propose the following solution. In cases of blame where the target's culpable wrongdoing strikes us as *warranting* or *justifying* punitive reactions without *obligating* or *requiring* them, the kind of desert at issue is indeed weak desert, but the fact that punishment is weakly deserved is not what makes it desirable. Something else plays that role. This solution risks falling afoul of the phenomenology of blame, which seems to present the target's desert as a core *reason* punishment is desirable. But it need not fall afoul of this phenomenology if we correctly understand the relationship between weak desert and the considerations that make punishment desirable.

When punishment is weakly deserved, an unrejectable principle permits it, and the principle is unrejectable in light of certain considerations. Some of those considerations can, and generally do, play a dual role: they serve to make the principle unrejectable *and* they also make hostility desirable in the case at hand. To illustrate, when a driver cuts in front of you to avoid the long car line, a punitive reaction can serve to teach him a lesson, to make clear (to him and others) that his behavior is wrong, and to vindicate your and other drivers' value as persons who deserve respect. These are among the considerations that explain why a principle permitting punishment is unrejectable in the case at hand. But they are also considerations that make punishment desirable.

I therefore suggest that in the common case where we think hostility is deserved, and this seems to be a reason that favors (but doesn't obligate) hostility, we are thinking about weak desert *and* at least one consideration that plays this dual role. If you were pressed, for instance, *what good is it to beep and glare like this?*, you might very well answer, *he needs to learn a lesson*, or *I can't just let that stand*, or *he disrespected me and everyone else who waited in this*

---

pick out the unrejectable principle from an infinity of nearby alternatives? In sum, while I am not sure this modification cannot be made to work, we have reason to avoid it if we can.

line. The worry, however, is that these considerations could make punishment desirable whether it is permissible or not. Granting this, blame's desire would arise whether we think punishment weakly deserved or not, and it would not evaporate when we cease to think it weakly deserved. It would arise and evaporate with our thoughts about the *considerations that make it desirable*.

But that's not quite right. I suggest, following John McDowell, that for minimally virtuous people, considerations that favor *impermissible* acts of punitive hostility toward others are *silenced*.[279] As Jeffrey Seidman explains the view, silencing has "two sides."[280] The first is "motivational": the virtuous agent will not desire the impermissible. For example, "*A* will not be tempted by the fact that it would be pleasant to sleep with someone with whom he ought not to, for instance – even if the prospective pleasure would strongly incline him to sleep with that person if there were nothing wrong with doing so."[281] The second side is "rational": the virtuous agent sees "the relevant [moral] reasons for acting, on occasions when they co-exist with considerations that on their own would be reasons for acting otherwise, as, not overriding, but silencing those other considerations – as bringing it about that, in the circumstances, they are not reasons at all."[282]

It follows that weak desert is indeed an extremely important reason for punishment. But it is not a *favoring* reason; it is an *enabling* reason.[283] This, I suggest, explains why blame's desire does not generally arise without the thought about weak desert, and why it generally evaporates when the thought evaporates. And to the extent that blame's desire *does* sometimes arise without

---

[279] *See* JOHN MCDOWELL, *The Role of* Eudaimonia *in Aristotle's Ethics* (1980), *reprinted in* MIND, VALUE, AND REALITY 3 (1998); *see also* Jonathan Dancy, *Reasons, Relevance and Salience: A Response to Hookway*, 130 PHIL. STUDIES 71 (2006) (discussing the related notion of enabling reasons); JOSEPH RAZ, PRACTICAL REASON AND NORMS ch. 1 (1999) (discussing the related notion of excluding reasons).
[280] Jeffrey Seidman, *Two Sides of 'Silencing,'* 55 PHIL. Q. 68 (2005).
[281] *Id.* at 68-69.
[282] *Id.* at 69 (quoting MCDOWELL, *supra* note 279, at 17).
[283] *See* Dancy, *supra* note 279.

its constitutive thought, and it *does* sometimes linger in part after the thought evaporates, the fact that we are not perfectly virtuous explains this.

Perhaps the most contentious part of the view I'm endorsing is that it suggests that normal blamers are sufficiently virtuous to experience "silencing." But I do not think it takes great virtue to experience this kind of silencing *to at least some degree*. That merely takes the absence of psychopathy and the normal inculcation of a moral conscience. Indeed, it is among the first moral lessons we teach our children, and in which we train them over many years: *don't inflict suffering on others*. That lesson comes with a caveat in which we also train them: *unless* they deserve it and you have standing to inflict it. We almost never teach them this caveat explicitly; we teach it to them with our actions, when we *punish* them and others, which is among the most common occurrences in a child's life. To the extent that normal adults have ingrained that lesson, not as a dictum with which to check themselves, but as a character trait, they are virtuous enough to experience the kind of silencing at issue in blame.

To tie my proposal together, recall my objection to Wallace: weak desert is not a reason in favor, and mere permissibility does not make something desirable. That objection is correct. However, weak desert does serve to unsilence or enable the reasons that favor punishment. And because some of those reasons are also central to explaining why hostility is weakly deserved, the thought that hostility is weakly deserved will present the same considerations as reasons hostility is deserved *and* reasons hostility is desirable. To the extent that a thought about desert sometimes seems to be a reason that favors but does not obligate hostility, this is the explanation.

This view is open to the possibility that one might think hostility is weakly deserved without adverting to considerations that play the dual role I described. In a simple version of this case, one has been trained to understand *when* weak desert obtains, but not why, so she thinks

hostility is weakly deserved without any considerations that favor it coming to mind. Unless this person has been trained to have a Pavlovian reaction to her weak desert thought, a desire for hostility will not arise, and this will be rational for her: she sees no reason to punish in the case at hand, even if punishment is permissible.

I suspect versions of this possibility obtain often. Sometimes, when we judge a person deserving of blame (from someone), blame does not arise because we are aware we lack standing or that something else undermines the appropriateness of our blame. But other times, blame does not arise because we have reached the judgment that the person is blameworthy without adverting to considerations that make punishment desirable. Very often, perhaps always, we could find such considerations if we looked for them. But we don't, and in a great many cases, we are beyond reproach for finding better uses of our time; to always seek out such considerations seems like a feature of vindictiveness.

IV. The Practice of Punishing

To review, blame essentially thinks that its target deserves punishment from the blamer or someone else who has standing. Blame is appropriate just when that thought is true. The thought is true when an unrejectable contractualist principle permits punishment from those with standing, including the blamer herself. This means, ultimately, that blame's appropriateness turns on a kind of *moral* principles. But can't blame's appropriateness sometimes *conflict* with morality?

For example, take a murderer who will respond to your punishing him by murdering an innocent third party. Plausibly, you owe it to that innocent, in the contractualist sense, not to punish the murderer. And if that's so, it seems to entail that there is no unrejectable principle that

permits punishing the murderer. Yet CB explains desert in terms of unrejectable principles. It would thus seem to follow that the murderer does not deserve punishment, and that blaming him is inappropriate, *because he would murder an innocent if you punished him*. That cannot be, and if CB entails it, the view fails.

Or take the inverse case: a demon will destroy the world unless you punish your innocent mother for the fact it is her birthday. She obviously does not deserve punishment for this, so CB would seem to entail that no unrejectable principle permits punishing her. Yet her life is on the line. Indeed, the world itself is at stake. Surely, she can reasonably reject a principle that does not permit punishing her *to save her life and the world*.

These cases raise a serious concern, and answering it, I hope, will serve to finally bring CB into clear focus. To begin, contractualism, as a view of all-things-considered morality, purports to determine rules for the general regulation of human conduct. In truth, I doubt it succeeds; I doubt that any system of *rules* can determine what we ought to do *all things considered*.[284] In my view, however, contractualism is a plausible theory of *role morality*: of some-things-considered moral rules for the general regulation of particular practices.

I aim to explain this by analogy. Take legal ethics—the role-morality governing lawyers *qua lawyers*. On one view, the binding rules of legal ethics—the ones that lawyers are obligated to follow in virtue of their role in the legal practice—are identical to the positive rules

[284] I implicitly refer to the rule-fetishism objection to contractualism—*see* Hanoch Sheinman, *Act and Principle Contractualism*, 23 UTILITAS 288, 302 (2011) (arguing that "Principle Contractualism" falls to the "principle-worship objection")—especially when understood alongside the so-called ideal worlds objection contractualism and rule consequentialism. *See* Abelard Podgorski, *Wouldn't It Be Nice? Moral Rules and Distant Worlds*, 52 NOÛS 279 (2018) (articulating the ideal worlds objection in its strongest form); Gideon Rosen, *Might Kantian Contractualism be the Supreme Principle of Morality?*, 22 RATIO 78 (2009); *see also* Caleb Perl, *Solving the Ideal Worlds Problem*, 132 ETHICS 89 (2021); Jussi Suikkanen, *Contractualism and the Conditional Fallacy*, *in* 4 OXFORD STUDIES IN NORMATIVE ETHICS 113 (2014); Nicholas Southwood, *Contractualism for Us as We Are*, 99 PHILOSOPHY AND PHENOMENOLOGICAL RESEARCH 529 (2018).

promulgated by the model rules of professional conduct and legislated at the state level.[285] But there is an alternative, contractualist answer: the rules governing the legal profession are those that none of the participants in the practice—clients, lawyers, and courts—can reasonably reject to play the role of rules that govern the practice.

It is possible, on a contractualist view of legal ethics, that the unrejectable rules diverge drastically from the positive rules. This would be so if the positive rules were profoundly abhorrent. But this is surely not so for us. The unrejectable rules of our system likely include most or all of the positive rules. This follows, *inter alia*, from the need for uniformity and predictability about what rules lawyers are going to follow, the great reputational harms of lawyers acting in violation of the letter of the written rules, and in our circumstances, the fact the positive rules do a fine job of addressing and balancing divergent interests.

And yet the unrejectable rules surely vary in some ways from the positive rules. After all, there are myriad gaps in the positive rules: situations to which the positive rules do not speak clearly enough to offer determinate guidance. In these cases, the unrejectable rules outstrip the positive rules, filling the gaps left by the latter.

Some positive rules, moreover, might be rejectable. Consider, the positive rules sometimes state obligations that most lawyers do not follow, such as an obligation never to assert in court something the lawyer does not know or believe to be true "on the basis of reasonably diligent inquiry."[286] Defense attorneys at arraignments constantly violate the strict letter of this rule by regurgitating claims just made to them by their client, who they met minutes ago and whose claims they have little reason to trust. ("Your honor, he has 3 kids that he takes to school

[285] *See* MODEL RULES OF PROFESSIONAL CONDUCT (AM. BAR ASS'N 2020) (first published in 1983).
[286] *Id.* cmt. 3.3.

every day.") Contractualism is the view that defense attorneys are not violating the *true* rules of legal ethics if the unrejectable rules allow for this convention.

You might reasonably ask: why have rules to govern lawyers *qua lawyers* at all? Why not, instead, require that lawyers act morally in the broad sense that applies to everyone? I take it the answer, at least in part, is that morality in the broad sense, which takes *all* things into consideration insofar as it is overriding, is too contentious, too complex, and too unclear in its application to many situations lawyers regularly face. Different moral views would result in unpredictable behavior and a drastic lack of uniformity, and lawyers would struggle to determine their obligations with speed and confidence. The practice of lawyering is too fraught—balancing zealous advocacy with integrity and honesty before the courts—to do without more streamlined rules.

This is why the unrejectable rules of legal ethics, like the positive rules, simplify the moral domain by considering the main concerns of a small group of generic participants in the practice—lawyers, clients, and courts. It is also why the unrejectable rules of the practice, on my view, are not overriding. If an evil demon will destroy the world unless you violate one of the unrejectable rules, there is no reason to think that the *truly* unrejectable rules *permit* you to do this. The rules are for the general regulation of the practice, not for every fantastical possibility that might ever obtain. Saving the world thus requires breaking an unrejectable rule. And your all-things-considered moral obligation is to do just that.

CB suggests that the practice of punishment is relevantly analogous to the practice of lawyering. This is to say, punishing one another for wrongs is a practice, and there are rules for

that practice just as there are rules for the practice of law.[287] At the interpersonal level, the positive rules of the practice of punishment are not written down, but they are embodied in conventions with which we are all more or less familiar (e.g., no punishing the innocent; blameless descriptive ignorance can exculpate; etc.). On one view—perhaps suggested by some passages in Strawson—the positive rules of the practice of punishment are the binding rules.[288] But again, there is a contractualist alternative: the binding rules of the practice of punishment are those that none of its generic participants—punishers and targets—can reasonably reject to play the role of the rules that govern the practice, i.e., that determine what's prohibited, permissible, and obligatory for those engaged in it.

Why is punishment a practice with its own binding rules? To start, punishment is arguably the weightiest and most fraught practice that humans *all* engage in—as targets if not as punishers. The practice involves intentionally hurting each other, and acting *justified* about it, too. It involves presenting oneself as *superior*, a member of the class that has standing to punish. Being on the receiving end of punishment is burdensome—it's *designed* to be burdensome—so targets have every reason to demand to understand the rules ahead of time.[289] This is so that they can avoid being targets, and so that when they're made targets, they can object by employing the rules in argument. The rules need to be streamlined to serve these purposes. If every instance of

[287] On the idea of blame and punishment as *practices*, I am indebted in part to Strawson, *supra* note 229, but more so to Philip Pettit, whose work helped me to see the deep importance of this idea. *See* PETTIT, THE BIRTH OF ETHICS ch. 6 (Kinch Hoekstra ed., 2018).
[288] For a reading of *Freedom and Resentment* along these lines, *see* Paul Russell, *Strawson's Way of Naturalizing Responsibility*, 102 ETHICS 287 (1992).
[289] This is a key theme of Scanlon's early work on blame and punishment. *See* SCANLON 1998, *supra* note 252, at 249, ch. 6 *passim* ("[T]he force of a person's objection to a principle imposing a burden on her, or permitting others to act in a way that would impose such a burden, can be diminished by the fact that she could avoid the burden by choosing appropriately.").

punishment were deserved just in case it was the thing to do *all things considered*, there would be little of the uniformity and predictability that targets are right to demand.

Moreover, just as legal ethics is torn between ends that are in deep tension with each other—zealous advocacy and candor to courts—so the practice of punishment is torn between its ends. On the one hand, the practice serves innumerable purposes related to moral education and character development, the vindication of victims' value, and the deterrence of wrongdoing. But most generally, it aims to prevent people from hurting each other. And how does it do so? By licensing people to hurt each other. This inescapable conflict would ensure uncertainty, even a degree of chaos, if it were left up to each of us to determine whether each possible instance of punishment is morally permissible *all* things considered.

But my aim, here, is less to justify having a practice of punishment with its own, some-things-considered rules than to point out that we *have* such a practice. And because this is so, we all face a question: what are the binding rules of the practice? As Strawson suggests, we begin our analysis by considering the positive rules. These are sometimes clear, but other times, they are vague and indeterminate because our conventions are unclear, fractured, and inchoate. This gives CB a clear role to play in filling the gaps. And of course, some of the positive rules may be rejectable; CB has a role to play in identifying the rejectable positive rules and their unrejectable alternatives.

