THE YALE LAW JOURNAL

GABRIEL S. MENDLOW

Why Is It Wrong To Punish Thought?

ABSTRACT. It's a venerable maxim of criminal jurisprudence that the state must never punish people for their mere thoughts—for their beliefs, desires, fantasies, and unexecuted intentions. This maxim is all but unquestioned, yet its true justification is something of a mystery. In this Essay, I argue that each of the prevailing justifications is deficient, and I conclude by proposing a novel one. The proposed justification captures the widely shared intuition that punishing a person for her mere thoughts isn't simply disfavored by the balance of reasons but is morally wrongful in itself, an intrinsic (i.e., consequence-independent) injustice to the person punished. The proposed justification also shows how thought's immunity from punishment relates to a principle of freedom of mind, a linkage often assumed but never explained. In explaining it here, I argue that thought's penal immunity springs from the interaction of two principles of broad significance: one familiar but poorly understood, the other seemingly unnoticed. The familiar principle is that persons possess a *right of mental integrity*, a right to be free from the direct and forcible manipulation of their minds. The unnoticed principle, which I label the *Enforceability Constraint*, is that the state's authority to punish transgressions of a given type extends no further than its authority to thwart or disrupt such transgressions using direct compulsive force. Heretofore unexamined, the Enforceability Constraint is in fact a signal feature of our system of criminal administration, governing the scope and limits of the criminal law.

AUTHOR. Assistant Professor of Law and of Philosophy, University of Michigan. I'm grateful to audiences at the American Philosophical Association Central Division Conference, the Analytical Legal Philosophy Conference, the Fordham Law School Criminal Law Theory Discussion Group, the London School of Economics Criminal Law and Criminal Justice Theory Forum, the University of Michigan Criminal Law Theory and Legal Theory Workshops, the University of Oxford Criminal Law Discussion Group, the Queen's University Faculty of Law, the University of Surrey School of Law Faculty Research Seminar, the University of Toronto Faculty of Law Criminal Law Theory Lab, the Yale Law School Legal Theory Workshop, and a symposium conference at Osgoode Hall Law School on the scope and limits of the criminal law. Special thanks to Larry Alexander, Paul Audi, Marcia Baron, Mitch Berman, David Brink, Sarah Buss, Jules Coleman, Nico Cornell, Susan Dimock, Antony Duff, James Edwards, Chris Essert, Rich Friedman, Stephen Galoob, Tom Green, Monica Hakimi, Daniel Halberstam, Scott Hershovitz, Don Herzog, Doug Husak, Brian Hutler, Vic Khanna, Josh Kleinfeld, Adrienne Lapidos, Kyle Logue, Daniel Markovits, Bill Miller, Michael Moore, Julian Mortenson, Federico Picinali, J.J. Prescott, Eve Primus, Richard Primus, Don Regan, Jed Rubenfeld, Alex Sarch, Scott Shapiro, Seana Shiffrin, Sonja Starr, Hamish Stewart, Malcolm Thorburn, Peter Westen, Gideon Yaffe, and Taisu Zhang. For research assistance, I'm grateful to Daniel Fryer, Andrew Jordan, Ross Macpherson, and Virginia Neisler, the Faculty Services Librarian at Michigan Law School. I also would like to thank the editors of the *Yale Law Journal*, especially Leslie Arffa. Research for this project was funded in part by the Cook Endowment.

ESSAY CONTENTS

> If there is any one proposition that commands general agreement among theorists and practitioners of the penal law, it is that judicial punishment ought not to be inflicted for private thoughts, wishes, inclinations, or states of character where these have not manifested themselves in conduct. Theorists from otherwise opposing philosophic schools converge on this principle.
>
> –Alan Brudner, *Punishment and Freedom: A Liberal Theory of Penal Justice*[1]

> [W]hat allegedly renders liability for [unexecuted] intentions objectionable is quite mysterious . . . .
>
> –Douglas N. Husak, *Philosophy of Criminal Law*[2]

INTRODUCTION

It's a venerable maxim of criminal jurisprudence that the state must never punish people for their mere thoughts—for their beliefs, desires, fantasies, and unexecuted intentions. This maxim is all but unquestioned, yet its true justification is something of a mystery. Jurists often say that mere thoughts are unpunishable because they're harmless, innocent, and unprovable. But, as I'll argue in Part I, certain thoughts are every bit as dangerous, wrongful, and provable as actions we readily criminalize. If mere thoughts are unpunishable, it's instead because they're *immune* from punishment despite deserving it. Unlike various legal immunities, however, the immunity of thought can't rest on a pragmatic foundation. Although the specter of intrusively oppressive policing may give us reason to *treat* thoughts as immune from punishment, it doesn't establish that they actually are. It doesn't establish that every act of punishment for thought involves an intrinsic (i.e., consequence-independent) injustice to the person punished: that every such act necessarily *wrongs the thinker*. In an influential set of books and articles, R.A. Duff has sought to ground the intrinsic injustice of punishment for thought in the value of moral autonomy. But, as I'll argue in Part II, Duff's argument presupposes something that Part I reveals as false: that no single thought is dangerous or wrongful enough to warrant punishment.

In place of these flawed rationales, Part III proposes that punishment for thought is intrinsically unjust because it's a form of indirect mind control. The proposed rationale captures the widely shared intuition that punishment for thought isn't simply disfavored by the balance of reasons but is morally wrongful in itself, an intrinsic injustice to the person punished. The proposed rationale

1. ALAN BRUDNER, PUNISHMENT AND FREEDOM: A LIBERAL THEORY OF PENAL JUSTICE 108 (2009).

2. DOUGLAS N. HUSAK, PHILOSOPHY OF CRIMINAL LAW 97 (1987).

also shows how thought's immunity from punishment relates to a principle of freedom of mind, a linkage often assumed but never explained. In explaining it here, I argue that thought's penal immunity springs from the interaction of two principles of broad significance: one familiar but poorly understood, the other seemingly unnoticed. The familiar principle is that persons possess a *right of mental integrity*, a right to be free from the direct and forcible manipulation of their minds. We'll see that this right undergirds a set of important principles governing the relationship between the mind and and the state (principles concerning such things as education, brainwashing, and forced medication), of which the ban on thought crime is merely one. The seemingly unnoticed principle is that the state's authority to punish transgressions of a given type extends no further than its authority to thwart or disrupt such transgressions using direct compulsive force. This principle, which I call the *Enforceability Constraint*, holds that the state may ensure compliance with a given norm through criminal punishment only when the state may in principle force compliance with that norm directly.

Heretofore unexamined, the Enforceability Constraint is in fact a signal feature of our system of criminal administration, governing the scope and limits of the criminal law. When conjoined with the principle that persons possess a right of mental integrity, the Enforceability Constraint entails that punishment for thought is intrinsically unjust: if using mind control to force compliance with a thought-proscribing norm would violate a potential norm-breaker's right to mental integrity, then so too would exposing the norm-breaker to punishment. That is why it's wrong to punish thought.

I. INADEQUATE RATIONALES FOR THE BAN ON THOUGHT CRIME

Theorists often claim that criminalizing mere thought would unleash the worst sort of tyranny and oppression. According to James Fitzjames Stephen, if we criminalized every improper thought, "all mankind would be criminals, and most of their lives would be passed in trying and punishing each other for offenses which could never be proved."[3] H.L.A. Hart adds: "Not only would it be a matter of extreme difficulty to ferret out those who were guilty of harboring, but not executing, mere intentions to commit crimes, but the effort to do so

3. 2 JAMES FITZJAMES STEPHEN, A HISTORY OF THE CRIMINAL LAW OF ENGLAND 78 (1883).

would involve vast incursions into individual privacy and liberty."[4] Quoting Stephen, Hart concludes: "[T]o punish bare intention 'would be utterly intolerable.'"[5]

These assertions are facile. To be sure, life would be intolerable under a regime that punished every improper mental state — every sadistic fantasy, evil desire, and hateful belief. But life also would be intolerable under a regime that punished every improper act — every unkindness and petty betrayal, no matter how harmless, innocent, or difficult to prove. That's an excellent reason not to punish every improper act. It's a terrible reason not to punish *any* act. In punishing acts, legal systems can and do discriminate between the grave and the paltry. If a legal system elected to punish thoughts (the word I'll often use to denote the entire class of mental states), the state could exercise like discretion, punishing only the rare thought that's dangerous, depraved, and provable. The key question is whether any such thought exists, and it's a question that Stephen and Hart evade.