Could it be that *all* the positive rules are rejectable? This is not the question whether we should *have* a practice of punishment with its own rules. It's the question whether the unrejectable rules of the practice might be completely different from the positive ones—most likely because there is only one unrejectable rule of the practice of punishment: *never punish*. CB

is consistent with an affirmative answer to this question, just as contractualism about the rules of legal ethics is consistent with the conclusion that every positive rule is rejectable.

But on the version of CB I support, the answer is almost certainly *no*. This is for two reasons. The first is that punishment serves innumerable essential purposes and we are in no position to replace it with a comprehensive alternative that would serve those purposes as effectively with less cost.[290] The second is that Strawson was correct that the norms of actual practice—what I've called the conventions or positive rules of the practice—are rightly important to us.[291] On my view, the practice of punishment needs rules that we *share*. The most shareable rules—the ones most feasible for *all* of us to understand and apply in the same predictable ways *together*—are the ones that everyone has already learned and practiced applying.

In effect, I suggest, the fact that a principle is a positive rule of our punishment practice is generally a pro tanto reason in favor of the principle's unrejectability, and the more of our positive rules that a proposed principle would overhaul, the stronger the objection it must overcome. If you claim that the unrejectable principles conflict with *all* our conventions, you

---

[290] I do not dwell here because the point merits a chapter of its own. But aside from its obviously important role in deterring wrongdoing, two essential roles I take punishment and blame to play are in moral education and vindicating the value of victims—two roles that Jean Hampton stressed in her seminal work. *See* Hampton, *The Moral Education Theory of Punishment*, 13 PHIL. & PUB. AFFAIRS 208 (1984); *A New Theory of Retribution*, *in* LIABILITY AND RESPONSIBILITY: ESSAYS IN LAW AND MORALS 377, 401 (R.G. Frey & Christopher W. Morris eds., 1991) (stressing the role punishment plays in vindicating victims worth); *see also* JEAN HAMPTON & JEFFRIE G. MURPHY, FORGIVENESS AND MERCY (1988). I discuss these issues more in chapter 4, section 2.

[291] *See* Strawson, *supra* note 229, at 133. Strawson understood the importance of our practices in terms of the fact they "really *are* expressions of our moral attitudes and not merely devices we calculatingly employ for regulative purposes. Our practices do not merely exploit our natures, they express them." *Id.* My own conception of the value of our practices is more aligned with their regulatory efficacy, but I share Strawson's sense that the fact we *have* the practice we have is of central importance to *justifying* that practice.

face an immense burden to show that the enormity of our positive rules merits doing without the unity, uniformity, and predictability that treating them as unrejectable brings.

Now return to the demon and the murderer. I claim you should punish your innocent mother and you should not punish the murderer. But those are claims about morality all-things-considered. The rules of the practice of punishment do not, and should not, take into account the idiosyncratic possibilities these cases imagine. So just as doctors and lawyers sometimes ought, all things considered, to violate the rules of their practices, we sometimes ought, all things considered, to violate the rules of the practice of punishment. The demon and the murderer are precisely the kinds of cases where this is so.

But the truth of the claim that *A* deserves punishment from *B* for *X* does not turn on what *B* ought to do all things considered. It turns on whether the unrejectable rules of the practice of punishment permit (or obligate) *B* to punish *A* for doing *X*. Desert, in this context, is tied essentially to the some-things-considered rules of the practice of punishment. In sum, for a person to deserve punishment, in the weak sense, is for the unrejectable rules of the practice to permit those with standing to punish her. For a person to deserve punishment in the strong sense is for the unrejectable rules to obligate it. And to have standing is for the unrejectable rules of the practice to put you in the group that's permitted or obligated to punish the deserving.

Because the some-things-considered rules of the practice of punishment can diverge from the dictates of all-things-considered morality, it can sometimes be all-things-considered obligatory to punish the undeserving or impermissible to punish the deserving. Blame attitudes, however, essentially involve thoughts not about all-things-considered morality, but about *desert*. Hence blame's appropriateness turns not on all-things-considered morality, but on the some-things-considered, unrejectable principles of the practice of punishment.

*V. **Conclusion***

In the next chapter, we turn to the process of determining what rules are unrejectable for the practice of punishment. CB will become clearer still, I hope. But let me conclude this chapter with a brief review, working backward to attitudinal blame.

There is a practice of punishment. This practice has rules particular to it. The unrejectable rules are those that are binding on participants. The rules speak in familiar terms: obligation, permission, and prohibition. To blame a person, at bottom, is both to desire to make a particular move in this practice—to punish the target—and to think that the binding rules of the practice permit or obligate *you* to make that move against *her*. To think the practice permits or obligates *you* to make that move is to think that you have standing; to think the practice permits or obligates those with standing to make that move against *her* is to think that she deserves punishment.

Insofar as we are virtuous, our desire to punish a target only arises when we think the target deserves that punishment. This is to say, it only arises when the principles of punishment we *take* to be unrejectable permit or obligate us to punish her. When the principles we take to be unrejectable do not permit this, the considerations that favor punishing the target are silenced. Thus, blame's constitutive desire will not arise unless we think the target deserves our punishment. And when the blamer comes to think the target does *not* deserve her punishment, the desire will evaporate—the considerations that supported it silenced once again.

Chapter 2 asked what makes a consideration relevant to blame's appropriateness. We now know. A consideration is relevant to blame's appropriateness just in case it is relevant to the deontic output of an unrejectable rule of the practice of punishment.

CHAPTER 4

HYPOCRISY AND SUPERIORITY

Abstract: The kind of punishment that blame desires presents its target as beneath the punisher with respect to the ostensible wrongdoing at issue. This presentation can be innocuous when the punisher's hands are clean. But the less clean the punisher's hands, the more that punishment presents its target as *fundamentally inferior*—i.e., morally inferior to the punisher in an invidious, categorical way. This chapter argues that targets' central objection to hypocritical punishment is that it presents them as fundamentally inferior. The success of that objection explains why hypocritical punishment is often impermissible, and given the apparatus of Contractualism about Blame, why hypocritical blame is often inappropriate. The chapter also begins the task of identifying the least rejectable principle that prohibits punishments susceptible to the central objection.

But as they persisted in their questioning, he straightened himself up and said to them, "Let the one among you who has never sinned throw the first stone at her." Then he stooped down again and continued writing with his finger on the ground. And when they heard what he said, they were convicted by their own consciences and went out, one by one, beginning with the eldest until they had all gone. Jesus was left alone, with the woman still standing where they had put her. So he stood up and said to her, "Where are they all—did no one condemn you?"

—John 8:6-10 (Phillips)[292]

We have all sinned. But we blame and punish others for their sins. In a broad sense, then, we are all hypocrites. John 8 suggests it follows that we all categorically lack the standing to blame and punish. This is a beautiful view, and appealingly simple.[293]

---

[292] This section of John 8, while not considered apocryphal, is part of a pericope that was interpolated into the New Testament around the 3rd and 4th centuries. For discussion, s*ee* Chris Keith, *The Initial Location of the Pericope Adulterae in Fourfold Tradition*, 55 NOVUM TESTAMENTUM 1 (2009); Wieland Willker, *The Pericope de Adultera: Jo 7:53 - 8:11*, *in* A TEXTUAL COMMENTARY ON THE GREEK GOSPELS (Bremen ed., 8th ed. 2011); *Pericope Adulterae*, in THE OXFORD DICTIONARY OF THE CHRISTIAN CHURCH (F.L. Cross & E.A. Livingstone eds., 3d ed. 2005).

[293] For more nuanced readings of John 8, *see* Gail R. O'Day, *John 7:53-8:11: A Study in Misreading*, 111 J. BIBLE LIT. 631 (1992).

But it is also radically revisionary. Consider a 70-year-old Ukrainian who condemns Putin for invading her homeland and killing her nephew when he fought to defend it. Are we to think her standing to blame Putin for these atrocities is undermined by the fact she cheated on her college girlfriend or stole a t-shirt from the mall last year, blameworthy wrongs ('sins') as those are? Given the conventions of our actual practice of blame and punishment, this would be surprising indeed.

This chapter investigates the relationship between hypocrisy and blame's appropriateness using the apparatus of Contractualism about Blame ("CB"). The focus is interpersonal (i.e., non-legal) blame and punishment. On CB, hypocrisy undermines blame's appropriateness just when it undermines the permissibility of punishment according to unrejectable principles of the practice of punishment. The rules governing hypocrisy's connection to *blame* are thus the rules that no one can reasonably reject to govern hypocrisy's connection to *punishment*. Our task, it follows, is to investigate the central objections to hypocritical punishment and to evaluate the relative rejectability of different rules that might play the role of unrejectable principles in the practice of punishment.

Section I reviews claims from chapters 2 and 3 and argues that we have a pro tanto reason to favor principles according to which hypocrisy sometimes, but only sometimes, undermines blame's appropriateness. Section II evaluates the principle encapsulated in John 8 and section III evaluates a principle that denies hypocrisy ever undermines standing. Section IV draws out the lessons from section III to suggest that the core objection to hypocritical punishment is that it presents the target as fundamentally inferior, i.e., morally inferior to the punisher in an invidious, categorical way. Section V considers how this relates to R. Jay Wallace's account of hypocritical blame and responds to an objection to Wallace's and my

accounts. Section VI suggests that blamers have personal reasons to prefer a principle that does not remove their standing too often. And finally, sections VII and VIII evaluate the relative rejectability of competing principles that are responsive, in different ways, to the core objection to hypocritical punishment. These sections argue in favor of a principle that gives *outward appearances* a surprisingly central role in determining whether a person has standing to blame and punish.

I. The Evaporation Heuristic and Pro Tanto Reasons for a Mixed View

Chapter 2 presented the evaporation heuristic as a method for testing whether a given consideration renders blame inappropriate (and not merely bad or odd). It also raised the case BIKER, in which you blame your target for illegally riding his bike down the sidewalk while, it turns out, you are illegally riding your bike down the same sidewalk.[294] I suggested that your blame, in that case, clearly tends to evaporate when you learn of your hypocrisy. The evaporation heuristic therefore supports the conclusion that hypocrisy *at least sometimes* undermines appropriate blame.

But now put yourself in the shoes of the Ukrainian elder, blaming Putin for the atrocities he's inflicted on you, your family, and your country. And say I remind you of that t-shirt you stole last year. I take it that your blame clearly does *not* evaporate upon the reminder of something so minor, given that you are blaming Putin for something so grave. In fact, my raising the t-shirt in this context is reasonably likely to cause you to blame *me*; it is offensive to raise a t-shirt to challenge your standing to blame Putin for atrocities.[295] The evaporation heuristic

[294] *See supra* ch. 3, sec. III.
[295] *See infra* sec. VI for a discussion of why this is offensive.

therefore supports the conclusion that hypocrisy *at least sometimes* does *not* undermine appropriate blame.

But given the view of blame put forward in chapter 3, can we trust the evaporation heuristic? We now know that blame evaporates when the blamer ceases to *think* the target deserves punishment from her. The target deserves punishment from the blamer just when an unrejectable principle of the practice of punishment permits the blamer to punish her. Thus, whether a given consideration causes blame to evaporate, for a given blamer, turns on what that blamer intuitively thinks the unrejectable rules of the punishment practice are. *If the blamer's intuitions are misguided, the evaporation heuristic will produce the wrong results.*

However, the evaporation heuristic remains useful as more than a guide to our intuitions. If I am right that most all of us intuit that blame is inappropriate for the biker but not the Ukrainian, it follows that our practice of punishment involves a clear convention, one that corresponds to a *positive rule* of our practice. The positive rule is precisely what the evaporation heuristic suggests: hypocrisy sometimes, but only sometimes, renders punishment impermissible and blame inappropriate.

Of course, this positive rule need not be unrejectable. But chapter 3 suggested that the fact *p* is a positive rule of our practice is a pro tanto reason in favor of *p*'s unrejectability. Because our system of punishment needs uniformity and predictability about the rules governing our system, there is a good reason to hew to the positive rules, for these are the ones we have already learned, integrated, and practiced, and these are the rules we reasonably *expect* one another to apply. In other words, the positive rules are the ones we already *share*, and that is a count in their favor.

It could still happen that the positive rules are rejectable. But arguments for this conclusion face a special burden. It follows, as we begin our investigation into the unrejectable principles to govern hypocritical punishment and blame, that we have reason to doubt at least two principles because they clearly conflict with the positive rules of our practice. The first is John 8's principle:

> **JOHN 8'S PRINCIPLE**: If *B* has ever committed a culpable wrongdoing, *B* lacks standing to punish (or blame) *A* for *X*.

And the second is a standing-skeptical view attributable to Macalester Bell.[296]

> **BELL'S PRINCIPLE**: Hypocrisy never undermines the standing to punish (or blame).

We should expect, instead, a mixed principle—one that sits between these two extremes—because that is the principle embedded in actual practice.

II. John 8's Principle

Whatever JOHN 8'S PRINCIPLE might entail for God's standing to punish, it entails that none of us have standing. People commit culpable wrongdoings. It's unavoidable. So JOHN 8'S PRINCIPLE is rejectable unless the principle *never punish* is unrejectable. Because that principle is rejectable, I suggest, JOHN 8'S PRINCIPLE is too.

Why is the principle *never punish* rejectable? One hint at the answer is suggested by this risible retelling of John 8, which made news in 2020 for its inclusion in a law-and-ethics textbook written by the Chinese Communist Party:

---

[296] *See* Macalester Bell, *The Standing to Blame: A Critique*, *in* BLAME: ITS NATURE AND NORMS 263, 272-76 (D. Justin Coates & Neal A. Tognazzini eds., 2013).

> *Once upon a time, Jesus spoke to an angry crowd that wanted to kill a guilty woman. "Of all of you, he who can say he has never done anything wrong can come forward and kill her." After they heard this, the crowd stopped. When the crowd retreated, Jesus raised a stone and killed the woman, and said, "I am also a sinner, but if the law can only be executed by a spotless person, then the law will die.*"[297]

This is comically offensive as a *retelling*, but as a *counterpoint*, it captures what is likely the chief concern for JOHN 8'S PRINCIPLE. One of the central functions of punishment is to protect first-order morality from desuetude—from degradation caused by disuse and disregard. Consider the practice of Old Norse religion. There was once a time when vast numbers of people knew the tenets of Old Norse religion, believed in those tenets, cared about them, and followed them. Over time, desuetude set in—people ceased to know, believe in, care about, or follow the tenets. Morality, similarly, is practiced to the extent that people know what it demands, believe it to be reason-giving, care about living up to its demands, and follow them. Punishment and blame play key roles in ensuring that morality is and remains widely practiced—that it does not fall into desuetude and "die"—through the functions of deterrence, moral education, and character development. And of course, this is just one of the important functional roles of that punishment and blame play.[298]

A full argument against JOHN 8'S PRINCIPLE would require a full argument against punishment-skepticism—the view that no one can reasonably reject the principle *never punish*. This is not an argument I make in this chapter or this dissertation. I take it that the thrust of the

---

[297] Translated by Harper's Magazine, from a now-retracted law and ethics textbook written by the CCP. *See The New New Testament*, HARPERS MAG., at https://harpers.org/archive/2020/12/the-new-testament/.
[298] *Consider, e.g.*, Jean Hampton, *A New Theory of Retribution*, *in* LIABILITY AND RESPONSIBILITY: ESSAYS IN LAW AND MORALS 377, 401 (R.G. Frey & Christopher W. Morris eds., 1991) ("I contend that punishment is uniquely suited to the vindication of the victim's relative worth, so that no other method of purporting to achieve vindication could be preferred to it."); Bell, *supra* note 296, at 281 ("If anything is protecting or shielding in this domain it is blame: blame helps to shield and protect us from the moral damage wrought by wrongdoing.").

argument is clear enough, however, and I assume it succeeds. Punishment may be terrible, and blame is among the darkest of human emotions. Yet the functional roles played by punishment and blame are essential, and without an alternative we can trust to play those roles as effectively but with less cost, everyone has sufficient reason to reject JOHN 8'S PRINCIPLE.

III. Bell's Principle

The standing to blame complicates our already-complicated practice of punishment. On BELL'S PRINCIPLE, we simplify that practice by removing hypocrisy from the picture altogether. This principle entails that no hypocrisy, no matter how egregious, can undermine the standing to punish or blame.

BIKER involved a kind of 'extreme' hypocrisy which I'll dub *pure hypocrisy*. In cases of pure hypocrisy, the blamer is currently committing the very same culpable wrongdoing for which she is blaming her target and the blamer does not blame herself. Roughly, a case of hypocrisy is "pure" if it is devoid of differences between the blamer and her target that could plausibly be invoked in an argument that the blamer is in a morally superior position to her target with respect to the wrongdoing at issue.

One implication of this definition is that, in pure hypocrisy cases, the target's *overall moral record* is not at issue.[299] By her overall moral record, I mean, roughly, her history of moral blameworthiness and praiseworthiness as it may relate to the question whether she is a "good" or "bad" person overall. The possibility that one's overall moral record is relevant to the standing to blame is underlined by a case (from Tom Hurka[300]) in which a committed vegetarian blames you

[299] For one important development of this notion, *see* MICHAEL ZIMMERMAN, AN ESSAY ON MORAL RESPONSIBILITY ch. 3 (1988).
[300] Princeton University Seminar (spring 2016).

for eating a roast beef sandwich, but that committed vegetarian happens to be Hitler. Because the *blamer's* overall moral record may be relevant to her standing to blame, it's plausible that the *target's* might be relevant, too. Hence, BIKER would not be a case of pure hypocrisy if the other biker's moral record were at issue.

Now, what is it about pure hypocrisy that makes punishment so objectionable? For example, if you give the biker the stink eye as you ride by him—an act of punitive hostility that aims to punish him—why can he object?

One thought is that in cases of pure hypocrisy, targets will often fail to get the moral message being expressed. Indeed, the other biker is not going to understand that you're blaming him *for bike riding*. Bell argues that hypocritical blame can still achieve many of the functional goods that non-hypocritical blame can achieve, and the same thought seems to hold for many cases of hypocritical punishment.[301] But there is doubtless a point on the spectrum at which we can begin to expect many of these goods—e.g., the communication of moral norms—to lapse.

Even still, the other biker's strongest objection to your blame and punishment is not that it harms him for insufficiently good reason. It is that it expresses a deeply troubling and offensive attitude of fundamental superiority. Here's what I mean.

Punishing, in general, expresses moral superiority; this, in my view, is part of what "the thoughts or messages that adopting such a stance conveys."[302] This might be a contingent fact about our practice of punishment, something that developed as part of the meaning of punishment because those who inflicted it so often adopted an attitude of superiority. But if that's so, it's no wonder why. Punishment involves asserting power over another. It often

[301] Bell, *supra* note 296, at 267-69, 272-76.
[302] I take this language from PHILIP PETTIT, THE BIRTH OF ETHICS 200 n.2 (Kinch Hoekstra ed., 2018), though he does not use it in the context of superiority.

presents the punisher as the moral educator, with the right to inflict harm on the one who has flouted what morality requires. The one punishing *teaches her target a lesson* and demands the target take that lesson to heart.[303] It is no wonder that *standing* would take on a connotation of *superiority* in our system; and once it takes on this connotation, every act of punishment will take on that meaning as well. But punishment expresses moral superiority whether this explanation is apropos or not. That is a background fact from which our contractualist analysis must begin.

Now, an expression of moral superiority can be relatively innocuous. (Terminologically, *moral* superiority is not necessarily objectionable; *fundamental* superiority, which I spell out presently and elaborate in sections IV and V, is always objectionable.) Say you are currently committing culpable wrongdoing *X*, whereas I have never done anything like *X*. If I present myself as superior to you with respect to *X*, you have little basis for offense. I am not degrading you as a person or suggesting that I am better than you in some highly general sense. I am suggesting that in this moment, I am morally superior to you with respect to culpable wrongdoing *X*, on the grounds that you are doing it and my hands are clean.

But now say this is a case of pure hypocrisy. There is no difference between us with respect to *X*. But my punishment still expresses moral superiority over you. And this expression, I suggest, cannot be innocuous: there's no basis for it but the superiority of *me* over *you*. This is to say, my punishment *does* degrade you as a person, and it *does* suggest I am better than you in a general sense that transcends anything about *X* in particular. What's more, since this is a case

[303] Jean Hampton, *The Moral Education Theory of Punishment*, 13 PHILOSOPHY & PUBLIC AFFAIRS 208 (1984) ("According to the moral education theory . . . punishment is intended as a way of teaching the wrongdoer that the action she did (or wants to do) is forbidden because it is morally wrong and should not be done for that reason. The theory also regards that lesson as public, and thus as directed to the rest of society.").

of pure hypocrisy, your moral record is not at issue, so the basis for my superiority cannot be that I'm generally spotless but you're generally a devil. So, what's left?

What's left, and what pure hypocrisy expresses, is the blamer's *fundamental superiority* over her target: her being somehow *categorically* superior in virtue of who she is and who her target is. Perhaps it's the target's race, class, gender, gender expression, sexual orientation, disability, attractiveness, or intelligence. Or perhaps it's some non-specific feature of the target's identity: Shelly is morally beneath us because she's Shelly. Whatever it is, the target is morally inferior to the blamer for reasons that have nothing to do with the wrongdoing at issue or the target's blameworthiness in general.

I take it that racism, classism, sexism, and so forth each involve—indeed, are largely *constituted by*—an attitude of fundamental superiority toward the group they diminish.[304] The history of moral development over time has, in substantial part, been a history of identifying and eradicating such attitudes. But such attitudes still linger, and where they do, we tend to find oppression and injustice, often profound. Every target has an enormously powerful objection to actions that present them as fundamentally inferior. And my claim is that purely hypocritical punishment does just that.

This idea of fundamental superiority is closely aligned with the idea that the superior persons deserve, because of who they are, to receive differentially positive treatment from those who would enforce legal or moral obligations through mechanisms like blame or punishment. (They are "above the law" in a way the inferior are not.) For example, the white supremacist thinks he should not be ticketed for jaywalking, but Blacks should be. The white supremacist

[304] R. Jay Wallace makes almost precisely this point in a passage discussed shortly. *See* Wallace, *Hypocrisy, Moral Address, and the Equal Standing of Persons*, 38 PHIL. & PUBLIC AFFAIRS 307, 334 (2010).

who jaywalks thinks he should not receive a hostile look from law-abiding Blacks. But he thinks Blacks may yet deserve a hostile look from law-abiding whites. It follows, in his eyes, that he is permitted to give the Black jaywalker a hostile look for jaywalking, to punish her for her wrongdoing, even while he jaywalks right past her—because the white jaywalker is fundamentally superior.

To be sure, I don't claim that every instance of pure hypocrisy expresses *the blamer's own attitude* of fundamental superiority. In BIKER, the blamer is merely a dunce who's briefly forgotten they are riding their bike. But punishment does express the blamer's thought that she has standing to blame, and the notion of standing to blame connotes moral superiority in our practice. In cases of pure hypocrisy, the only thing that could justify the blamer's moral superiority is her fundamental superiority. So whether the blamer believes in her fundamental superiority or is simply a dunce, her act of punishing has a meaning that transcends her actual intentions and attitudes.

I therefore suggest that targets of purely hypocritical punishment have an overwhelmingly powerful objection to BELL'S PRINCIPLE: it permits people to inflict punishments that present and treat their targets as fundamentally inferior. Actions that present and treat persons as fundamentally inferior are profoundly disrespectful and harmful. BELL'S PRINCIPLE would even permit people to use the practice of punishment intentionally for the very purpose of creating and maintaining a culture in which certain groups are widely seen and treated as fundamentally inferior.[305] But we need not focus on those who would work to create such a

[305] True, first-order morality would bar them from this; but first-order morality is distinct from the unrejectable principles of punishment, and it is those principles, not first-order morality, that ground blame's appropriateness.

culture *intentionally*. Purely hypocritical punishment would help to create and maintain such a culture whether it was intended to do so or not.

IV. A Provisional Hypothesis

Purely hypocritical punishment is a rarified phenomenon. It has dialectical force against BELL'S PRINCIPLE, extreme as that principle is. But does it cast light on why the principles of punishment prohibit any "impure" cases of hypocritical punishment?

Yes. The characteristic objection to hypocritical blame is, who are *you* to blame me?[306] When I imagine myself in the position of a target uttering those words to a hypocritical blamer, I find it natural to imagine myself angry and offended. I also find it clear what I am angry and offended *about*. It's not that you are blaming or punishing me 'without standing,' or that you're inflicting these burdens despite having no answer to my question. It's that I think you do have an answer to my question—or at least, that there is only one answer that could make sense of how you're treating me: *You think you're better than me.* You think you're morally superior to me *as a person*. In other words, you think you're *fundamentally superior.*

This sentiment seems to be at the heart of targets' emotional reaction to hypocritical blame. That suggests that a deep-rooted concern about fundamental superiority reaches well beyond cases of pure hypocrisy. And it makes sense that it would. Many cases of impure hypocrisy differ sufficiently little from the pure cases that they serve to present their targets in the same light: as fundamentally inferior.

Of course, in cases of impure hypocrisy, it is never the case that an attitude of fundamental superiority is the *only* thing that the punisher could sincerely take to justify an

[306] *See* Patrick Todd, *A Unified Account of the Moral Standing to Blame*, 53 NOÛS 347, 348-51 *passim* (2019) (making much of this point).

attitude of superiority. By definition, impure hypocrisy involves some potentially relevant, moral difference between the blamer and her target. Perhaps I am on my bike now, but you got off a few seconds ago. You could sincerely believe that that minor difference is sufficient to make you morally superior in the sense that punishment expresses.

But the problem is that this is unbelievable even if it is true. Blame and punishment in such cases will express fundamental superiority whether the blamer has an attitude of fundamental superiority or not. If purely hypocritical blame and punishment serve to create and perpetuate a world in which their target is seen and treated as fundamentally inferior, this kind of hypocritical blame and punishment will, too. This is to say, in an important sense, it does not matter what the blamer *thinks* gives him the right to act morally superior to his target. What matters is what his actions suggest, communicate, or express to the target and the community at large.

This suggests a provisional hypothesis: targets' *central objection* to hypocritical punishment is that it presents them as fundamentally inferior. Pure hypocrisy is the greatest offender in this regard, so the objection is strongest in that case. Then, as a general rule, the strength of the objection weakens as hypocrisy becomes less and less pure. It becomes more and more believable that the impurities are what you take to justify your stance of moral superiority—that you don't think you're superior to Shelly because she's Shelly or because she's disabled or gay, but that you think you're superior to Shelly *about her littering* because you haven't littered in a decade and Shelly just threw her garbage out the window. And as the risk that your stance will communicate your target's fundamental inferiority becomes weaker, the central objection fades in force.

We should ask: why would the principles governing the practice of punishment turn on concerns about the harms of expressing fundamental superiority? Punishment can have innumerable bad consequences that play no positive role in moral education, deterrence of wrongdoing, etc. Expressing or entrenching the idea that a certain person or group is fundamentally inferior is just one of them.

Here, for example, is another: punishment causes blood feuds and wars. Group *B* punishes group *A* for *X* and, predictably, group *A* appreciates that not at all. So *A* inflicts a worse punishment on *B*. *B* judges *A* deserving of an even worse punishment in turn, and we're off to the races.[307] And here is yet another example: punishment can be used to entrench tyranny. *B* punishes his adversaries for *X* but not his friends and uses punishment as a tool to stigmatize those who oppose him. So why, then, do we have principles that prohibit *hypocritical* punishment out of a concern about expressions of fundamental superiority, but not principles that prohibit punishments likely to spur blood feuds or punishments in service of tyranny?

My reply is that the rules of our practice *are* responsive to the concerns about these very consequences. The rules of the practice do not permit *A* and *B* to inflict harsher and harsher punishments on one another. If *B* punishes *A* for *X*, then either *B* does so in violation of the rules of the practice, or *A* violates those rules when she inflicts punishment back. The rules of the practice prohibit the situation I described. And because the rules of the practice are generally

[307] Almost everyone is familiar with fictional tales that embody this plot. Take, for example, the musical *West Side Story* (Jerome Robbins, Leonard Bernstein & Stephen Sondheim, 1957), which portrays an escalating war between the Jets and the Sharks, rival gangs. The war climaxes after Maria's lover, Tony, is killed by the Shark Chino. Holding a gun and capable of killing Chino as punishment—continuing the pattern set out in the musical up to that point—Maria first yells, "Now I can kill too, because now I have hate!" But then, after a pause, she drops the gun, collapsing and crying out with grief. It is only because Maria chooses *not* to punish her lover's killer that the bloody war ends and the gangs reconcile over their shared grief. This general plotline, and the lesson it's usually meant to convey, is captured in the well-known proverb, *an eye for an eye leaves the whole world blind*.

permitting rather than obligating, *B* is not usually obligated to punish *A* for *X* even where *B* has standing and *A* deserves punishment.

As for tyranny, the situation I described supposes that the tyrants' adversaries and his friends are all blameworthy for *X*. This is unusual, I suspect; tyrants tend to punish those who are not blameworthy to create the perception that they are. And of course the principles of punishment forbid *that*. After all, we can safely assume they forbid punishing the blameless. But in the situation I described above, the tyrant's selective punishments of his adversaries but not his friends seems like precisely an exercise in presenting one group as fundamentally inferior to the other. In that situation, the basis for the fundamental inferiority is not race or class; it is *loyalty to the tyrant*. Those who are loyal to him are the superior; those who are disloyal are beneath them. If I am right to think that the principles of punishment serve to prohibit punishments that present their targets as fundamentally inferior, then the tyrant's actions are impermissible for the same reason as the hypocrite's.

The principles governing punishment must focus on the generic concerns and interests of those affected by the principles. That is why evil demons are not accounted for. The reason that the risks of blood wars and tyrannical entrenchment are different is that these are realistic risks that the general rules of practice must account for. My suggestion is that it *does* account for them. It is thus little surprise that it would account for the risk that punishment would serve to express fundamental superiority, too.