I'll argue that the answer is yes. Contrary to the received wisdom, certain thoughts are dangerous, depraved, and provable. Thus, the ban on punishing mere thoughts can't be justified by any of the leading rationales: the harm principle, the requirement that criminal transgressions be culpable wrongs, or the requirement that criminal transgressions be proved beyond a reasonable doubt.

I'll consider these rationales in turn.

A. The Harm Principle

Reporting a common view, P.J. Fitzgerald notes that "[t]he comparative harmlessness of mere thoughts and intentions by themselves is considered sufficient reason for not punishing them. The small degree of harm likely to result from such intentions is not thought to justify the interference with liberty which punishment would involve."[6] If thoughts aren't more than minimally harmful, then criminalizing them violates John Stuart Mill's harm principle. According to the harm principle, "the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others."[7] But is Fitzgerald right that thoughts never risk more than a "small degree of harm"?

---

4. H.L.A. HART, PUNISHMENT AND RESPONSIBILITY: ESSAYS IN THE PHILOSOPHY OF LAW 127 (2d ed. 2008).

5. *Id.* (quoting 2 STEPHEN, *supra* note 3, at 78).

6. P.J. FITZGERALD, CRIMINAL LAW AND PUNISHMENT 97 (1962).

7. JOHN STUART MILL, ON LIBERTY 9 (Elizabeth Rapaport ed., Hackett Publishing Co. 1978) (1859).

Actions sometimes risk a large degree of harm, and actions typically flow from thoughts. So the question isn't whether thoughts ever risk harm. It's *which* thoughts risk harm, and how much.

As a class, thoughts vary greatly with respect to how much harm they risk because they vary greatly with respect to the likelihood that they'll lead to actions. Some thoughts are basically inert. For example, a non-normative belief (e.g., that there's water in my cup) won't incline me to act unless accompanied by an "active thought"[8] like a normative belief (e.g., that I should avoid dehydration) or a desire or intention (e.g., to drink). Even within the sub-class of "active thoughts," mental states come in two fundamentally different varieties, as Duff explains:

> First, there are those the completion of which requires no world-impacting action: fantasising or contemplating, for instance, might lead to overt action but are not necessarily frustrated without it; they can be completed within the realm of thought. Secondly, there are kinds of thought the completion of which requires overt action. Decision and intention formation are obvious examples: whilst I can fail to do what I decide or intend to do, such lack of overt action frustrates my decision or intention; such thinking demands overt action in a way that the first kind does not.[9]

Conceivably, we'd contravene the harm principle if we criminalized thoughts of the first kind, thoughts "the completion of which requires no world-impacting action"—although the threat that heterodox beliefs pose to authoritarian governments and the resultant zeal with which they're criminalized both bespeak a darker and not altogether implausible view of the dangerousness of thoughts that by their nature "might lead to overt action but are not necessarily frustrated without it." The harm principle presents far less of an obstacle to criminalizing mental states of the second kind, those "the completion of which requires overt action." It's for this reason that the Essay will focus primarily on a particular aspect of the prohibition on punishing mere thought: namely, the harder-to-justify prohibition on punishing mere *intent*.

Fitzgerald is simply wrong to assume that unexecuted intentions risk only a "small degree of harm." Consider a person's intention to kill, particularly when formed after extensive reflection and deliberation. Is such an intention really less likely to cause harm than driving recklessly or possessing volatile explosives—activities that we don't hesitate to criminalize on account of their dangerousness?

---

8. I borrow the term from R.A. Duff. *See* R.A. DUFF, ANSWERING FOR CRIME: RESPONSIBILITY AND LIABILITY IN THE CRIMINAL LAW 102 (2007) (distinguishing between two kinds of "active thought").

9. *Id.*

If no lethal intention were more than minimally dangerous, it would be irrational for me to fear you simply because you intended to kill me. But it's difficult to accept that such fear is irrational. There would be little point to forming intentions if intentions didn't generally increase the likelihood of actions. It's one thing for you to *want* to kill your enemy, or to *believe* that killing him has something to be said for it. Wanting and believing these things are common enough occurrences, which don't necessarily indicate a propensity to violence. It's another thing entirely for you to *intend* to kill your enemy, to make killing him your goal. To make killing someone your goal is to embrace a distinctive and unusual set of rational commitments. It's to commit to watching for an opportunity to kill him, to seizing such an opportunity when practicable, and to refraining from conduct that would make performance impossible. Rational commitments of this sort are what distinguish intending to kill, which is rare, from desiring to kill, which is sometimes said to be common. Rationality doesn't demand of one who desires to kill that she abandon all contrary intentions. Rationality doesn't even demand that she abandon all contrary desires. But rationality does demand that an intending killer kill, or else abandon her intention.

It's true that intentions can be rescinded, decisions rethought, and plans discarded, but it doesn't follow that your intending to do something never increases the likelihood that you'll do it. A characteristic effect of forming an intention is to place yourself under rational and psychological pressure to follow through, pressure compounded by a range of familiar cognitive biases that further reduce the likelihood you'll change your mind. The more invested we feel in a decision, the less likely we are to reconsider it.[10] We also tend to remember our past decisions as being more justified than they actually were.[11] When confronted with new evidence, we tend to revise our opinions insufficiently.[12] And we generally

10. *See generally* Hal R. Arkes & Catherine Blumer, *The Psychology of Sunk Cost*, 35 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 124 (1985) (documenting the "sunk cost effect" that leads actors, after making a substantial investment, to continue endeavors they might have otherwise abandoned, in order to avoid appearing wasteful).

11. *See generally* Jack W. Brehm, *Postdecision Changes in the Desirability of Alternatives*, 52 J. ABNORMAL & SOC. PSYCHOL. 384, 389 (1956) (showing a tendency among test subjects to reduce dissonance following decisions by "making the chosen alternative more desirable and the unchosen alternative less desirable"); Mara Mather, Eldar Shafir & Marcia K. Johnson, *Misremembrance of Options Past: Source Monitoring and Choice*, 11 PSYCHOL. SCI. 132, 132 (2000) (revealing "choice-supportive memory distortion" of past decisions).

12. *See* Ward Edwards, *Conservatism in Human Information Processing*, *in* FORMAL REPRESENTATION OF HUMAN JUDGMENT 17, 18 (Benjamin Kleinmuntz ed., 1968).

tend to place more credence in evidence that confirms our beliefs than in evidence that contradicts them.[13] Evidence that contradicts our beliefs sometimes perversely strengthens them.[14]

If the rational pressures intrinsic to intention and the cognitive biases that reinforce those pressures all increase the odds that you'll do what you intend to do, forming a lethal intention creates a risk of death.[15] If you're a competent person with the means to kill, the danger posed by your lethal intention could be at least as great as that posed by many risky activities we seldom think twice about punishing, such as driving recklessly and possessing volatile explosives.[16]

B. The Requirement That Criminal Transgressions Be Culpable Wrongs

Not only can lethal intentions be dangerous, but for that very reason they also can be culpably wrongful, at least potentially. If it's sometimes culpably wrongful to create a risk of nondeadly injury inadvertently, then presumably it's sometimes culpably wrongful to create a risk of deadly injury knowingly—which is what you do when you form the intention to kill, assuming you're a competent person with the necessary means. Knowingly creating a risk of death is a serious wrong, a wrong the public seemingly has standing to condemn. It's hard to accept that the public could lack standing to complain of some risk just because the risk originates inside a person's head rather than on the outside. The site of the risk seems to lack independent moral significance.