V. Wallace, the Principle of Equal Consideration, and the Private Blame Objection

The account I've offered thus far shares many similarities to R. Jay Wallace's. Wallace identifies a number of objections to which hypocritical blame and punishment are subject but

claims that the central objection is that it "offends inherently against the principle of equal consideration."[308] His account of what this means has numerous threads, and at times, he seems to suggest a view on which this inherent offense is not a matter of inflicting harm, but about failing to live up to the obligations the blamer takes up. For example,

> Blame carries with it a kind of practical commitment to critical self-scrutiny. It is not necessarily objectionable to be subject to resentment or indignation about offenses committed by others when you have yourself committed similar offenses in the past. What is objectionable is to experience these forms of blame without subjecting your own attitudes and behavior to critical assessment.[309]

And,

> By blaming others in ways that mobilize the reactive sentiments, hypocrites are participating in a system of social sanction and constraint that essentially involves the distribution of esteem and disregard. It is the way they do this—and in particular *their failure to subject their own behavior to the critical attention they train on others*—that effectively attaches differential significance to the interests of the persons whom they blame and to themselves.[310]

In these passages, the point seems to be that blamers take up a commitment to self-scrutiny when they blame, and by blaming hypocritically, they fail to meet this commitment. These offenses against the principle of equal consideration would occur whether they cause harm or not.

But at other times, Wallace suggests a view very much like the one I've put forward, in which the harm caused by acts of hypocritical blame and punishment is central.

> I have suggested that the concern for equal consideration and regard is violated when hypocrites attach greater importance to their own interests than to those of the target of hypocritical address. But what exactly is involved in this differential "attachment" of significance to the interests of different parties? At issue, I believe, are not merely theoretical commitments regarding the comparative importance of

[308] Wallace, *supra* note 304, at 335.
[309] *Id.* at 326.
[310] *Id.* at 333 (emphasis added).

> the different parties' claims. It is not obvious that one offends against the principle of equal consideration if one scrupulously treats people as if their interests were of equal significance, even while one rejects the proposition that this is the case with regard to some particular class of individuals. Conversely, someone who assents to the thesis that persons are fundamentally equal while behaving as if the interests of some count for less would intuitively seem to run afoul of the principle of equal consideration. *This suggests that the activities that flout that principle involve forms of consequential treatment, whereby, e.g., benefits or burdens are distributed differentially to people, in ways that operationalize an attachment of unequal significance to their basic interests*.[311]

My view seems to mirror these claims, especially the last sentence. My view also agrees with Wallace's that hypocritical blame and punishment offend against the equality of persons in the same way that certain other behaviors do, too.

> Racists and sexists, for instance, often engage in behavior that flagrantly and selectively denigrates the moral standing of people who belong to certain ethnic or gender groups, openly depriving members of those groups of certain benefits or advantages, or subjecting them to hostile and demeaning treatment that is not displayed toward people who belong to other ethnic or gender categories. A pattern of behavior with this structure would be objectionable in much the way that hypocrisy is.
>
> …. [It is] illuminating to interpret the moral objection to hypocrisy as continuous with the objection to these more flagrant ways of denying the equal standing of others.[312]

And elsewhere, Wallace describes a hypocritical blamer as "act[ing]" as if she deserves protection from social disapprobation, despite the fact she's committed the same wrong in the same way. "As long as it goes uncorrected, this complex stance attaches to my interests greater importance than it ascribes to yours, affording my interests a higher standard of protection and consideration than it affords to yours. This offends against a presumption in favor of the equal standing of persons that I take to be fundamental to moral thought."[313]

---

[311] *Id.* at 332-33 (emphasis added).
[312] *Id.* at 334.
[313] *Id.* at 328.

Between these two sets of threads—one stressing the blamer's failure to meet a self-imposed obligation, and the other stressing the consequences of operationalizing fundamental inequality—it is the first set that has been given uptake in the literature. This is because there is an objection, attributable to Kyle Fritz and Daniel Miller, that is widely thought to be decisive against the second set, namely, that the harms of *overt* blame and *open* opprobrium cannot explain the inappropriateness of hypocritical *attitudinal* blame, which can be kept fully *private*.[314] In effect, hypocritical private blame can itself be inappropriate, and the harmfulness of any given *actions* cannot be what makes the *private attitude* inappropriate.

Patrick Todd adds an additional dimension to the objection. Responding to Wallace's claim about the "complex stance" that hypocritical blame embodies, Todd asks, "how could this explanation be applied to cases in which the hypocritical blame is . . . directed only towards a fictional character?" He goes on,

> Do I owe it to King Joffrey (a chief villain in *Game of Thrones)* to waive my claim of protection from the negative social effects of the kind of blame I now direct towards him, if I have committed infractions of just the same kind? Of course not; no one could possibly owe it to a fictional character to do anything at all.[315]

Where do these objections go wrong? The answer is clear on Contractualism about Blame. The objections wrongly assume that the harms Wallace and I identify must be harms of attitudinal blame if they're to make attitudinal blame inappropriate. But on CB, blame's appropriateness turns on whether the *punishment* blame desires is permissible. To punish or engage in punitive hostility is to *do* something—to take an *action*. Thus, private attitudinal blame

[314] *See* Fritz & Miller, *Hypocrisy and the Standing to Blame*, 97 PAC. PHIL. Q. 118, 124 (2015) (hereafter Fritz & Miller 2015); *see also* Fritz & Miller, *The Unique Badness of Hypocritical Blame*, 6 ERGO 545, 551 (2019) (repeating the objection) (hereafter Fritz & Miller 2019). Todd, *supra* note 306, at 365-66 also endorses this objection.
[315] Todd, *supra* note 306, at 366.

is appropriate only if its public "expressions"—i.e., acts of punitive hostility and punishment—are permissible. And if *B*'s punishing *A* for *X* is impermissible because the punishment would express fundamental superiority, then *B*'s blaming *A* for *X*, even privately, is inappropriate.[316]

Now Todd is right, of course, that you cannot punish a fictional character. But you can certainly *desire* to punish him. Moreover, you can *think* that such punishment would be deserved and permitted by the unrejectable principles of the practice of punishment. Therefore, blame for a fictional character can be appropriate or inappropriate depending on whether, if you were to punish the character in the situation you're imagining, that punishment would be permitted by the unrejectable rules of the practice of punishment. The fact you *cannot* punish the character because he is fictional is perfectly consistent with these points. Todd's point about fictional characters, therefore, has no purchase against this chapter's account of hypocrisy given CB.[317]

In sum, it follows from CB that there is no problem with explaining the inappropriateness of hypocritical blame attitudes in terms of hypocritical "expressions" of blame, i.e., hypocritical punishment. To the extent that Wallace's view, like mine, does just this, it has been wrongly dismissed.

VI. Blamers' Objections

Say that my provisional hypothesis is correct: targets' central objection to hypocritical punishment is that it presents them as fundamentally inferior. One question is how this should be captured in an unrejectable principle. But before turning to that, we should consider whether

[316] This is because blame's constitutive thought—that *A* deserves punishment from *B* or someone *else* with standing—is false when *B* lacks standing to punish. *See supra* ch. 3, sec. 1.4.
[317] I take it that much the same explanation applies to the possibility of blame for the dead, who can no more be punished than a fictional character, and for whom one's *blame*, we can assume, cannot be harmful.

blamers have anything to say about all this. Do they have reason to want the standing to blame, and thus to reject principles that prohibit them from blaming on the basis of hypocrisy too broadly?

We should not conflate this with the question whether we all have reason to want a system of punishment that can achieve its functional ends. We do. And for that reason, we all have an important objection to any principle, JOHN 8'S PRINCIPLE most of all, that would deny standing too widely.

There is a separate question, however, whether the loss of the standing to blame is ever a harm to those *individuals* who lose it. I suggest that it is. Return to the 70-year-old Ukrainian. I suggested that if I were to challenge her standing to blame Putin on the ground that she culpably stole a t-shirt from the mall last year, she'd be reasonably likely to blame *me*. My challenge is offensive. Why?

One line of thought focuses on *power*. To refuse the Ukrainian the right to blame and punish in such a case is effectively to deny her the right to punish anyone for anything. If her theft of a t-shirt last year undermines her right to punish Putin for atrocities, it undermines her right to punish others for more mundane wrongs, and given a year has passed already, it suggests she may be powerless to do so indefinitely. Punishment is a way for persons to stand up for themselves, for others, and for the good. Blame can benefit victims by giving them strength and courage to fight back. And it is a burden to be excluded from blame when others are not so excluded. When the Ukrainian's friends and family members angrily blame Putin together, her loss of standing means she cannot stand together with them and do the same.

But these points seem a bit cerebral when we think about my challenge to the Ukrainian's standing. Yes, punishment and blame can be valuable goods for individuals, particularly individual victims. But I do not think the Ukrainian's objection to my challenge is,

How dare you deprive me of something important?

It's,

How dare you suggest I'm morally on a par with Putin?

Or,

How dare you suggest my wrongs are morally on a par with his?

If that's right, it suggests her objection to my challenge to her standing is very much like targets' central objection to hypocritical blame.

Now, it seems true that denying that someone has standing to blame presents them as morally inferior to those that do. But the Ukrainian is not offended that we've suggested she's morally inferior to her sister or her friends. She's offended that we've presented her as *not* superior to Putin, her target. In effect, our challenge to her standing presents her as *on his level*.

This is enormously insulting, of course. But it's a bit difficult to parse apart two distinct insults it might communicate. The first is an insult of *false accusation*: that she has committed a culpable wrongdoing that is on a par with Putin's atrocities. It is objectionable to paint someone's wrongdoing as vastly more significant than it is, e.g., to suggest that your wrongly raising your voice at your child constituted serious child abuse. But when I challenge the Ukrainian's standing to blame Putin over a stolen t-shirt, the false accusation simply seems unbelievable.

It is *because* such an accusation is so outlandish, I suggest, the central insult of my challenge is that it presents her as fundamentally inferior. Her wrongs are so much less severe

than Putin's that my challenge suggests something *must* bridge the wide gap between them. What could this be? It could only be that *she* is simply that far beneath *him*. To deny her moral superiority in this context is to suggest that a categorical inequality between her and Putin allows a t-shirt to undermine her moral superiority over a man who's committed atrocities against her and her loved ones.

In effect, just as hypocritical blame and punishment can serve to create a world in which the target and those like her are seen and treated as fundamentally inferior, so can charges of hypocrisy that deny the standing to blame or punish when the blamer's wrongs are not on a par with her target's. Targets can object to a principle that permits hypocritical blame in too many cases on the ground that it allows them to be presented as fundamentally inferior. And blamers can object to a principle that prohibits hypocritical blame in too many cases on the ground that it allows *them* to be presented as fundamentally inferior, too.

In sum, the threat of fundamental inferiority cuts both ways.

VII. The Provisional Hypothesis Expanded

Say all this is correct. How could it be captured in a principle? Consider, to start,

> **ACTUAL ATTITUDE**: *B*'s punishment of *A* for *X* is impermissible if *B*'s attitude of superiority is partly grounded in an attitude of fundamental superiority.

Any principle that turns on whether the blamer has an attitude of fundamental superiority jars with one of this chapter's key suggestions: the *harm caused* by acts that express fundamental superiority is at the heart of the core objection to hypocrisy. The kind of *expression* at issue is objective—what it communicates—not whether it's an expression of the blamer's actual attitude.

It could conceivably turn out that the least rejectable principle in this area uses a heuristic for the kind of objective expression we're concerned with. If that were so, then whether the blamer *has* an attitude of fundamental superiority seems as good a heuristic as any for whether her blame will *express* such an attitude. But it seems to me ACTUAL ATTITUDE must fail. This is because, in BIKER, you *don't* have an attitude of fundamental superiority toward the other biker. You simply forgot you're on your bike. ACTUAL ATTITUDE would make your blame appropriate, in this case, so long as you continue to forget yourself. It follows that ACTUAL ATTITUDE is, at minimum, remarkably inconsistent with the positive norms of our practice, which hold that your blame is inappropriate throughout the case, including before you recall that you're on your bike. And I can only report, for my part, that I find it too unbelievable that your blame is initially appropriate in BIKER to take this suggestion seriously at the outset.

Consider, next, a principle from Fritz and Miller, who also take concerns about fundamental equality to be at the heart of this topic. They claim that *B* lacks standing to blame others for violations of type *X* if *B* has a *differential blaming disposition* ("DBD"): a disposition to blame some but not others for *X* without a justifiable reason for this difference.[318] "[L]ike other dispositions," they claim, "DBDs can be quite finely grained, and need not have any lengthy temporal duration."[319] To have such a disposition, on their view, is to reject the fundamental equality of persons with respect to violations of type *X*, and thus to lose the right to blame which rests on that equality.[320]

In the form of a principle to govern punishment, Fritz and Miller's view suggests,

> **DBD PRINCIPLE**: It is impermissible for *B* to punish *A* for *X* if *B* is disposed to punish some but not others for *X* without a justifiable reason for this difference.

[318] Fritz & Miller 2015, *supra* note 314, at 122.
[319] Fritz & Miller 2019, *supra* note 314, at 556.
[320] *See* Fritz & Miller 2015, *supra* note 314, at 125-26.

The first thing to say about this principle is that it is schematic until we fill in the notion of a "justifiable reason." Unfortunately, Fritz and Miller do little in this regard; their view seems to be that a reason is justifiable if it keeps a differential blaming disposition from being such as to reject the fundamental equality of persons.[321]

But even in this schematic form, it seems the principle faces a very serious concern related to followability and overgeneralization. Consider a real possibility: You have a moral blind spot. You are disposed to let *otherworldly beautiful* people off the hook for most wrongs. The fact someone is otherworldly beautiful is certainly not a justifiable reason for this. You therefore have an unjustified differential blaming disposition with respect to most wrongs. The DBD PRINCIPLE thus entails you lack the standing to blame for most wrongdoings.

What's the problem? The first is that you are in no position to know whether you have such blind spots. Because your prejudice in favor of or against *any* given group is sufficient to spoil your standing to blame with respect to a given kind of wrong, and none of us can be confident we have *no* such prejudices, the only ones of us who can be confident in applying the DBD PRINCIPLE are those who know they *do* have such prejudices; they, at least, can be confident that the principle widely prohibits their blame. The rest of us are in the dark.

And the second problem, closely related to the first, is that this principle casts far too wide a net. The fact you have a prejudice in favor of otherworldly beautiful people, of whom there may be none on Earth, is miles far afield from whether your punitive hostility toward

[321] They address this matter in part when they say, "A father need not reject the moral equality of persons when he feels more affinity toward his own child than a stranger's child. And a mother is not unfairly inconsistently blaming when she publicly lashes out at a drunk driver who hits her child, even though she feels only a bit of indignation toward a drunk driver who hits the child of another person she reads about in the news." *Id*. at 563. It is not clear to me whether, on their view, the mother would have an unjustified DBD if she did not feel *even* a bit of indignation toward the other drunk driver.

Shelly for her littering is likely to express fundamental superiority. In conjunction with the fact that the DBD PRINCIPLE conflicts with the positive rules of our practice by casting so wide a net, and thus denying standing on grounds many grounds our practice takes to be irrelevant, we can conclude that the principle is rejectable.

Given this chapter's arguments, a natural suggestion to consider next is,

> **ACTUAL APPEARANCE**: *B*'s punishment of *A* for *X* is impermissible if, were *B* to stand by that punishment, then a fully rational person, aware of everything relevant that *A* and the broader community are aware of, but not aware of the answer to the question whether *B*'s punishment is partly explained by *B's* attitude of fundamental superiority, would answer yes by a preponderance of the evidence.

This is the first principle we've considered that tries to capture the notion of *objective expression*, i.e., what the punishment *actually serves to communicate* to others, and whether it serves to communicate an attitude of fundamental superiority.

One problem for ACTUAL APPEARANCE is that the preponderance of the evidence standard seems arbitrary.[322] The standard is met only if it is *more* than 50% likely that *B*'s punishment is partly explained by her attitude of fundamental superiority. This means that if the rational person's answer is precisely 50%, then the principle does not prohibit *B*'s blame. But the principle might have said, instead of "by a preponderance of the evidence," that the rational person would answer *yes rather than no if forced to choose between the two options*. Such a principle would refuse to allow the 50% case. Alternatively, the principle might have said that the fully rational person would answer that it *is at least 50% likely the answer is yes*, in which case 50% likelihood would entail that *B* lacks standing to blame. Why land on the preponderance of the evidence standard rather than one of these nearby options?

[322] *See infra* ch. 5, sec. IV (arguing that arbitrariness of this sort may be a serious objection to a principle).

In answer, note first that insisting a rational person answer yes or no refuses to speak to a coherent and reasonably likely possibility: she judges the two equally likely. A principle that refuses to speak to a coherent and reasonably likely possibility is objectionable relative to a principle that speaks to it. Note second that if a rational person thought it exactly 50% likely that *B*'s punishment is partly explained by her attitude of fundamental superiority, then well-recognized principles of charity suggest treating the punishment as *not* involving such an attitude. This is analogous to the legal principle of lenity which instructs criminal law judges, when two readings of a statute are equally plausible, to adopt the reading that favors the defendant and avoids penalizing him.[323] In this context, the favorable reading is that *B*'s punishment is not partly explained by an attitude of fundamental superiority, and it is the reading that avoids penalizing *B* by denying him standing. I take these to be intuitive and plausible reasons to prefer the preponderance standard over these nearby alternatives.

Why a preponderance standard rather than something substantially weaker (e.g., 'probable cause') or substantially stronger (e.g., 'beyond a reasonable doubt')? A substantially stronger standard seems problematic because it will fail to prohibit many instances of punishment that do, in fact, communicate fundamental superiority to targets and observers. A substantially weaker standard is a more viable suggestion. But I suggest the preponderance standard is the most reasonable and natural suggestion given the conflicting interests that push in both directions. To the extent a rational person would *not* judge that an act of punishment involves an attitude of fundamental superiority, challenging the punisher's standing on the basis of hypocrisy risks presenting *her* as fundamentally inferior. This standard tries to strike a balance in the most intuitive way—by treating 50% as a turning point. For these reasons, I suggest, the

[323] For a helpful discussion of lenity and its philosophical and legal underpinnings, *see* Lawrence M. Solan, *Law, Language, and Lenity*, 40 WM. & MARY L. REV. 57 (1998).

preponderance of the evidence standard is the *least arbitrary* of the possibilities. Therefore, in an important sense, it is not arbitrary at all.

Unfortunately, this principle faces a second, more damning problem. Say that *B* appears to everyone to be an upstanding, virtuous person. But *B* also happens to be the masked villain who has been committing heinous crimes by night. Since no one is aware of the villain's identity, ACTUAL APPEARANCE entails that *B* has the standing to punish others for the milquetoast versions of the very crimes he regularly commits. For example, he can blame and punish a person for stealing a t-shirt just before he slips on his mask and commits grand larceny.

There is something extremely distasteful about a principle that would allow this. I suspect it is rejectable for three reasons. First, secrets may always come to light, so what seems an innocuous expression of moral superiority now, given what's known, will seem an invidious expression of fundamental superiority if the truth comes out. Second, a principle governing punishment should not reward individuals for keeping their own flaws hidden from others. Maintaining the standing to blame and punish should not be an incentive for *B* to make sure he's never discovered. And third, the standing to blame is a kind of power over others—a power to inflict harm on them when deserved—and our principles of punishment should not give someone like *B* any sort of *power* over others.

This response raises the reasonable concern that this chapter has treated fundamental inferiority, until now, as if it were the *only* objection to hypocritical blame. We can grant that it is an important objection, even the "central" objection, without thinking all others can be ignored or addressed in a single paragraph. I think this is a fair concern; it is doubtless possible my focus has been too narrow. The question, of course, will be what I have ignored, what I have failed to fully analyze, and what difference these make to the rejectability of principles to govern

hypocritical punishment. For now, I can only suggest that whatever the strength of the objections to this chapter's focus, it is valuable to see where it leads.

And I suggest it leads here:

> **ACTUAL ATTITUDE OR APPEARANCE** (or "TRIPLE A"): *B*'s punishment of *A* for *X* is impermissible if, either,
>
> a) *B*'s attitude of superiority is partly grounded in an attitude of fundamental superiority. Or,
>
> b) were *B* to stand by that punishment, then a fully rational person, aware of everything relevant that *A* and the broader community are aware of, but not aware of the answer to the question whether *B*'s punishment is partly explained by *B's* attitude of fundamental superiority, would answer yes by a preponderance of the evidence.

TRIPLE A is the principle this chapter officially endorses. Condition (a) accounts for the case of the masked villain and disempowers anyone with an attitude of fundamental superiority from blaming. This is itself incentive to avoid such attitudes, and it goes some way toward minimizing the chance that one's actions will appear to express them. Condition (b) continues to focus on the actual appearance of the punishment to those at risk of seeing it as a presentation of the target's fundamental inferiority. TRIPLE A thus homes in on what I've suggested is the core issue, without ignoring nearby concerns that the masked villain lays bare.

One of TRIPLE A's notably attractive features, I suggest, is how it explains a case from Todd, the likes of which initially seems to be a counterexample to it:

> Jonas is a Nazi commander working in a death camp[,] . . . secretly opposed to the Nazi regime. He thus does everything within his power to save the lives of as many prisoners as possible, consistent, of course, with maintaining his position as a committed Nazi; Jonas (correctly) reasons that he can do much more good secretly sabotaging the Nazi efforts as a trusted commander than he could by open defiance. Jonas hears rumors of an escape. In order to keep appearances, he must order someone to investigate the fence. Jonas thus orders Thomas to investigate the fence

> and sound the alarm should he see anyone attempting escape. Jonas chose Thomas for this task because he (blamelessly, though incorrectly) thought that, of all the people he might choose, Thomas would be the most likely to have mercy and not sound the alarm should he actually find prisoners escaping, and instead report back that there was nothing to the rumors. Instead, however, Thomas discovers the escaping prisoners, sounds the alarm, and the prisoners are caught and executed.[324]

Todd contends that Jonas has the standing to blame Thomas in this case.[325] On CB, that means he has the standing to punish. And yet consider a fully rational person, aware of everything that Thomas and the broader community are aware of—which does *not* include Jonas's committed opposition to Nazism. Such a person would think it clear that Jonas's punishing Thomas *for following Jonas's own order* must be explained by the fact Jonas thinks Thomas fundamentally inferior to him. TRIPLE A thus seems to *deny* Todd's eminently plausible contention that Jonas has standing to punish.

But now compare two ways that Jonas might punish Thomas. First, in the story Todd goes on to tell, Jonas condemns Thomas for not showing mercy and confronts him hostilely. Thomas says, "But you ordered me to do it!" And Jonas replies, "Yes, I ordered you to do it, but that gives you no excuse; you should have disobeyed my orders, even at great risk to yourself."[326] Now there is something odd about this story, namely, that Jonas is breaking character: he's coming clean to Thomas bout his true values and who he really is. And he's likely to get himself killed by the Nazis he's infiltrated.

But it is no fluke that Todd tells the story this way. As TRIPLE A suggests, it is *because* Jonas comes clean in this dangerous way that he has the standing to punish and blame Thomas for following his order. TRIPLE A permits Jonas to punish Thomas *in a way that blocks the*

[324] Todd, *supra* note 306, at 334.
[325] *Id.* at 335.
[326] *Id.*

suggestion of fundamental superiority, and this is precisely what Jonas does in the story Todd tells. It follows that Jonas can appropriately blame Thomas by desiring to punish him in this way; the thought that such punishment is permitted by the principles that govern punishment *is true.* This is no counterexample to TRIPLE A.

Now consider a second version of the story. Again, Jonas condemns Thomas and confronts him hostilely. Again, Thomas says, "But you ordered me to do it!" This time, however, Jonas does not break character; he does not reveal his true values. So how does he respond to Thomas? It seems to me that he can't: there is simply nothing that Jonas can say that can answer Thomas's objection short of breaking character, laying out his true values, and blocking the suggestion of fundamental superiority. Jonas's punitive hostility, in this situation, will communicate that Thomas is fundamentally inferior to him. And TRIPLE A entails that that is impermissible. Jonas does not have the standing to punish Thomas *like this*. And to the extent his blame desires to punish Thomas *like this*, in a way that hides what makes Jonas morally superior and leaves Thomas and others to conclude that Jonas thinks Thomas is fundamentally inferior, his blame is inappropriate.

I suggest this is not merely a plausible analysis of Todd's case. I suggest it is the *right* analysis. TRIPLE A correctly holds that Jonas has the standing to punish Thomas the way Todd imagines, but not in the alternative way I've described. The principle correctly holds that a blame attitude that desires to punish Thomas in the way Todd imagines is appropriate. A blame attitude that desires to punish Thomas in the alternative way I described is not. The fact TRIPLE A analyzes Todd's case in just this way is a mark in its favor.

VIII. Conclusion

I don't claim to have fully developed or defended TRIPLE A.[327] Moreover, it is only one of the principles we might have landed on by taking seriously the idea that expressions of fundamental superiority are at the heart of the objection to hypocritical punishment. For example, we might have considered:

> **APPEARANCE GIVEN EVERYTHING:** *B*'s punishment of *A* for *X* is impermissible if, were *B* to stand by that punishment, then a fully rational person, aware of *everything* relevant except the answer to the question whether *B*'s punishment is partly explained by *B's* attitude of fundamental superiority, would answer yes by a preponderance of the evidence.

This is an eminently plausible principle. But it takes us far afield from the central concern about the harm caused by punishments that communicate fundamental superiority. A fully rational person *aware of everything relevant* will know many things that are almost always unknown to targets and onlookers. But this suggests an improvement:

> **APPEARANCE GIVEN THE EXTERNAL:** *B*'s punishment of *A* for *X* is impermissible if, were *B* to stand by that punishment, then a fully rational person, aware of everything relevant *external to the blamer's mind*, including the answer to the question whether *B*'s punishment is partly explained by *B's* attitude of fundamental superiority, would answer yes by a preponderance of the evidence.

And I think APPEARANCE GIVEN THE EXTERNAL is a viable alternative to TRIPLE A. My preference remains for the latter because it fixes its attention more directly on the central concern about the harm caused by actual appearances. In other words, I suspect that APPEARANCE GIVEN THE EXTERNAL will diverge too greatly and too often from appearance *in fact*.

Yet I raise APPEARANCE GIVEN THE EXTERNAL not to argue that it is inferior to TRIPLE A, but to highlight the breadth of potentially viable principles and the need for more work in this

[327] More needs to be said, in particular, about the awareness clause—"aware of everything relevant that *A* and the broader community are aware of"—both in terms of defense and elaboration.

area. I also raise the APPEARANCE principles to highlight once more that CB is not tied to any *particular* principle to govern hypocritical blame and punishment. A CB theorist is free to dismiss my analysis of the central objection to hypocritical blame or to dismiss my analysis of the rejectability of principles that try to account for it. This chapter does not purport to be the last word, and whether its arguments should succeed or fail, Contractualism about Blame remains a valuable framework within which to understand the connection between hypocrisy and blame.

CHAPTER 5

THE EPISTEMIC CONDITION ON APPROPRIATE BLAME

Abstract: Does appropriate blame require the blamer to meet an epistemic condition? The evaporation heuristic provides little help on this question, but Contractualism about Blame suggests an affirmative answer. The harder question is what, precisely, the condition amounts to. This chapter uses CB to evaluate several competing principles and rejects the most natural candidate, which would prohibit the punishment blame desires if the blamer does not *know* her target is blameworthy. Instead, the chapter tentatively suggests, the unrejectable principle prohibits the punishment blame desires unless the blamer (1) has sufficient evidence to form an all-out belief that her target is blameworthy and (2) has satisfied all applicable epistemic-procedural requirements in developing and assessing her evidence.

You are sitting in your cubicle when a coworker draws near. Wearing his mask around his chin—a violation of office rules—he sneezes explosively in your direction. You can almost *feel it*. The next day, he calls off work: Covid. A few days later, you begin to cough. And sure enough: Covid. You're furious. You'd done everything in your power to avoid the virus because you live with an immunocompromised person. But now you have the virus, and your housemate's health is in danger. And it's all because of your coworker's recklessness.

Or is it?

You can't be sure you got the virus from him. It's a pandemic; you might have picked it up elsewhere. In at least one sense of the term, you do not *know* your coworker gave you Covid. And you certainly haven't investigated everyone else you'd come into contact with this past week—every other coworker, the cashier at the store, the chef who made Thursday's takeout—to see whether *they* might have given you the virus instead of him. You're right that *if* your coworker gave you Covid through his recklessness, he's blameworthy for it. But even if your

suspicion is correct, is it appropriate for you to blame him given your imperfect epistemic situation? More generally, what epistemic conditions must a blamer meet, if any, for her blame to be appropriate?

Questions of this sort should trouble us. Given the epistemic limitations we mortals tend to face, blamers' epistemic situations are always imperfect. But while much has been said about the epistemic requirements that a target must meet to be blameworthy for his wrongdoing,[328] and while several writers have suggested there is an epistemic condition on appropriate blame,[329] there is almost no work arguing for the existence of such a condition or analyzing it in depth.[330] This chapter begins the task of systematically investigating whether there is an epistemic condition on appropriate blame, and what, precisely the condition amounts to. It uses the apparatus of Contractualism about Blame, according to which blame is appropriate just when the

---

[328] For a taste of this literature, *see* RESPONSIBILITY: THE EPISTEMIC CONDITION (Philip Robichaud & Jan Willem Wieland eds., 2017) (collecting 16 new papers on the topic).

[329] *See, e.g.*, Dana Kay Nelkin, *Rational Belief and Statistical Evidence: Blame, Bias, and the Law*, *in* LOTTERIES, KNOWLEDGE, AND RATIONAL BELIEF: ESSAYS ON THE LOTTERY PARADOX 5, 25 (Igor Douven ed., 2021) ("[B]lame, to be justified, depend[s] on the rational belief that the agent in question is culpable."). Patrick Todd denies that there is an epistemic condition on the *standing* to blame but seems to agree that there is an epistemic condition on *appropriate* blame. *See* Todd, *A Unified Account of the Moral Standing to Blame*, 53 NOÛS 347, 350-51 (2019) ("[I]n order appropriately to blame a given wrongdoer, one must be justified in believing that the given agent is indeed blameworthy."). Gideon Rosen also seems to rely on the proposition that blame has an epistemic appropriateness condition when he claims that "we are never entitled to any significant confidence" that an agent is blameworthy for a given act, and infers, on that basis, that we should "presumably stop blaming" our friends and neighbors and "presumably also take steps to revise or reconceive any social practice whose justification depends on the assumption that people are morally responsible for their bad actions." Rosen, *Skepticism About Moral Responsibility*, 18 PHIL. PERSPECTIVES 295, 309-11 (2004). *See also* Amy L. McKiernan, *Standing Conditions and Blame*, 32 SW. PHIL. REV. 145 (2016) (defending the non-meddling condition on the standing to blame because of the limited "epistemological relationship between the would-be blamer and the wrongdoer").