It's true that a risk generated externally (e.g., by driving recklessly or possessing volatile explosives) typically isn't within a person's exclusive control, whereas a person can control the risk generated by her malevolent intention, in that she can extinguish the risk just by abandoning the intention. Does this mean, as Larry Alexander and Kimberly Ferzan argue, that an intending criminal

---

13. *See generally* Raymond S. Nickerson, *Confirmation Bias: A Ubiquitous Phenomenon in Many Guises*, 2 REV. GEN. PSYCHOL. 175 (1998) (providing a broad overview of how confirmation bias acts in a variety of contexts to "account for a significant fraction of the disputes, altercations, and misunderstandings that occur among individuals, groups, and nations").

14. *See generally* Lawrence J. Sanna, Norbert Schwarz & Shevaun L. Stocker, *When Debiasing Backfires: Accessible Content and Accessibility Experiences in Debiasing Hindsight*, 28 J. EXPERIMENTAL PSYCHOL.: LEARNING, MEMORY, & COGNITION 497 (2002) (showing that when test subjects were asked to list alternatives to a decision that they had made, they were more likely to feel justified in making their original decision).

15. *Cf.* DOUGLAS HUSAK, *Does Criminal Liability Require an Act?*, *in* THE PHILOSOPHY OF CRIMINAL LAW: SELECTED ESSAYS 17, 50 (2010) (arguing that some "firm intentions impermissibly increase the risk of a subsequent harm").

16. *Cf.* Frederick Schauer, *On the Distinction Between Speech and Action*, 65 EMORY L.J. 427, 439-47 (2015) (comparing the potential harm from dangerous thoughts to the potential harm from preparatory acts that are in fact criminalized).

isn't culpable for her mere intentions?[17] That she isn't culpable for anything at all until she does something to "unleash" a risk over which she lacks complete control?[18]

I don't see why. When a person forms the intention to kill, she culpably creates in herself a psychological condition the purpose and possible effect of which is to cause a death. Although she can eliminate the risk of death by abandoning the intention, we shouldn't pretend that abandoning an intention is as easy as flipping a mental switch. As I noted a moment ago, intentions carry substantial mental inertia.[19] When a person forms the intention to kill, she sets herself on a path that makes someone's death at least a little bit more likely—just as a person may do when she acquires a safely stored but very deadly weapon or appropriates the nuclear launch codes. Like forming a lethal intention, these activities may properly be subjected to public condemnation even though the risks they create remain exclusively within the actor's control. It's everyone's business when someone knowingly creates an impermissible risk, wherever and by whatever means.

But riskiness is only part of what makes lethal intentions wrongful, and probably not even the largest part. If, thanks to fortuity or incompetence, your intention to kill me creates no appreciable risk that I'll die, you wrong me nonetheless, just by aiming at my death. The wrongfulness of your intention derives not only from the risk it creates, but also—and perhaps more fundamentally—from the wrongfulness of the action toward which it aims. Ordinarily, you have a conclusive moral reason not to kill me, which is virtually always[20] also a conclusive moral reason not to *try* to kill me, *prepare* to kill me, *plot* to kill me, *plan*

17. *See* LARRY ALEXANDER & KIMBERLY KESSLER FERZAN WITH STEPHEN MORSE, CRIME AND CULPABILITY: A THEORY OF CRIMINAL LAW 197-216 (2009); *cf.* Federico Picinali, *A Retributive Justification for Not Punishing Bare Intentions or: On the Moral Relevance of the 'Now-Belief,'* 32 L. & PHIL. 385, 386 (2013) (arguing that unexecuted intentions are categorically less culpable than executed ones). The most prominent defenders of the claim that intentions aren't culpable wrongs, Alexander and Ferzan, put forward several arguments that warrant more attention than I can give them here. (For an extended discussion of their arguments, see R.A. Duff, *Risks, Culpability and Criminal Liability*, *in* SEEKING SECURITY: PRE-EMPTING THE COMMISSION OF CRIMINAL HARMS 121, 126-42 (G.R. Sullivan & Ian Dennis eds., 2012).) What Alexander and Ferzan's arguments ultimately show, I believe, isn't that intentions aren't culpable at all, but that a given intention is *less* culpable when unexecuted than when acted upon. If unexecuted intentions are culpable to some degree, then the pivotal question is whether they're culpable to a *sufficient* degree to warrant criminalization. I'm unpersuaded that the answer to this last question is always no, as I indicate in the text.

18. ALEXANDER & KESSLER, *supra* note 17, at 197.

19. *See supra* notes 10-14 and accompanying text.

20. In a bizarre scenario like Gregory Kavka's toxin puzzle, *see* Gregory S. Kavka, *The Toxin Puzzle*, 43 ANALYSIS 33, 33-35 (1983), your conclusive moral reason not to kill me might be no more than a nonconclusive (i.e., defeated or outweighed) moral reason not to intend to kill me.

to kill me, or *intend* to kill me.[21] When you form the intention to kill me, you therefore do something you have a conclusive moral reason not to do. And when you do something you have a conclusive moral reason not to do, you do something wrongful — even if all you do is form a mental state.

It's therefore unsurprising that the wrongfulness of malevolent intentions is presupposed by a range of moral judgments and emotional reactions both natural and inevitable. Consider the host of attitudes and demands we'd have to disclaim if your unexecuted intention to kill me weren't a culpable wrong. I couldn't resent you for your intention. I couldn't demand that you abandon it. I couldn't even demand that you apologize for it. I could think the worse of you on account of your intention, but I couldn't say, "How dare you intend to kill me?" If you've done me no wrong, I lack the standing to condemn you. Although I could view your intention as a moral failing — a character flaw — I couldn't view it as a moral transgression. I couldn't view it as a moral transgression even if you unquestionably formed it voluntarily. And it seems clear that at least some intentions are subject to a person's voluntary control, particularly intentions that a person forms after reflection and deliberation.[22] Indeed, even mental states *not* within a person's voluntary control can qualify as culpable wrongs — apt objects of moral blame, if not of criminal punishment.[23] If I discover that you believe I'm miserly and dishonest because of my ethnicity, I'll resent you and demand an apology. And I won't be remotely inclined to forgive you even if (*especially* if) you protest with apparent sincerity that your beliefs about my character spring from a prejudice that's ingrained and involuntary. It therefore strikes me as implausible that your prejudiced beliefs don't wrong me, that they're merely immoralities of character rather than moral transgressions. It strikes me as more implausible still that your (voluntarily-formed) intention to kill me isn't a proper object of resentment, apology, or forgiveness. But if intentions aren't culpable wrongs, these implausible consequences follow.

Suppose an eccentric billionaire offers to pay you a million dollars if, at midnight tonight, you intend to kill me tomorrow afternoon. He emphasizes that the money will be in your bank account by 10 a.m. tomorrow morning, so you don't actually have to go through with the killing. You just have to *intend* to. In this scenario, you've got a conclusive moral reason not to kill me, but a defeated moral reason not to intend to kill me. I assume that scenarios with these rational implications are exceedingly rare.

21. *See* Duff, *supra* note 17, at 135-36.

22. *Cf.* HUSAK, *supra* note 15, at 48 ("[W]hether a person decides to rob a bank, deliberates over time, anticipates every contingency, and carefully plans an escape seems no less under individual control than whether he or she commits any number of acts that unquestionably are the proper objects of criminal liability.").

23. *See generally* Robert Merrihew Adams, *Involuntary Sins*, 94 PHIL. REV. 3 (1985) (arguing that we should be held morally, but not legally, accountable for involuntary "bad states of mind").

What's not implausible is that your intention to kill me is less wrongful than your partial attempt to do so, and your partial attempt less wrongful than your completed attempt.[24] As William Blackstone remarked, "[G]enerally, a design to transgress is not so flagrant an enormity, as the actual completion of that design. For evil, the nearer we approach it, is the more disagreeable and shocking; so that it requires more obstinacy in wickedness to perpetrate an unlawful action, than barely to entertain the thought of it . . . ."[25] But even if Blackstone is right to think that every wrongful action is more wrongful than the intention from which it flows, it doesn't follow that every wrongful action is more wrongful than any possible intention.

If, as I've just argued, some intentions are at least as wrongful as certain punishable actions, then we should demand to know why the intentions alone are off limits to the criminal law. If mere intentions are categorically immune from punishment, they must possess a kind of privilege. What is the nature of that privilege and why do actions lack it?

Belying the difficulty of this question, many people seem to accept something like the simplistic answer offered by Francis Wharton, who suggests that the performance of an action constitutes a *forfeiture* of the privilege:

> The mere unexecuted purpose of thieving does not make a thief. . . . It is sure that as soon as the intention touches and makes an impress on the outside world, this immunity from prosecution ceases. He who, intending to steal, has false keys made for the purpose of entering a room where money is kept, and who begins the work of purloining by applying the keys to the door, may be indicted for the attempt; and so may he who lays an ambuscade for another, arranging the materials of ambush so as to facilitate the surprise; and so may he who puts in operation a plan by which a forgery will be effected, unless some extraneous influence intervenes.[26]

If performing an action forfeits one's immunity from punishment, we need some explanation of *why* this is so. Is it because an action "touches and makes an impress on the outside world," as Wharton suggests—because an action has direct

---

24. *See* David O. Brink, *The Path to Completion*, *in* 4 OXFORD STUDIES IN AGENCY AND RESPONSIBILITY 183, 184 (David Shoemaker ed., 2017) (arguing for a "historical and scalar" conception of attempt, according to which partial attempts vary in heinousness depending on, among other things, how much the actor has already done in furtherance of his criminal intention, and how much remains to be done).

25. *See* 4 WILLIAM BLACKSTONE, COMMENTARIES *15.

26. Francis Wharton, *Comparative Criminal Jurisprudence*, 4 CRIM. L. MAG. 1, 5 (1883).

physical effects? This explanation explains nothing: it identifies a difference between actions and mere mental states but gives no account of why the difference matters. Why is it significant that actions have physical effects? Our actions, thanks to their physical effects, can risk causing harm and can constitute public wrongs, wrongs of proper concern to the polity. But, as we've seen, so too can our intentions — before they've had any physical effects.

C. *The Requirement That Criminal Transgressions Be Proved Beyond a Reasonable Doubt*

Even if dangerous and wrongful, lethal intentions would be inapt for punishment if, as Stephen asserts, they "could never be proved."[27] The idea that mere intentions are unpunishable because unprovable goes back at least to Blackstone, who wrote:

> [A] fixed design or will to do an unlawful act is almost as heinous as the commission of it, yet, as no temporal tribunal can search the heart, or fathom the intentions of the mind, otherwise than as they are demonstrated by outward actions, it therefore cannot punish for what it cannot know.[28]

The scope of Blackstone's claim is uncertain, thanks to the potential ambiguity of the phrase "intentions . . . demonstrated by outward actions." If the phrase refers to intentions manifested somehow in a person's acts — possibly including expressive acts that the person doesn't intend as a means of bringing his intentions closer to fruition — then Blackstone's claim is an exceedingly narrow one. It's the claim that a given malign intention is practically unprovable and therefore unpunishable if it's not manifested through any outward act whatsoever, even an expressive but nonexecutory act like a confession. This amounts to the claim that an intention is practically unprovable if it's utterly secret, a claim that's nearly tautological. To quote Giorgio Del Vecchio, "when it is said that an act of thought is not punishable, reference is [made] . . . to a *known* act of thought."[29] He continues: "If the maxim . . . refers to secret thoughts it is absurd, because not thought alone, but every act whatsoever, is unpunishable as long as it is hidden."[30] Rightly or not, Blackstone generally is understood to be making a

---

27. 2 STEPHEN, *supra* note 3, at 78.

28. 4 BLACKSTONE, *supra* note 25, at *21.

29. GIORGIO DEL VECCHIO, THE FORMAL BASES OF LAW 140 (John Lisle trans., Legal Classics Library spec. ed. 2001) (1878) (emphasis added).

30. *Id.* at 140 n.30.

stronger claim: that a person's malign intention is unprovable and therefore unpunishable not only when wholly unmanifested but also when wholly *unexecuted*. In other words, a person's intention is unprovable and therefore unpunishable when the person has performed no act in furtherance of his intention—no act that he intends either as a means of fulfilling the intention or as a means of facilitating its fulfillment.[31] I'll refer to this claim as *Blackstone's Principle*.[32]

Is Blackstone's Principle true? It's instructive that criminal doctrine doesn't fully reflect it. Take, for example, § 2119 of Title 18 of the U.S. Code, which makes it a crime to engage in carjacking "with the intent to cause death."[33] As the Supreme Court has interpreted § 2119, a carjacker's lethal intention will satisfy the statute's mens rea requirement even if the intention is conditional: even if the carjacker intends *to kill the car's owner only on the condition that doing so is necessary to accomplish the carjacking*.[34] Thus, a carjacker violates § 2119 even when he plans and hopes not to kill but is reluctantly willing to do so if there's no other way of wresting the car from its owner. To violate the statute, the carjacker needn't do anything at all with the purpose of enabling himself to carry out his conditional intention to kill. Neither the statute nor the rules of evidence require that the prosecution prove the carjacker's conditional intention from evidence of acts the carjacker performed in furtherance of his intention. As far as the law of evidence is concerned, the intention can be proved entirely by the carjacker's confession, by a statement to his accomplice, or by an entry in his private diary.[35]

---

31. *See, e.g.*, Patrick A. Broderick, *Conditional Objectives of Conspiracies*, 94 YALE L.J. 895, 896 n.11 (1985) (ascribing to Blackstone the thesis that mere intentions are unpunishable because unprovable); Leo Katz, *Villainy and Felony: A Problem Concerning Criminalization*, 6 BUFF. CRIM. L. REV. 451, 466 n.43 (2002) (same); Ron Shapira, *Structural Flaws of the "Willed Bodily Movement" Theory of Action*, 1 BUFF. CRIM. L. REV. 349, 376 n.101 (1998) (same); Alec Walen, *Criminalizing Statements of Terrorist Intent: How To Understand the Law Governing Terrorist Threats, and Why It Should Be Used Instead of Long-Term Preventive Detention*, 101 J. CRIM. L. & CRIMINOLOGY 803, 836 (2011) (same); Mark Zingale, *Fashioning a Victim Standard in Mail and Wire Fraud: Ordinarily Prudent Person or Monumentally Credulous Gull?*, 99 COLUM. L. REV. 795, 816 n.116 (1999) (same).

32. Blackstone's Principle pertains only to unexecuted intentions—as it must, because the law treats executed intentions as paradigmatically provable. *See* CHARLES L. BLACK, JR., IMPEACHMENT: A HANDBOOK 26 (1974) ("An old English judge said that 'The Devil himself knoweth not the heart of a man.' But courts have to try, and continually do try, to work out the truth about intents and motives, for these are often . . . of the very essence of [a] charge.").

33. 18 U.S.C. § 2119 (2012).

34. *See* Holloway v. United States, 526 U.S. 1, 12 (1999) ("The intent requirement of [18 U.S.C.] § 2119 is satisfied when the Government proves that at the moment the defendant demanded or took control over the driver's automobile the defendant possessed the intent to seriously harm or kill the driver if necessary to steal the car (or, alternatively, if unnecessary to steal the car).").