[330] The major exception is D. Justin Coates, *The Epistemic Norm of Blame*, 19 ETHICAL THEORY & MORAL PRAC. 457 (2016), who argues that expressed blame is governed by norms of assertion, which include an epistemic norm. He recognizes that his argument only appears naturally to apply to expressed blame, *id.* at 464, but attempts to bridge the gap by arguing that the norms of assertion also apply to "judgments or beliefs[, which] are, in many ways, assertions we make to ourself," *id.* at 465. Autobiographically, I do not find this convincing, and my own way of thinking about this issue, as detailed below, differs drastically from Coates's. That said, what I go on to argue suggests the epistemic condition on appropriate blame is not derivative of a norm governing assertion.

principles of the practice of punishment permit the blamer to punish her target. The focus throughout is interpersonal (i.e., non-legal) blame and punishment.

In what follows, I argue that the evaporation heuristic is ill-suited to guide our analyses in this context (section I). However, CB militates clearly in favor of an epistemic condition (section II). The harder question is what the condition amounts to, and the remainder of the chapter begins the task of analyzing competing alternatives. I argue against a condition that requires *knowledge* of blameworthiness (section III) and a condition that requires a particular rational *credence* that a target is blameworthy (section IV). I then suggest we should favor a condition that prohibits blame unless the blamer (1) has sufficient *evidence* to form an all-out belief that her target is blameworthy and (2) has satisfied all applicable epistemic-procedural requirements in developing and assessing her evidence (section V). In the concluding sections (VI and VII), I show that outstanding questions remain and suggest that this chapter's favored condition deserves substantially more debate and inquiry.

I. A Glitch in the Evaporation Heuristic

A natural first question is whether the evaporation heuristic suggests there is an epistemic appropriateness condition, and if so, what it is. The most obvious candidate is a principle that prohibits blame unless the blamer *knows* her target is blameworthy, i.e., deserving of blame and punishment.[331] To apply the heuristic, we need a case. So consider,

> **CONVERTIBLE**: You look out the window to see that Smith is driving what seems to be your stolen red convertible. You judge that he's blameworthy for stealing it, and you blame him. But then the Oracle of Delphi stops by and tells you, "whether or not this matters to your blame's appropriateness, you do not, in fact, *know* Smith

[331] Chapter 2 defined blameworthiness as moral responsibility for a wrong. It then argued that there is more to *deserving blame* than blameworthiness so defined. Partly in light of that lesson, but also for ease of discussion, this chapter treats blameworthiness as identical to deservingness of blame and punishment.

> is blameworthy." (You implore her to tell you more, but she refuses, except to reassure you that she's using "know" in the sense familiar to analytic philosophers.) How does your blame attitude respond to your recognition of the fact that you do not know Smith is blameworthy?

Now, if you're like me, then after getting clear on the situation you're in, your blame is inclined to evaporate. The problem is that there are two competing explanations for why. The first is that we take there to be a knowledge condition on appropriate blame. Blame thinks that punishment from the blamer is permissible according to the unrejectable rules of the punishment practice, and when we learn the blamer does not know her target is blameworthy, we conclude that the thought is false. This is how the evaporation heuristic is supposed to work; it tells us that we take the deontic outputs of the principles of punishment to turn on the blamer's lack of knowledge that her target is blameworthy.

The second explanation begins from the fact blame has a cognitive component—thought. Thought is belief-like in that it tends to arise with evidence of its truth and evaporate in the absence of such evidence.[332] When we learn that we lack knowledge that our target is blameworthy, we learn that we cannot trust our evidence. And once our evidence is called into question like this, our thought that the target is blameworthy evaporates. But this is not because we've learned the thought is false. It's because we've learned we lack evidence sufficient to justify thinking it.

Each of these possibilities can explain why blame evaporates in CONVERTIBLE. If the first obtains, it casts light on what we take the principles of punishment to be. If the second obtains, it casts no light on those principles. Instead, it shows us that the cognitive component of blame is

[332] For a review of 'thought,' *see supra* ch. 3 § 1.2.

sensitive to considerations that call our evidence into question, and that blame evaporates alongside its cognitive component.

Can we discern which of these explains why blame evaporates in CONVERTIBLE? Speaking for myself, I cannot. More generally, I doubt that anyone's introspection can elicit a justifiably confident verdict one way or another. Because of this "glitch" in the evaporation heuristic, I suggest it is not a reliable starting point for our analysis.

II. The Argument for an Epistemic Appropriateness Condition

Without the evaporation heuristic to guide us, we must turn immediately to CB to determine whether there is an epistemic condition on appropriate blame.

The argument favoring an epistemic condition starts with the fact that punishment is burdensome for targets. There is much that is good about punishment, of course; but these goods rarely obtain when a target is blameless. Moreover, blameless targets have an especially strong objection to facing punishment: not only is it a burden to them, but they lacked reasonable opportunity to avoid punishment by remaining blameless. We can safely assume the principles of punishment prohibit punishing the blameless, after all, so the blameless cannot be charged with putting themselves in the position of permissible targets of punishment. And yet they are punished nonetheless.

Any set of principles that sometimes permit punishment creates a predictable risk of punishers targeting the blameless on the false thought that they are blameworthy. The fact the principles prohibit punishing the blameless is an insufficient response given the predictability of such incidents. Any set of principles that sometimes permit punishment is therefore objectionable to the extent it fails to incorporate reasonable mechanisms to prevent or minimize

the risk of mistaken punishments for the blameless.[333] An epistemic condition on permissible punishment is precisely such a mechanism. Unless all candidate epistemic conditions face powerful objections that outweigh the arguments in favor of such a mechanism, no one can reasonably reject *some* epistemic condition on permissible punishment and (*ipso facto*) on appropriate blame. The question is simply which candidate condition is the least rejectable.

The main counterarguments are twofold. The first is that any epistemic condition will sometimes bar blame and punishment in cases where the target deserves punishment (and the punisher meets all other conditions on permissible punishment). To the extent that blame and punishment for the blameworthy have beneficial effects (e.g., moral education, deterrence of further wrongs, etc.), an epistemic condition is objectionable because it lessens these effects on net. But this argument is impersonal and aggregative, and thus discounted on CB. After all, punishing innocent Ian substantially burdens *Ian*; failing to punish blameworthy Bob does not substantially hurt *Bob* or any other individual. So while this argument should not be ignored, it is weak compared to the considerations raised above.

The second argument focuses on the burdens an epistemic condition places on individual victims. This argument rests on three related worries: (1) victims sometimes benefit from blame

[333] An objection: If you have accepted and internalized a rule that prohibits punishing the innocent, doesn't conscientious and scrupulous application of the rule already require you not to punish a person unless you *know*, or are reasonably justified in believing, that they are guilty? The reply: Perhaps a kind of conscientious and scrupulous application of the rule would *involve* this, but if so, without an independent epistemic requirement, conscientious and scrupulous application of the rule is not required to blame appropriately and punish permissibly. Blame, moreover, involves the *thought* that the principles that permit punishment permit you to punish your target. The epistemic requirements on rational thought are, arguably, low. *See* Ben Holguin, *Thinking, Guessing, and Believing*, 22 PHILOSOPHERS' IMPRINT 1 (arguing that the rational requirements on thought are, roughly, the same as those of guessing). Without an epistemic requirement, blamers would not even need to meet the requirements on rational thought to blame appropriately. But even if we assume that they must meet those requirements "by their own lights," in some sense, the requirements are too low to sufficiently mitigate the risk of blame and punishment of innocents.

and punishment, and will sometimes be barred from these benefits by the epistemic condition; (2) victims may sometimes be unsure whether they satisfy the epistemic condition, causing ambivalence and uncertainty at a time when they'd benefit from adopting a stable emotional outlook; and (3) victims who do not meet the condition, but strongly suspect a potential target is blameworthy, will face one of two burdens—(i) accepting that they cannot blame or punish the one they strongly suspect is blameworthy *now*, despite knowing this might change at any point when more information comes to light, or (ii) working to acquire additional information, which may be difficult and costly.

We should not underestimate the strength of this second argument. To elaborate, an epistemic condition on appropriate blame *will* put many victims in a worse position at a time when they're already suffering. Punishment serves to expressively defend and vindicate victims' value,[334] and blame, as a kind of anger, can provide victims with much-needed strength and help them break attachments to the friends, lovers, and family members who've wronged them. An epistemic condition will block some victims from the benefits of blame and punishment, and moreover, it will block others from settling on a stable attitude in response to their victimization. Some will be unsure whether they meet the epistemic condition and find themselves vacillating between blaming and not blaming. Others, who recognize that they don't meet the epistemic condition, will only know that their blame is inappropriate *right now*. It might become appropriate at any point in the future if additional information comes to light, so those who fail to meet the condition cannot confidently put the question *whether to blame* to rest. Victims who desire to put that question to rest, or desire to blame or punish a target if he is blameworthy, will

[334] *See* Jean Hampton, *A New Theory of Retribution*, *in* LIABILITY AND RESPONSIBILITY: ESSAYS IN LAW AND MORALS 377 (R.G. Frey & Christopher W. Morris eds., 1991).

have to investigate further, and this will often require them to confront the one who wronged them—often a substantial burden.

That said, contractualism commits us to adjudicating the strength of competing objections, even when they are powerful on both sides. And it commits us to doing so without being able to explain, in more fundamental terms, why one set of objections—one set of *reasons*—outweighs the other. In some cases, this is very difficult. But this case does not seem to be one of them. We ought not to ignore victims' arguments against an epistemic condition. But innocents' arguments in favor of an epistemic condition are stronger.

The question, it follows, is which candidate condition is the least rejectable. And while any number of considerations might affect our answer, this section underlines two to keep in mind: an epistemic condition should effectively minimize punishment and blame for the blameless, but it should do so without exacerbating the burdens it will place on victims, for they are the ones who must shoulder its heaviest burdens.

III. Followability and the Knowledge Principle

The most obvious candidate for an epistemic condition is,

> **KNOWLEDGE**: *B*'s punishing *A* for *X* is impermissible unless *B* knows that *A* is blameworthy for *X*.[335]

To evaluate this principle, we need to understand how it operates to achieve its aim: protecting those who are not blameworthy from facing blame and punishment. To the extent the

[335] There is a good question, which I do not address, of whether an epistemic condition should focus not on the blamer's epistemic relationship to the proposition that the target is blameworthy or deserving of punishment, but instead on her epistemic relationship to the proposition that the principles of punishment permit *the blamer* to punish in the case at hand. This is an issue to consider in future work.

principle also protects those who *are* blameworthy, this is a bad side effect. The most natural way for a principle to protect a class from facing burdens is by changing the deontic status of the acts that burden the class, i.e., by making it the case that it is impermissible to distribute the burdens in question to that class. We might thus expect that if KNOWLEDGE achieves its aim, it does so by making impermissible some instances of otherwise-permissible punishment of those who are not blameworthy.

But that is not how KNOWLEDGE operates. The class the principle aims to protect, the blameless, are *already* impermissible targets of punishment and inappropriate targets of blame. KNOWLEDGE might serve as an independent explanation of why it is impermissible to punish the blameless. But it never makes a difference to the fact that it is impermissible to punish them or inappropriate to blame them. Instead, there is only one situation in which KNOWLEDGE makes a difference to the deontic status of punishment. When a target is blameworthy, but a would-be blamer does not know it, KNOWLEDGE prohibits punishments that would otherwise be permissible. This is to say, KNOWLEDGE makes it the case that some punishment for the blame*worthy* is impermissible. These are remarkable features of a principle that aims to protect the blameless without protecting the blameworthy.

How, then, does KNOWLEDGE operate to achieve its aim? Consider the relationship between a principle that requires safe driving, on the one hand, and principles that require stopping at red lights, obeying speed limits, etc., on the other.[336] The former principle has a narrower deontic extension than the second set of principles. To the extent their deontic extensions differ, this is because principles like "stop at red lights" apply in cases where it would be perfectly safe to run the light. But principles like "stop at red lights" are nonetheless valuable

[336] For this example, and on the importance of *followability* to contractualist principles generally, *see* Daniel Weber, *Being Realistic About Rules*, unpublished manuscript (Apr. 2022).

because principles like "drive safely" are hard to follow without more specific principles that can guide drivers more readily. If "drive safely" were our only principle to govern the road, there would be more accidents—not because "drive safely" leaves any of the *unsafe* behavior causing these accidents *permissible*, but because it is too hard to follow and apply correctly, especially when surrounded by other drivers who each may have their own thoughts about what the principle requires of them.

In effect, "stop at red lights" protects drivers because it is *followable* where "drive safely" is not. I suggest that KNOWLEDGE protects innocents for the same reason. A principle proscribing punishment for the blameless is hard to follow when you don't know whether the target is blameworthy. KNOWLEDGE is more followable in these circumstances: you cannot answer the question whether the target is blameworthy, but you can answer the question whether you *know* she is. And just as "stop at red lights" will sometimes force drivers to waste time when running the light would be safe, so KNOWLEDGE will sometimes force would-be blamers to withhold punishment for the blameworthy. But just as we should prefer to sometimes waste drivers' time rather than risk accidents, so we should prefer to sometimes proscribe punishment for the blameworthy rather than risk punishment for the blameless.

One upshot of this discussion is that *followability* is a key desideratum for an epistemic appropriateness condition. How followable, then, is KNOWLEDGE? The first thing to say is that the real question is only how followable the principle is *relative* to its (yet-unstated) competitors. The second is that the question highlights how fraught this chapter's topic is due to its connection to outstanding disputes in epistemology. Some accept a version of the so-called KK principle: if you know that *p*, you're at least *in a position* to know that you know that p.[337] Those

[337] The literature on the KK principle is prodigious. For a particularly influential argument against the KK principle, *see* TIMOTHY WILLIAMSON, KNOWLEDGE AND ITS LIMITS ch. 5 (2000). *See also* Simon

who endorse KK will presumably think KNOWLEDGE more followable than those who reject it, for those who endorse KK will think we more often (indeed, always) have access to the consideration on which KNOWLEDGE's application turns.

Whether it's possible for this chapter's analysis to proceed without taking contentious stands on outstanding epistemological debates is an open question. But I will attempt to avoid such stands in what follows. Partly for that reason, the remainder of this chapter is unavoidably speculative and at times cursory, aimed more at *beginning* the task of analyzing the main competitors rather than *completing* it.