35. In *Holloway*, the defendant's conditional intention to kill seems to have been proved entirely by his accomplice's testimony about their "plan." *See id.* at 4 ("The accomplice testified that

If a violation of § 2119 strikes you as at least theoretically provable in circumstances where the carjacker hopes not to kill, plans not to kill, has performed no action with the purpose of enabling himself to kill, but is willing to kill if necessary, then you probably don't accept Blackstone's Principle. Should you? Only if you should accept that unexecuted intentions are categorically unprovable. And I don't see why you should, given that evidence of unexecuted intentions isn't categorically inferior to evidence of other mental states.

It's true that unexecuted intentions can't be proved from *executory evidence*, evidence of actions done to fulfill one's intentions or to make their fulfillment more likely. Unexecuted intentions instead must be proved from *expressive evidence*, such as a person's self-reports. But as a general matter we don't regard expressive evidence as inferior to executory evidence.[36]

Indeed, expressive evidence is arguably the superior form. One reason for the widespread rejection of the "equivocality test" for the actus reus of attempt is that almost no action meets the test's demand of being "in itself sufficient evidence of the criminal intent with which it is done."[37] Criticizing the equivocality test, Glanville Williams suggests that it's in fact executory evidence that's the inferior type:

> D goes up to a haystack, fills his pipe, and lights a match. The act of lighting the match, even to a suspicious-minded person, is ambiguous. It may indicate only that D is going to light his pipe; but perhaps, on the other hand, the pipe is only a "blind" and D is really bent on setting fire to the stack. We do not know . . . . But suppose that as a matter of actual fact D, after his arrest, confesses to the guilty intent, and suppose that that confession is believed. We are now certain of the intent . . . . That the act is ambiguous, which in itself might have created a doubt as to the *mens rea*, no longer matters, for the *mens rea* has been proved by the confession.[38]

While Williams's example illustrates how the probative force of expressive evidence can equal or surpass that of executory evidence, it doesn't show that expressive evidence is ever sufficient on its own to prove an offender's intent. A proponent of Blackstone's Principle could insist that *D*'s confession proved his

---

the plan was to steal the cars without harming the victims, but that he would have used his gun if any of the drivers had given him a 'hard time.'").

36. For a discussion of the comparative strengths of expressive and executory evidence of intentions, see GIDEON YAFFE, ATTEMPTS: IN THE PHILOSOPHY OF ACTION AND THE CRIMINAL LAW 224-28 (2010).

37. *The King v. Barker* [1924] NZLR 865, (CA) at 874 (N.Z.) (Salmond, J.).

38. GLANVILLE WILLIAMS, CRIMINAL LAW: THE GENERAL PART 630 (2d ed. 1961).

intent to start a fire only because we considered the confession in the light of *D*'s executory conduct: going up to a haystack and striking a match. Indeed, a proponent of Blackstone's Principle could claim more generally that, without evidence that a person has taken steps to fulfill his purported intention, we can never know whether the purported intention is anything more than a mere desire or fantasy.[39]

The trouble with this more general claim is that our everyday experience contradicts it. We routinely rely on what a person says he intends to do, well before he's begun to act on his stated intention. Examples range from the mundane to the vital: from the friend who says she'll meet you for lunch at noon to the outlaw country singer who, without moving a muscle, advises you not to let the sun set on you in Tulsa.

To be clear, our credence in a given statement of intention can't rationally stem from that statement alone. Belief in such a statement isn't rational unless we've also got evidence that the person making the statement is minimally trustworthy. Now, it might be (although I don't see why it must be)[40] that evidence of a person's trustworthiness isn't rationally adequate unless it includes evidence of the person's prior actions—evidence that she generally does what she says she intends to do. But that doesn't mean we can never rationally conclude that a person possesses a given intention unless we've seen her take steps to execute it. Usually, it's enough if we've seen her follow through on her past statements of intention. Thus, the body of evidence on which we rationally conclude that a person possesses a given intention needn't include any action done to execute *that very intention*. And that's enough to falsify Blackstone's Principle. The principle's truth or falsehood is a matter of whether unexecuted intentions can be proved at all, not of whether they can be proved from expressive evidence alone.

That Blackstone's Principle is probably false doesn't mean that proving unexecuted intentions is easy. Proving them is difficult, and that difficulty alone is reason not to criminalize them. "[Purely] mental acts being private," writes Michael Moore, "proof of them requires intrusive investigative methods[,] . . . good evidence of [such] mental states is rare, and the required lines (between mental acts and passive states, between intentions and wishes) are fuzzy . . . ."[41] A law criminalizing mere thought is therefore vulnerable to the same sort of objection that Duff mounts against the "first act" test for criminal

39. On the importance and difficulty of distinguishing fantasy from true intention in criminal conspiracy prosecutions, see United States v. Valle, 807 F.3d 508, 522 (2d Cir. 2015) (the "Cannibal Cop" case).

40. Couldn't we rationally conclude that a person is trustworthy based on the demonstrable truth of his past statements even if none of them was specifically about his own intentions?

41. MICHAEL S. MOORE, ACT AND CRIME: THE PHILOSOPHY OF ACTION AND ITS IMPLICATIONS FOR CRIMINAL LAW 48 (1993).

attempts, which deems a defendant guilty of an attempt as soon as he's performed a single act in furtherance of his criminal intention. As Duff observes, a "first act" test would "encourage even greater reliance on confessions, in the likely absence of other conclusive evidence of guilt, with all the dangers of abuse and miscarriage of justice which that involves."[42]

These risks surely give us some reason not to punish thought. Indeed, they give us some reason to treat punishment for thought as though it were morally forbidden. But they don't establish that punishment for thought is morally forbidden in fact. The risk of intrusively oppressive policing doesn't establish that there's an intrinsic (i.e., consequence-independent) injustice in every act of punishment for thought,[43] any more than the unreliability of coerced confessions establishes that there's an intrinsic injustice in every act of interrogational torture. Although it might be politically expedient to oppose torture on instrumental grounds, the basic moral reason to refrain from torture isn't that torture produces unreliable information, or that torturing our adversaries encourages them to torture us when we fall into their hands, or that engaging in torture tends to undermine other legal norms against state brutality.[44] All of these things are probably true, and all of them give us good reason to conduct ourselves as though torture were morally forbidden. But none of them shows that torture actually is forbidden in itself—that each act of torture, irrespective of its consequences, is unjust.

***

We've now seen the inadequacy of the conventional rationales for the ban on thought crime—the harm principle, the requirement that criminal transgressions be culpable wrongs, and the requirement that criminal transgressions be proved beyond a reasonable doubt. We might be tempted to abandon the maxim that it's *categorically* unjust to criminalize mental states and to fall back on a *non-categorical* claim: that it's unjust to criminalize beliefs, fantasies, and other mental states "the completion of which requires no world-impacting action" (to quote Duff), yet not unjust to criminalize firm criminal intentions and certain other

42. R.A. DUFF, CRIMINAL ATTEMPTS 37 (1996).

43. *Cf.* HUSAK, *supra* note 2, at 96 ("Herbert Morris has argued, somewhat convincingly, that a system that punished persons solely for intentions, and never for actions, would deviate in a number of important specified respects from standard and familiar examples of legal systems. But such a conclusion fails on its own terms to explain what is objectionable about each and every statute that punishes persons for intentions." (footnote omitted) (discussing Herbert Morris, *Punishment for Thoughts*, *in* ESSAYS IN LEGAL PHILOSOPHY 95 (Robert Summers ed., 1968))).

44. For a discussion of the mutual dependence among the various legal norms against state brutality, see Jeremy Waldron, *Torture and Positive Law: Jurisprudence for the White House*, 105 COLUM. L. REV. 1681, 1726-34 (2005).

mental states "the completion of which requires overt action," whenever such mental states are dangerous, culpably wrongful, and provable.[45] Douglas Husak endorses basically this claim, asking: "When there is overwhelming evidence that a defendant firmly intends to commit a crime, why should punishment be undeserved simply because the act requirement is unsatisfied? I am unaware of a principled answer to this question."[46]

I'm not aware of a principled answer either. As I've argued, malevolent criminal intentions can be dangerous, culpable, provable wrongs. Still, I hesitate to give up on the idea that it's always unjust to punish thought. The revulsion many commentators—indeed, most people—express at the prospect of punishment for mere mental states seems to emanate from a source firmer than the dubious assumption that no single mental state is culpably wrongful. Commentators vehemently assert that punishing mere mental states transgresses a principle of "natural justice"[47] founded in "the inviolability of thoughts,"[48] a principle whose disregard constitutes a "monstrous"[49] intrusion into a person's "private world"[50] and an invasion of her "essential . . . human right to freedom of thought."[51] These remarks describe a supposed injustice both narrower and deeper than that of punishing someone for a transgression undeserving of punishment. The supposed injustice is narrower in that it's peculiar to the mind, and deeper in that it transcends the injustice of punishing someone for a transgression that isn't culpably wrongful. If punishing someone for a mental state is a "monstrous" intrusion into her "private world," it presumably remains so even when the mental state in question is a dangerous, culpable wrong.

II. THE BAN ON THOUGHT CRIME AS A CATEGORICAL MORAL IMMUNITY

Even if none of the conventional rationales suffices on its own to ground a categorical ban on thought crime, the collective weight of these considerations might well support fidelity to a categorical ban. If it's simply too costly, too risky,

45. DUFF, *supra* note 8, at 102.

46. HUSAK, *supra* note 15, at 51.

47. G.A. Endlich, *The Doctrine of Mens Rea*, 13 CRIM. L. MAG. & REP. 831, 832 (1891).

48. Meir Dan-Cohen, *Harmful Thoughts*, 18 L. & PHIL. 379, 379 (1999).

49. Gideon Yaffe, *Criminal Attempts*, 124 YALE L.J. 92, 101 (2014).

50. Andrew Ashworth, *Attempts*, *in* THE OXFORD HANDBOOK OF PHILOSOPHY OF CRIMINAL LAW 126, 134 (John Deigh & David Dolinko eds., 2011).

51. Clay Calvert, *Freedom of Thought, Offensive Fantasies and the Fundamental Human Right To Hold Deviant Ideas: Why the Seventh Circuit Got It Wrong in* Doe v. City of Lafayette, Indiana, 3 PIERCE L. REV. 125, 125 (2005).

and too oppressive to try to distinguish the few mental states that merit punishment from the many that don't, then, on balance, we shouldn't criminalize any. But to adopt a categorical ban on these grounds alone is to give up on the idea that there's an intrinsic (consequence-independent) injustice in each act of punishment for thought. It's to dismiss as hyperbole commentators' assertions about "the inviolability of thoughts"[52] and the "monstrous"[53] intrusion into a person's "private world"[54] that occurs when her thoughts are made the object of punishment. To give up on these ideas and to dismiss the associated rhetoric as hyperbole is akin to giving up on the idea that there's an intrinsic injustice in torture, the idea that torture's injustice isn't solely a function of its downstream consequences.

To view torture's injustice as intrinsic isn't necessarily to see the moral ban on torture as *absolute*. It's instead to see every act of torture as involving a grievous moral sacrifice, even in the hypothetical circumstance in which the state's vital ends supposedly justify its torturous means.[55] I submit that any purported justification of the ban on torture is morally deformed if it gives no account of this moral sacrifice, if it makes no effort to elucidate torture's intrinsic injustice and speaks instead only of torture's instrumental shortcomings. The basic moral reason not to torture is that torturing a person does an injustice *to that person*. The torture victim's signal complaint is that he himself has been wronged, not that the practice to which he's been subjected engenders various other abuses. A person punished for his thoughts is prone to lodge a similar complaint, to complain that he himself has been wronged. This complaint is sound if, but only if, there's an intrinsic injustice in every act of punishment for thought. My question is whether there really is, and, if so, why.

The conventional rationales would yield a satisfactory answer if they supported a categorical ban, as there's little doubt that it's intrinsically unjust to punish someone for a supposed transgression that's harmless or innocent or that can't be proved beyond a reasonable doubt. But, as I argued in Part I, certain thoughts are dangerous, wrongful, and provable — so the conventional rationales can't explain what's intrinsically unjust about *every* act of punishment for thought. What's needed is an explanation of why it's intrinsically unjust to punish mental states that are provable, dangerous, and culpably wrongful: mental states that bear the chief hallmarks of paradigmatic punishable actions.

---

52. Dan-Cohen, *supra* note 48, at 379.

53. Yaffe, *supra* note 49, at 101.

54. Ashworth, *supra* note 50, at 134.

55. For an analysis of the moral sacrifice involved in interrogational torture, see David Sussman, *What's Wrong with Torture?*, 33 PHIL. & PUB. AFF. 1, 19-33 (2005).

In itself, there's nothing especially puzzling about the idea that a class of dangerous and culpably wrongful transgressions is immune from punishment. Criminal law contains a miscellaneous assortment of what Paul Robinson calls "nonexculpatory defenses," defenses like diplomatic, judicial, legislative, and executive immunity, all of which preclude liability "where the actor by all measures deserves condemnation and punishment."[56] These defenses provide a poor analogy to the prohibition on punishing thought, however, because none of them takes its primary justification from the notion that withholding the defense would perpetrate an intrinsic injustice *on defendants*. Rather, as Robinson explains, "[n]onexculpatory defenses arise where an important public policy other than that of convicting culpable offenders, is protected or furthered by foregoing trial or conviction and punishment."[57]

Certainly, the ban on thought crime furthers important public policies — as does the ban's closest counterpart, the ban on punishing speech and other forms of expression. In fact, the most famous of all arguments for freedom of expression, Mill's marketplace-of-ideas argument in Chapter 2 of *On Liberty*, is a classic example of what lawyers call a "policy argument." Mill writes,

> the peculiar evil of silencing the expression of an opinion is, that it is robbing the human race; posterity as well as the existing generation; those who dissent from the opinion, still more than those who hold it. If the opinion is right, they are deprived of the opportunity of exchanging error for truth; if wrong, they lose, what is almost as great a benefit, the clearer perception and livelier impression of truth, produced by its collision with error.[58]

No part of Mill's argument credits the idea that suppressing speech is wrong *because it wrongs the speaker*.

If we're to vindicate the notion that punishing pure thought is wrong *because it wrongs the thinker*, we can't rely on any sort of policy argument. We need an argument that depicts thought's immunity from punishment not as an immunity based in good public policy but as an immunity based in the thinker's status as a moral being.

Duff has propounded just such an argument across a set of books and articles, which collectively constitute the most sophisticated effort to answer this Essay's titular question.[59] At first blush, Duff's argument appears to offer exactly what we need. Instead of denying that criminal intentions are harmful, culpably

---

56. 2 PAUL H. ROBINSON, CRIMINAL LAW DEFENSES § 201 (1984).

57. *Id.*

58. MILL, *supra* note 7, at 16.

59. DUFF, *supra* note 42; Duff, *supra* note 8; Duff, *supra* note 17.

wrongful, or provable, Duff seeks to ground thought's immunity from punishment in the value of moral autonomy. As I'll show, however, Duff's argument covertly presupposes what we've already seen is false: that mere thoughts can never qualify as culpable wrongs.