How, then, might we determine the rejectability of KNOWLEDGE without determining the truth of KK? Note, first, that there are vastly divergent conceptions of knowledge and its major features, like whether it's closed under entailment. A principle whose application turns on such a contentious concept therefore promises to be systematically misapplied by those who conceive of it incorrectly. It also guarantees disagreements about how to apply the principle during high-stakes and heated interpersonal exchanges—as exchanges motivated by wrongdoings tend to be—where disagreement about a governing principle's applicability is reasonably likely to undermine productive resolutions. These predictable misapplications and disagreements might be avoidable, or at least mitigated, by competitor principles that turn on credence or epistemic justification rather than knowledge *per se*.

Additionally, even if KK is true, it is not obvious that those who do *not* know that *p* are in a position to recognize that they do not know that *p*. When you see Smith driving what appears to be your stolen red convertible and form a justified true belief that he is the thief, you may very

---

Goldstein and Daniel Waxman, *Losing Confidence in Luminosity*, 52 NOÛS 962 (2021) (evaluating Williamson's argument and concluding that KK is false on different grounds); Kevin Dorst, *Abominable KK Failures*, 128 MIND 1227 (2019) (arguing that several major arguments against the KK principle are unconvincing).

well think you know he's the thief, but you'll be mistaken—this is a Gettier case—and it is anything but clear how you would be in a position to see this. Moreover, even if KK is true, and you are always *in a position* to know that you know that *p*, the work of attaining this higher-order awareness may be burdensome and often left undone. (*No one*, as far as I'm aware, accepts the strong version of KK on which knowing that *p* entails knowing that you know that *p.*)

This discussion raises the hope that, whether or not KK is true, and whether or not you *take* it to be true, there may be an alternative we can all accept is superior. And whether or not this pans out, there is some reason to think there *must* be a superior alternative. Say we discover that the causal theory of knowledge is true, along with a version of meta-normative realism on which we have no causal access to normative facts or reasons. And say that, on the true versions of these views, we continue to have *justified* beliefs about normative matters, including the fact that so and so is blameworthy, but we never have *knowledge* of such things.[338]

It seems to me clear that, if this were so, blame would still sometimes be appropriate. This is partly a matter of sheer intuition. But it seems, too, that CB would militate in the same direction, for it would turn out that the reason we categorically lack *knowledge* of blameworthiness has nothing to do with the likelihood that our potential target is blameworthy. If an epistemic condition on blame is meant to minimize the risk of misplaced blame and punishment, then *knowledge*, it seems, is at best a proxy for what matters.

IV. A Credence Principle?

Since what matters is minimizing the risk of blame for innocents, a credence-based principle is a natural candidate. There are many questions about how such a principle should be

[338] I am grateful to Gideon Rosen for raising this idea.

constructed. A particularly salient one is the degree of credence it should specify. But also salient is how we specify the kind of credence at issue. It is presumably not subjective credence, after all. If I have a credence of 1 that you are blameworthy because Donald Trump said so, or because I have a .5 credence that you did wrong and a .5 credence that you are culpable, which I've added together to reach a credence of 1, then any principle that turns on subjective credence will not bar my blame. Yet it presumably should, so the best contenders will concern *rational* credence.

But we have reason to think that the epistemic condition on appropriate blame does not turn on rational credence, either. Consider a case from Lara Buchak:

> You leave the seminar room to get a drink, and you come back to find that your iPhone has been stolen. There were only two people in the room, Jake and Barbara. You have no evidence about who stole the phone, and you don't know either party very well, but you know (let's say) that men are 10 times more likely to steal iPhones than women. I contend that this isn't enough to make you rationally believe that Jake stole the phone. If you accused Jake, he could, it seems to me, rightly point out that you don't have evidence that he in particular stole the phone. He could protest that you only know something about men in general or on average. But you should have a high credence that Jake stole the phone: if you had to place a bet with only monetary gain and loss at stake, it is clear that you should bet on Jake (given the statistics, you can expect to do better in general by betting on the man in these kinds of cases: assuming there are an equal number of men and women in the population, then for every 11 cases of iPhone-stealing, 10 are perpetrated by men). On the other hand, if we modified the case so that you know that men and women are equally likely to steal, but a fairly but not perfectly reliable eyewitness (let's say, 90% reliable) tells you she saw Jake take it, it seems that you can rationally form the belief that Jake took it, even though you have a lower credence in this case. A similar point holds if Jake has a guilty look and if guilty looks indicate strongly but not perfectly that the individual has perpetrated the crime in question.[339]

[339] Buchak, *Belief, Credence, and Norms*, 169 PHIL. STUDIES 285, 292 (2014).

As Buchak notes, "[s]tatistical evidence . . . presumably produces a rational credence,"[340] in this case, greater than .9 that Jake is the thief. And yet it is clear that blaming Jake in the case at hand is inappropriate. This goes despite the fact that blaming Jake could be appropriate in an alternative case where you have non-statistical evidence that supports a lower credence than that which is rational to have in the original case. This is a strong reason to doubt that the epistemic condition on permissible punishment and appropriate blame turns on rational credence.

Those who favor a credence-based principle might suggest a principle that turns on both rational credence *and* the kind of evidence that supports that credence. A principle like this might specify a certain rational credence *and* require that the credence not be based on statistical evidence alone. I think this remains a viable possibility. Someone who took this line might argue that if statistics alone were sufficient for permissible punishment, then persons who are statistically likely to be culpable in circumstances *C* will be punished and blamed time and again whenever circumstances *C* arise—a substantial burden. One might also argue that blame based on purely statistical evidence disrespects persons *as persons* by treating them as, instead, instances of the category *male* or *gay* or *Black*, etc. Because this chapter cannot be comprehensive, however, we cannot give these arguments the discussion they deserve.

Instead, my hunch, developed further in section V, is that Buchak's case shows us it is not rational credence we are after. Rather, it is evidence sufficient for full-blown, rational belief. On this view, Buchak's case is just like the lottery paradox. What these show us is that statistical evidence *and rational credence* are not sufficient for appropriate blame and permissible punishment because they are not sufficient for full-blown, rational belief. The question for this

[340] *Id.*

view is why the unrejectable epistemic condition would turn on evidence sufficient for full-blown belief. In other words, why is a principle that turns on rational credence rejectable?

One possibility is that credence-based principles have followability problems because our subjective credences are generally opaque to us, and in general, it is no less opaque what credences we rationally ought to have. But, at the risk of spoiling what's to come, every epistemic condition on appropriate blame has followability problems, and it's not obvious that these are the most severe.

A second possibility, which I'm inclined to endorse, is that any credence-based principle has a serious *accessibility* problem. Normal human beings need to be able to determine which principles are rejectable and which are unrejectable. To the extent a principle makes this impossible, it is *inaccessible*. Accessibility requires that persons be able to think of the principle in the first place; find it plausible enough to merit serious consideration; and distinguish the principle as superior to nearby alternatives.

The central accessibility problem for any credence-based principle is that it draws an arbitrary line in the sand. Consider a principle that prohibits punishment absent a rational credence of at least .9 that the target is blameworthy. This principle has a *followability* problem because rational credences tend to be opaque to us. But it has a more severe *accessibility* problem because there is no good reason to land on that number. How is a person supposed to discern whether the unrejectable principle involves the number .9 rather than .75 or .8 or .95 or .91 or .87? Appeal to a vague border will not help us here, because the questions will then be where the vagueness begins and where it ends, and we'll face additional followability problems when credences fall within the part of the range where there is no answer to the binary question whether the principle prohibits punishment.

One might object that these arguments don't succeed against a principle according to which *B*'s punishing *A* is permissible only if *B* is justified in being *highly confident* that *A* is blameworthy. But this is a verbal formulation of a vague range. My argument is that there is no clear answer to the questions, why not just *confident*? Why not *extremely confident*? Why not *reasonably confident*? Why not *confident beyond a reasonable doubt*? "Highly confident" remains an arbitrary choice. To the extent this is not apparent, I suspect that is because it is so unclear what these verbal formulations *mean* that they blend together, making it seem as if there are a small number of options when there are in fact an infinity of options that we need numbers, not words, to differentiate.

Whether or not this gets at the central problem for a credence-based principle, I take it that an accessible epistemic condition on appropriate blame should avoid drawing lines in the sand. KNOWLEDGE is appealing, in part, because it does not succumb to such problems. It is a natural principle to land on, in a way that credence-based principles struggle to be.

V. Evidential Principles

Our discussion of KNOWLEDGE pushed us toward a principle that turns more directly on risk of error. But I've argued against credence-based principles, too, partly on the basis of Buchak's case and partly on the ground that they draw arbitrary lines in the sand and are thus inaccessible. We should therefore consider an *evidence*-based principle that specifies the extent of evidence necessary for appropriate blame in an intuitive way.

Consider:

> **EVIDENCE FOR BELIEF**: *B*'s punishing *A* for *X* is impermissible unless *B* has evidence sufficient for epistemically rational belief that *A* is blameworthy for *X*.

The belief specified in this principle is what philosophers sometimes call "full-blown" or "outright" belief. It is a mental state with which we are all intimately familiar as it is often said to be the "mental state expressed by sincere, intentional assertions."[341] It is also the mental state that Buchak and others have suggested is not rational on the basis of statistical evidence alone.[342]

This principle does not invoke the concept of knowledge, but it is nearly equivalent to

> **EVIDENCE FOR KNOWLEDGE**: *B*'s punishing *A* for *X* is impermissible unless *B* meets the epistemic justification requirement for knowledge that *A* is blameworthy for *X*.

These principles are equivalent on the natural assumption that the epistemic justification required for all-out belief is the same as the epistemic justification required for knowledge.[343] And even though EVIDENCE FOR KNOWLEDGE *invokes* knowledge, neither principle requires us to meet the full conditions for knowledge. For example, assuming Gettier cases cannot be explained in terms of insufficiency in epistemic justification, neither principle bars blame in such cases. And if it should turn out that knowledge requires something beyond epistemic justification and true belief, like a causal connection that is lacking in this context because normative realism is true, neither principle will categorically bar blame.

---

[341] Ralph Wedgwood, *Outright Belief*, 66 DIALECTICA 309, 310 n.1 (2012) (quoting PATRICK MAHER, BETTING ON THEORIES 131 (1993)).
[342] *See* Buchak, *supra* note 339, 291-94.
[343] I have two possibilities in mind for how one might reject this assumption. One is that knowledge requires *stronger* epistemic justification than rational belief. The second is that knowledge requires a certain *kind* of epistemic justification that rational belief does not. For example, we can imagine a form of the causal theory of knowledge according to which the kind of *justification* knowledge requires is analyzed in causal terms, but where the kind of justification that rational belief requires is not. If this were so, then assuming meta-normative realism, EVIDENCE FOR KNOWLEDGE will require something that never obtains—a causal connection to the fact the target is blameworthy—but EVIDENCE FOR BELIEF might often still be true.

I take it that EVIDENCE FOR BELIEF and EVIDENCE FOR KNOWLEDGE will be indistinguishable to lay people. There might be an accessibility problem if the two principles appeared to diverge, for then we would face difficult questions about which is unrejectable. But because the principles are equivalent on a plausible, natural assumption, there is little need to adjudicate which is unrejectable. Each also fixes the required epistemic justification to things with which we are quite familiar: belief and knowledge. Far from drawing lines in the sand, these principles point to a line already drawn by belief and knowledge.

Do these principles explain Buchak's point that purely statistical evidence is insufficient for appropriate blame? The answer is certainly yes if we grant that purely statistical evidence is insufficient for rational belief.[344] But those who deny this view will have to point to something specific to blame and beliefs about blameworthiness. The most plausible line, described above, is likely that if purely statistical evidence is sufficient for rational belief that a target is blameworthy, then those who fall into the statistically culpable group will be made targets again and again, with little opportunity to avoid this substantial burden. It follows that the EVIDENCE principles we're considering can explain Buchak's point whatever one's broader views on the relationship between statistical evidence and rational belief.

The main problem for these principles is familiar: followability. It is often unclear whether we have sufficient evidence for belief or knowledge. And yet by now it should be clear: every epistemic condition on appropriate blame will face followability issues. The fact these principles are no exception does not mean they are rejectable. My key contention is that whether or not KK is true, these principles are *more* followable than a principle requiring knowledge because these principles turn on *just one part* of what knowledge requires. Debates and unclarity

[344] Nelkin endorses this view, which she calls the Statistical Thesis: "For any proposition P, it is not rational to believe P solely on the basis of a high statistical probability of P." *See supra* note 329, at 8.

surrounding what *else* knowledge might require, and certain followability problems, like those distinctive to Gettier cases, do not arise on these principles. My hope, and my suggestion, is that these principles are thus *less* rejectable than KNOWLEDGE regardless of your views on that principle's followability when considered in isolation.

These principles face a separate issue, however: they are underspecified. Return to the case of Jake. Assume that Buchak is right that it would be rational, in at least one sense, for you to form an all-out belief that Jake is the thief based on a guilty look. We might nonetheless think: the belief may be epistemically justified *given the evidence you have*. But you have yet to engage in even a cursory investigation that might reveal additional evidence. Until you do so, your belief is not epistemically justified *given the evidence you* should *have*.

In this vein, Gideon Rosen distinguishes a broad and narrow sense in which a belief can be epistemically justified. In the narrow sense, "a belief is justified when the subject's total evidence constitutes a good and sufficient reason for it," where "facts of which the subject is unaware do not count as part of his total evidence and are thus irrelevant to whether he is epistemically justified in believing what he believes."[345] This narrow sense of epistemically justified belief leaves out the "active" side of epistemic life, namely, "*inquiry*, a process in which we ask questions, formulate hypotheses, gather evidence, perform experiments, reflect on our results, double-check our calculations, and so on."[346] A belief is epistemically justified "*in the broad sense*," then, "if (a) the subject has complied with every pertinent procedural requirement in developing and assessing his evidence, and (b) his opinion is justified by that evidence in the narrow sense."[347]

---

[345] Rosen, *Kleinbart the Oblivious and Other Tales of Ignorance and Responsibility*, 105 JOURNAL OF PHILOSOPHY 591, 600 (2012).
[346] *Id.*
[347] *Id.* at 601.

In the case where you conclude that Jake is guilty based solely on his guilty look, your belief may be epistemically justified in the narrow sense. But if you'll accuse or punish Jake on the basis of your belief, the stakes are moderately high. It follows, at least if there are relatively low-cost steps you could take to investigate further, that your belief is not epistemically justified in the broad sense. With this in mind, distinguish two versions of EVIDENCE FOR BELIEF:

> **NARROW EVIDENCE**: *B*'s punishing *A* for *X* is impermissible unless *B* has evidence sufficient, in the narrow sense, for epistemically rational belief that *A* is blameworthy for *X*.

> **BROAD EVIDENCE**: *B*'s punishing *A* for *X* is impermissible unless (i) *B* has evidence sufficient, in the narrow sense, for epistemically rational belief that *A* is blameworthy for *X*, and (ii) *B* has complied with all applicable epistemic-procedural requirements in developing and assessing her evidence.

Our question is, which is the unrejectable principle?

The reason to prefer BROAD EVIDENCE is obvious: it does more to protect the blameless from misplaced blame and punishment. Innocents in Jake's shoes have a powerful objection to NARROW EVIDENCE: it fails to protect them from punishment in cases where would-be blamers would learn of their innocence by taking steps they are already epistemically obligated to take.

The most obvious thing to say against BROAD EVIDENCE is that it burdens victims by placing an additional set of restrictions on their ability to punish permissibly and blame appropriately. Yet it is important not to overstate this point. None of the principles we're considering is "narrow scope,"[348] so BROAD EVIDENCE does not obligate *B* to comply with the applicable epistemic-procedural requirements in developing and assessing their evidence for the proposition that *A* is blameworthy for *X*. It requires *B* to *either* comply with these requirements

[348] On the distinction between wide-scope and narrow-scope principles, *see* John Broome, *Wide or Narrow Scope?*, 116 MIND 359, 359-61 (2007).

or not to punish *A*. BROAD EVIDENCE does not force *B* to investigate in cases where the investigation would be too costly or burdensome (although it does force *B* either to investigate or forego blame, which may also be costly).

For similar reasons, BROAD EVIDENCE might also restore a bit of autonomy to victims. I suggested above that an epistemic condition on appropriate blame will generally place victims in a position where their blame might become appropriate at any moment when more evidence comes to light, so they will not be able to settle the question *whether to blame*. This goes somewhat less on BROAD EVIDENCE, where choosing not to investigate will more often allow the would-be blamer to confidently move forward knowing that her blame is not *and will not become* appropriate; after all, she's chosen not to inquire further.

That said, it is hard to evaluate this consideration. As Rosen notes, epistemic-procedural requirements "vary massively from case to case, and it is hard to say anything very general about their content."[349] Presumably, moreover, the content of the epistemic-procedural requirements can vary along with the evidence one *has*. For example, blaming Jake on the basis of a guilty look may be inappropriate until you investigate further. So if you decide not to investigate, then *holding fixed your evidence*, you can be confident your blame will remain inappropriate. But if Jake suddenly confesses to the crime and pulls the stolen iPhone out of his pocket, it's doubtful that your blame remains inappropriate. True, you did not meet the procedural requirements that applied when your only evidence was Jake's guilty look. But those requirements almost certainly ceased to hold when you were suddenly faced with this additional evidence.

The fact that epistemic-procedural requirements can vary so significantly points to the deepest problem for BROAD EVIDENCE, namely, that it may be extremely unclear, in a given case,

[349] Rosen, *supra* note 345, at 601.

what the epistemic-procedural requirements demand. This makes BROAD EVIDENCE substantially harder to follow than NARROW EVIDENCE: victims will sometimes be unsure whether they've met the procedural requirements because they will be unsure what they require.

A counterpoint is that blamers are well versed in applying these procedural requirements *to their targets*. It often happens that targets are ignorant of morally salient, descriptive features of their actions, and when this is so, the question is whether they are blameworthy for (or at least in spite of) their ignorance. A central way that targets can be blameworthy for their ignorance is by (culpably) failing to meet the epistemic-procedural requirements applicable to their belief formation. Insofar as blamers are well versed in judging *others* for meeting or failing to meet these procedural requirements, they can complain only so much that the requirements are too difficult to discern and apply to themselves.

What's more, we are familiar with a concept of *negligence* on which the applicable procedural-epistemic requirements in a given case are those that a would-be blamer must meet to avoid the charge of negligence in belief formation. The law regularly charges juries to determine whether certain harms were negligently caused, for example, where negligence is understood in terms of what a *reasonable* person would and would not do, what they would believe and not believe given their evidence, and whether they would or would not investigate or analyze the situation further.[350] The law refuses to define *reasonable* in more fundamental terms because it

[350] *See* Stephen G. Gilles, 80 VA. L. REV 1015, 1017 n.6 (1994) (collecting examples of jury instructions that tell juries to determine negligence based on whether the defendant exercised the degree of care that a reasonable person would have exercised under the same circumstances); Ashley M. Votruba, *Will the Real Reasonable Person Please Stand Up?*, 45 ARIZONA STATE L.J. 703, 705 (2013) ("In most cases the [jury] instructions are only passively heard once and do no more than tell the jury to evaluate what the ordinary reasonable person would do given the circumstances. The jury generally hears little more than the term 'reasonable person,' and rarely is the term actually defined . . . . Instead they are left largely to their own devices to decide what is considered negligent behavior in this circumstance with only the vague, undefined concept of the reasonable person as their guide."); Patrick J. Kelley & Laurel A. Wendt, *What Judges Tell Juries About Negligence: A Review of Pattern Jury Instructions*, 77 CHI.-KENT L. REV.

takes it that normal human beings understand the concept; after all, they use it in everyday life as a tool to evaluate their own actions and the actions of others, and certain concepts do not admit of further definition.[351]

With this in mind, I suggest BROAD EVIDENCE is well understood as equivalent to two other formulations:

> **NON-NEGLIGENCE**: *B*'s punishing *A* for *X* is impermissible unless (i) *B* has evidence sufficient, in the narrow sense, for epistemically rational belief that *A* is blameworthy for *X*, and (ii) *B* has *not been negligent* in developing and assessing her evidence.

> **REASONABILITY**: *B*'s punishing *A* for *X* is impermissible unless (i) *B* has evidence sufficient, in the narrow sense, for epistemically rational belief that *A* is blameworthy for *X*, and (ii) *B* has developed and assessed her evidence in the way a *reasonable person* would.

587, 587-89 (2002) ("[T]he negligence standard requires the jury to determine two things: what the defendant did, and whether an ordinary reasonable person would have done what the defendant did under the circumstances. The standard itself, then, does not give the jury rule-like guidance.").

[351] This idea is captured in Baron Alderson's famous statement, in Blyth v. Birmingham Waterworks Co., 156 ENG. REP. 1047, 1049 (1856), "Negligence is the omission to do something which a reasonable man, guided upon those considerations which ordinarily regulate the conduct of human affairs, would do." *See* 1 Special Comm. On Uniform Court Instructions, Iowa Uniform Jury Instructions No. 2.1 (1970) (Negligence is "the failure to do something which a reasonably prudent person, *guided by those considerations which ordinarily regulate the conduct of human affairs*, under the circumstances would do." (emphasis added)); *see also* Catharine Pierce Wells, *Tort Law as Corrective Justice: A Pragmatic Justification for Jury Adjudication*, 88 MICH. L. REV. 2348, 2385 (1990) (The jury "has no choice but to apply its own rough intuitive sense of responsibility to a situation whose elements are so common as to be within the ready imagination of each individual juror."); Walter W. Steele Jr. & Elizabeth G. Thornburg, *Jury Instructions: A Persistent Failure to Communicate*, 67 N. CAROLINA L. REV. 77, 100 (1988) (describing "reasonable person" as the first example of an "inherently general" legal concept, which the law does not define in order to maintain its flexibility and its ability to evolve with changing times and changing community standards; "[b]ecause of these problems," the authors say, "we would not propose a redefinition of the concept[] to achieve greater clarity."); Benjamin C. Zipursky, *Reasonableness In and Out of Negligence Law*, 163 U. PENN. L. REV. 2131, 2134 (2015) ("American negligence law—which gives the question of breach in a negligence case to the jury—can be understood as delegating the normative decision about which risks are justifiable to the jury.").

The law makes many dubious assumptions, of course, and the idea that normal humans grasp these concepts might be one of them. But even if this assumption is overly simple, it likely points to the fact that some normative concepts are basic, and these particular concepts are as followable as any because of their prevalence in everyday life.

Because the epistemic-procedural requirements BROAD EVIDENCE invokes can be understood in terms of concepts with which we are as familiar as any, and because requiring that would-be blamers meet these conditions before inflicting punishment or blame is a substantial step toward protecting the innocent, I suggest the costs of this more demanding principle are worth their price. This is a normative judgment that might be disputed, but contractualism commits us to ending our story with a judgment of just this sort. I therefore suggest that BROAD EVIDENCE, understood as equivalent NEGLIGENCE and REASONABILITY, is the least rejectable principle we've considered.

VI. An Outstanding Issue: Moral Procedural Requirements?

But even if this is correct, difficult questions remain. Particularly salient is the question whether moral considerations can affect the content of our epistemic-procedural requirements, and if not, whether BROAD EVIDENCE might be too narrow yet.

To explain, Rosen argues that ignorance can be blameworthy even when the ignorance is epistemically justifiable in the broad sense, i.e., narrowly justified and the subject has met all her procedural-epistemic requirements. He considers the case of Goldberg, who's charged with criminally investigating Himmelfarb. Goldberg investigates scrupulously, doing "everything a responsible investigator ought to do, and his evidence amply supports the conclusion that

Himmelfarb is a crook."[352] So Goldberg forms the belief that Himmelfarb is criminally blameworthy, files a report, and accuses Himmelfarb of a crime. As it turns out, Himmelfarb has been framed by a malicious colleague. "[I]there is nothing more to the story," Rosen claims, "this is a straightforward case of blameless wrongdoing from excusable ignorance."

There is more to the story, however.

> Goldberg had promised Himmelfarb's mother that he would spend an extra day on the investigation digging more deeply into the evidence, despite the fact that he (and she) had no reason to believe that further inquiry would reveal anything new. Since he was in a position to keep his promise without great cost, Goldberg was morally obligated to put in the extra time. But he did not. . . . We may add that if he had kept his promise, he would have discovered the fraud and exonerated Himmelfarb.[353]

Rosen claims that Goldberg's ignorance is culpable in light of these additional details.[354] This is despite the fact the promise he made does *not* support an additional procedural-*epistemic* obligation.

> The active side of epistemic life is constrained by a range of obligations, only some of which are recognizably epistemic. A belief is justified in the broad sense when the agent has complied with his procedural epistemic obligations and his opinion is supported by his total evidence. But some of our nonepistemic obligations (like Goldberg's obligation to spend the extra day) have something important in common with our epistemic obligations properly so called. In a somewhat elusive sense, these obligations are grounded in a concern for truth and the avoidance of error. When one violates such obligations, one is reckless or negligent in a familiar sense: one has failed to take a precaution against a certain bad outcome—ignorance—that one was morally obliged to take in the circumstances. Since ignorance is a foreseeable (if unlikely) consequence of this sort of negligence, one will normally be at fault for any ignorance that results from such a failure, even if the resulting ignorance is fully justified in the epistemic sense.[355] (603)

[352] Rosen, *supra* note 345, at 601.
[353] *Id.* at 601-02.
[354] *Id.* at 602.
[355] *Id.* at 603.

We could accept Rosen's claims in the context of culpability for ignorance without accepting them in the context of the epistemic condition on appropriate blame. Rosen, after all, is not speaking directly to this chapter's concerns; it just so happens that the wrongdoing Goldberg commits is punitive in nature.

But Rosen's analysis also happens to speak directly to why BROAD EVIDENCE might be too narrow. Goldberg was negligent, in a sense: he failed to take a precaution against a certain bad outcome—inflicting blame and punishment on the innocent—that he was morally obligated to take in the circumstance. Rosen's analysis also suggests it may have been too fast to equate BROAD EVIDENCE with NON-NEGLIGENCE and REASONABILITY: the first makes clear that the relevant procedural requirements are purely epistemic, whereas the latter two leave open whether the negligence or reasonability at issue is exclusively epistemic.

Goldberg's case is idiosyncratic. Plausibly, CB should treat a promise like Goldberg's as the kind of idiosyncratic feature that contractualism's generic reason requirement instructs us to ignore.[356] But this is too fast. For while Goldberg's promise is certainly idiosyncratic, it is commonplace for a situation to have special features that seem to ground similar moral obligations. Say, for example, that your child has been accused of a terrible crime. I have no relationship with your child, but each of us takes the same investigative steps and comes to the same compelling evidence in favor of their guilt. I meet the standards of BROAD EVIDENCE, so I blame your child. Now it seems we might say one of three things. Either,

(1) You may not meet the standards of BROAD EVIDENCE, despite the fact I meet those standards, because the extent of the inquiry the principle requires of each of us is different.

[356] *See* SCANLON, WHAT WE OWE TO EACH OTHER 202-206 (1998) (explaining that contractualist reasoning must not take into consideration idiosyncratic features of particular agents when determining the general rules to govern behavior).

Plausibly, this is because a form of moral encroachment theory is true: the principle's requirements turn, in part, on what's morally at stake, and the moral stakes of your treating your own child with punitive hostility are higher than the stakes of my doing so.[357] Or,

(2) We both meet the standards of BROAD EVIDENCE, but you face an additional moral obligation to investigate further. Whether you meet this additional obligation does not affect the appropriateness of your blame because the permissibility of punishment does not turn on whether you've met your *non-epistemic* obligation. Or,

(3) We both meet the standards of BROAD EVIDENCE, but you face an additional moral obligation to investigate further. If you do not meet this obligation, then your blame is inappropriate. This is because BROAD EVIDENCE is too narrow, and the unrejectable principle requires meeting both the applicable epistemic- *and moral*-procedural requirements.

I am inclined toward (1), myself. On this view, BROAD EVIDENCE is equivalent to NON-NEGLIGENCE and REASONABILITY because in the context of belief formation, there are only epistemic-procedural requirements; moral considerations like promises to investigate and special relationships come into play in determining the stringency of those requirements. Barring further inquiry, my view is thus that BROAD EVIDENCE is equivalent to NON-NEGLIGENCE and REASONABILITY, and BROAD EVIDENCE is the epistemic condition on appropriate blame.

But I do not raise these difficult questions to answer them. Instead, I raise them to underline the fact that there remains much work to do in this area, even if BROAD EVIDENCE is largely correct. Some of this work can be done using the apparatus of CB. But some of it falls

[357] For an argument to this effect, *see* James Fritz, *Moral Encroachment and Reasons of the Wrong Kind*, 177 PHIL. STUDIES 3051, 3057-59 (2020) (arguing that moral reasons can affect whether an agent is required to withhold belief until she acquires more evidence). For an argument that moral considerations can raise the evidential bar that must be met for belief to be *narrowly* justified, *see* Rima Basu & Mark Schroeder, *Doxastic Wronging*, *in* PRAGMATIC ENCROACHMENT IN EPISTEMOLOGY 181, § 4.1 (Brian Kim & Matthew McGrath eds., 2018).

squarely into the realm of epistemology. Whether my preferred option, (1), is available turns on the outcome of debates in the theory of moral encroachment. These debates are not to be had within the CB framework, and yet the application of the framework to this context remains beholden to their outcome.

VII. Conclusion

In sum, we should be confident that CB militates in favor of *some* epistemic appropriateness condition. And we should be confident that this is, in at least large part, to protect the blameless. Moreover, because the blameless are already protected by principles that proscribe punishment for those who are not blameworthy, we should be confident that *followability* is a central concern for any principle that would play the role of an epistemic appropriateness condition.

I have suggested that BROAD EVIDENCE is unrejectable, but whether it must be supplemented to take account of non-epistemic procedural requirements, whether it is less rejectable than NARROW EVIDENCE, whether a modified credence-based principle might be superior, and whether the EVIDENCE principles are superior to KNOWLEDGE, are all difficult questions that this chapter does not purport to have settled. But it should be no surprise that much work remains to be done in this important but fraught and underexplored area of philosophical thought. In the spirit of BROAD EVIDENCE, this is no time to settle our beliefs. It is time to inquire further.