Duff's argument proceeds from the claim that respect for autonomy precludes punishing anyone who isn't yet "in the process of committing" a crime, a category of people that includes not only those who merely intend to commit a crime and haven't acted, but also those who've taken preparatory steps but haven't yet "crossed the Rubicon."[60] Duff says that if a person intends to commit a crime or is preparing to commit a crime but hasn't yet begun to commit it, we must treat him as a responsible and autonomous agent, someone who has the capacity to change his mind and guide his conduct by the right reasons.[61] If we're to treat him that way, we may remonstrate with him and hope he'll change his mind. But we may not threaten to punish him for his unexecuted intention (in order to induce him to abandon it), nor may we actually punish him for it (in the belief that he won't change his mind).[62] Duff accordingly concludes that if a

60. DUFF, *supra* note 42, at 390 (citations and quotation marks omitted).

61. *Id.* at 388-89 ("To respect someone's freedom as a responsible agent is to see and treat her as someone who is in principle susceptible to rational persuasion; this requires that we seek to modify her conduct only by offering her good reasons to modify it for herself. Suppose we know that someone intends to commit, or is preparing to commit or taking initial steps towards committing, a substantive crime. If we have the moral standing to intervene (which we might claim when what he intends is a crime), we can properly do so by trying to dissuade him from continuing in this criminal enterprise: by appealing to the moral reasons for which he ought to obey the law; or perhaps by reminding him of the prudential reasons for desistance provided by the threat of punishment for the substantive crime. In trying thus to persuade him, we treat him as someone who *could* be persuaded to desist for himself, and who still has time (a *locus poenitentiae*) to desist. We should treat him thus, not necessarily because we think that we might in fact persuade him to desist (we might harbour no real hope of this), but because this is what it is to respect him as a responsible agent. If instead we intervene forcibly to prevent him advancing his criminal enterprise, we cease to treat him as a responsible agent: we deny him the freedom to decide for himself whether to desist; we pre-empt his future actions by force, and thus infringe his autonomy. If the law is to treat its citizens as responsible agents, it must leave them free to decide for themselves, not merely whether to embark on a criminal enterprise, but whether to continue with it.").

62. DUFF, *supra* note 8, at 104-05 ("If the state is to treat its citizens as responsible agents who can be guided (who can guide themselves) by reasons, it should be slow to coerce . . . since that is to treat them as if they will not be guided by the reasons that should dissuade them from such wrongdoing. This is most obviously true when the grounds for that prediction of wrongdoing do not include a present intention to do wrong, as when people are diagnosed as 'dangerous' on the basis of other indicators; but it is also true when the prediction is grounded on the agent's present criminal intention. It is one thing for a fellow citizen or a police officer to warn him that he should abandon his plan: that is still to treat him as a responsible agent who can be moved by the reason for not committing the wrong of which we remind him. It is quite

person merely intends to commit a crime or is preparing to commit a crime but hasn't yet begun to commit it, we may not punish (or threaten to punish) him for his mere intention. "Only when an intending criminal passes beyond the stage of 'mere preparation' and embarks on the commission of the crime itself can we bring the coercive powers of the criminal law to bear on her without infringing her status as a responsible agent."[63] Before that point, "the law should leave intending criminals a *locus poenitentiae*: the chance to decide for themselves to abandon their criminal enterprises."[64]

Duff's argument is attractive because it vindicates two seemingly inconsistent but widely endorsed ideas about the injustice of criminalizing mere intent, and it shows how these two ideas are in fact compatible. The first idea is that the injustice of criminalizing mere intent springs from the same source as the injustice of criminalizing preparatory conduct. Duff's argument implies that both kinds of criminalization are unjust for the same reason: they deprive would-be criminals of a *locus poenitentiae*, a fair chance to change their minds.

Duff's argument also vindicates a second idea that is seemingly (but not actually) inconsistent with the first: while there's no absolute bar to criminalizing preparatory conduct (the law may criminalize such conduct when the prospective crime is unusually dangerous or difficult to detect), the law must never criminalize mere intent—that's prohibited absolutely.[65] Although Duff doesn't say so, his argument supplies a straightforward explanation for this asymmetry. The bar to criminalizing preparatory conduct is non-absolute because criminalizing preparation doesn't eliminate an intending criminal's opportunity to desist; it merely diminishes it. By contrast, the bar to criminalizing mere intent is absolute because criminalizing mere intent would virtually destroy an intending criminal's *locus poenitentiae*.[66] (This contrast assumes the absence of an abandonment defense for those who renounce their criminal intentions before the law intervenes.)

Despite these virtues, Duff's argument falls short of justifying a categorical ban on punishing mere intent. At the most fundamental level, the argument begs

---

another thing to hold him guilty of a criminal offence at so early a stage in his intended criminal enterprise: that is to treat him as someone who will not be dissuaded, or dissuade himself, from carrying the wrong through.").

63. DUFF, *supra* note 42, at 390.

64. *Id.* at 387.

65. *Cf. id.* at 388 ("[T]o say that the law should treat its citizens as responsible agents is to assert freedom as a categorical limit which should be *respected*, rather than as a consequential good which should be *maximized*.").

66. More precisely, criminalizing mere intent destroys an intending criminal's *locus poenitentiae* as regards crimes of mere intent. As regards crimes of conduct, the intending criminal still may have time to repent.

the question. Even if it succeeds in showing that we must always give the benefit of the doubt to those who intend to commit (but haven't yet begun to commit) a punishable wrong—and this is debatable, for reasons I'll explore below—the argument still doesn't show that an intention to commit a punishable wrong can never be a punishable wrong itself. Duff's basic idea is this:

> If the state is to treat its citizens as responsible agents who can be guided (who can guide themselves) by reasons, it should be slow to coerce them on the ground that they are likely to commit a wrong if not thus coerced . . . . We cannot wait until [an intending criminal] has completed his enterprise: but we should wait until he has more definitively constituted himself as a wrongdoer by coming closer to completing his plan.[67]

The problem with this claim is that it presupposes without justification that one who "intends to commit, or is preparing to commit or taking initial steps towards committing, a substantive crime"[68] hasn't already committed a punishable wrong *simply by preparing or intending*. Duff evidently assumes that a person's malign intention is a candidate for punishment only insofar as the intention constitutes an incipient *attempt* to commit a "substantive crime." But why assume that no intention is ever dangerous or wrongful enough to qualify as a punishable wrong in itself? As I've argued, certain unexecuted intentions seemingly possess all of the characteristics of paradigmatic punishable actions: they are dangerous, wrongful, culpable, provable, and subject to a person's voluntary control.

Duff has suggested to me[69] that he could refine his argument by saying that the state disrespects a person's moral autonomy if it intervenes punitively before the person has committed a "primary wrong." In essence, the state should not intervene at a point when the person's supposed transgression is wrongful only derivatively—only insofar as the transgression aims at some consummate wrong that the person has not yet begun to commit.

Earlier, I raised doubts about whether mere intentions are wrongful only in this derivative sense: as I argued in Part I, lethal intentions seem *non-derivatively* wrongful insofar as they culpably create a risk of death. But even if this were not so—even if mere intentions were wrongful only insofar as they aimed at "primary wrongs"—the refined version of Duff's argument still would leave a fundamental matter unexplained. It would leave unexplained why respecting an ill-intentioned person's freedom as a responsible and autonomous agent requires that we seek to modify his conduct only "by appealing to the moral reasons for which he ought to obey the law; or perhaps by reminding him of the prudential

67. DUFF, *supra* note 8, at 104-05.

68. DUFF, *supra* note 42, at 388.

69. Email from R.A. Duff to author (June 29, 2017) (on file with author).

reasons for desistance provided by the threat of punishment for the substantive crime [i.e., the primary wrong]."[70] At this early stage in the intending criminal's project, Duff observes, "there [still] is logical space for the thought that [the intending criminal] might yet . . . abandon[] the attempt voluntarily."[71] But regardless of whether such logical space exists, it isn't obvious that we respect an intending criminal as a responsible and autonomous agent only if we do nothing more aggressive than seek to dissuade him from moving forward with his plan.[72] If a person can be self-governing even when irredeemably committed to a criminal enterprise,[73] then it isn't at all clear that we disrespect a person's autonomy when we intervene to stop him from pursuing a criminal plan that we're certain he won't reconsider.

To intervene on the expectation that he'll follow through on his criminal choice is arguably to show *respect* for his capacity for self-government. To intervene on this ground is to act on the assumption that he's a person whose intentions do what a person's intentions are supposed to do: ensure that his conduct

---

70. DUFF, *supra* note 42, at 389.

71. *Id.* at 358.

72. *See* Douglas Husak, *Attempts and the Philosophical Foundations of Criminal Liability*, 8 CRIM. L.F. 293, 307 (1997) (reviewing R.A. DUFF, CRIMINAL ATTEMPTS (1996)) ("The mere logical possibility of a change of heart, when unaccompanied by any empirical likelihood, does not seem to me to require that the criminal law can offer no realistic protection for [the would-be victim of an intending criminal]. Friends of [the intending criminal] may have tried for years to persuade him to renounce his [violent] plan—to no avail. Yet Duff would insist that our respect for [the intending criminal's] autonomy as an agent who is susceptible to rational persuasion continues to be owed him—until the moment at which he reaches for his gun. Notice that even at this late point there is 'logical space' for a change of heart. Why does our respect for [the intending criminal's] autonomy become consistent with punishment only at *this* time? Clearly, we would need a more detailed account of responsible agency if Duff's conclusion is to be established.").

73. A person is autonomous if he is self-governing. *See* Sarah Buss, *Personal Autonomy*, *in* THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY (Edward N. Zalta ed., 2016), http://plato.stanford.edu/entries/personal-autonomy [http://perma.cc/HS3P-UQP3]. Only on the moralistic conception of autonomy sometimes attributed to Kant does self-governance require action in accordance with the dictates of morality. *See, e.g.*, Michael S. Moore, *Liberty's Constraints on What Should Be Made Criminal*, *in* CRIMINALIZATION: THE POLITICAL MORALITY OF THE CRIMINAL LAW 182, 187 (R.A. Duff et al. eds., 2014) (describing the Kantian conception of autonomy as one according to which "an autonomous action is a right act done for a right reason"). On most other conceptions of autonomy, a person can act autonomously—can exhibit self-government—even if he acts immorally, indeed, even if he lacks the capacity to appreciate the difference between right and wrong. *See* Buss, *supra*.

is coherent over time by settling practical questions and minimizing reconsideration.[74] If, despite the firmness of his present intention, we refrain from intervening because we think he might reconsider, that's in a sense to *disparage* his autonomy. Duff's argument therefore leaves unexplained how respecting autonomy requires affording a *locus poenitentiae* to someone who by all appearances has made a firm commitment to pursue a criminal enterprise.

Considerations of autonomy aside, Duff's argument in fact leaves unexplained how *any* capacity of human beings—whether properly called autonomy or responsible agency or rationality—requires that we give intending criminals the benefit of the doubt in the way that Duff claims we must. It's one thing to say that respect for a person requires that we be charitable when predicting his conduct—that we proceed on the assumption that he's minimally susceptible to the force of moral reasons, rather than completely beyond the reach of appeals to conscience. If that's what respect for a person requires, then we should be slow to conclude that a person is hell-bent on an evil path even when the evidence strongly indicates that he is. But when the evidence is unambiguous, continuing to give him the benefit of the doubt doesn't seem respectful. It seems naïve.

Duff or a defender might respond that, no matter how strong the evidence that a person is hell-bent on an evil path, if he hasn't yet begun to commit a "primary wrong" but has only aimed at one, we'll disrespect him unless we're slow to treat what he's already done as criminal.[75] The problem with this response is that it's unlikely to justify a *categorical* ban on criminalizing mere intentions.[76] If respect for persons did justify such a ban, it seemingly would justify a categorical ban on punishing *all* derivative wrongs, including ordinary criminal attempts, which are paradigmatically punishable.[77] So respect for persons almost certainly doesn't require a categorical ban on criminalizing mere intentions. At most, it affords a consideration that *weighs against* criminalizing such intentions—a consideration that countervailing reasons might outweigh or defeat.[78]

---

74. *See* MICHAEL E. BRATMAN, INTENTION, PLANS, AND PRACTICAL REASON 20 (1987) ("Intentions are conduct-controlling pro-attitudes, ones which we are disposed to retain without reconsideration, and which play a significant role as inputs into reasoning to yet further intentions.").

75. Email from R.A. Duff to author, *supra* note 69. Duff suggested this response to me, as well as the rejoinder.

76. *Id.*

77. We could still criminalize attempts on the ground that they're non-derivatively wrongful insofar as they *risk* harm. But this very reasoning would also justify criminalizing mere intentions, as I argued in Part I.

78. I presume that the reasons that justify criminalizing ordinary attempts outweigh respect-based considerations decisively, or else the criminalization of attempts would be far more controversial than it is.

III. MENTAL IMMUNITY AND FREEDOM OF MIND

We've yet to uncover a principled basis for the idea that punishing thought is categorically impermissible. So it remains a mystery what commentators are actually describing when they speak of "the inviolability of thoughts"[79] or when they call punishment for mere mental states a "monstrous"[80] intrusion into a person's "private world"[81] and an invasion of her "essential . . . human right to freedom of thought."[82]

I aim in what follows to mine the foundations of this rhetoric and lay bare the premises of an argument of my own. The argument gives analytical clarity to the attractive but heretofore unexplained idea that thought's immunity from punishment relates to a principle of freedom of mind. Although I hope to render the argument's premises plausible, my primary objective is to show that our legal order presupposes these premises, and thus to explain why the conclusion they entail seems so intuitive.

A. The Basic Idea

Given how often and how fervently theorists associate the ban on thought crime with a principle of freedom of mind,[83] it's somewhat surprising that no one has bothered to show how the second principle might undergird the first. Theorists may think the linkage is just obvious. Or they may assume there is so little conceptual space between the two principles that any demonstration of the linkage would be uninteresting. As we'll see, the linkage is both interesting and unobvious.

---

79. Dan-Cohen, *supra* note 48, at 379.

80. Yaffe, *supra* note 49, at 101.

81. Ashworth, *supra* note 50, at 133-34.

82. Calvert, *supra* note 51, at 125.

83. *See supra* notes 47-51 and accompanying text; *see also* Anthony M. Dillof, *Punishing Bias: An Examination of the Theoretical Foundations of Bias Crime Statutes*, 91 NW. U. L. REV. 1015, 1016-20 (1997) (explaining the puzzle of bias crimes given the principle of freedom of thought); Markus Dirk Dubber, *Toward a Constitutional Law of Crime and Punishment*, 55 HASTINGS L.J. 509, 554 (2004) (contending that the principle of freedom of thought prevents the state from punishing nonactualized thoughts); Adil Ahmad Haque, Lawrence v. Texas *and the Limits of the Criminal Law*, 42 HARV. C.R.-C.L. L. REV. 1, 37 (2007) (arguing that there is no right to have others believe or voice "only appropriate attitudes toward" you, as that would violate their freedom of thought and expression); Adam J. Kolber, *Two Views of First Amendment Thought Privacy*, 18 U. PA. J. CONST. L. 1381, 1399 (2016) (arguing that "a plausible case can be made that the First Amendment prohibits pure thought crimes" because the Amendment is generally understood to protect freedom of thought).

In brief, I propose that the injustice of punishment for mere mental states takes its character from the injustice of a more literal breach of the "inviolability of thoughts": namely, a direct and forcible intrusion into the mind.

This more literal breach of the "inviolability of thoughts" is the sort of intrusion that the state would perpetrate if it exposed you to a mind-altering drug in order to disrupt your criminal intentions. It's natural to suppose that this sort of direct and forcible mind control is unjust insofar as it violates your *right of mental integrity*, your right to be free from unwanted mental interference or manipulation. I'll say more about the contours and limits of this right in a later Section. For now, an example will convey the basic idea. Suppose you're an intending criminal. Without invading your right to mental integrity, the government may question you about your criminal intention, try to persuade you to abandon it, surveil you, tail you, and stand ready to thwart you if you attempt to carry your intention out. But the government will invade your right to mental integrity if it causes you to abandon your intention by forcing you to ingest mind-altering drugs, by exposing you to psychotropic gas, or by employing some other form of forcible mind control.

To be sure, many of these intrusions also may invade your right to *bodily* integrity.[84] Forcing you to ingest or inhale an unwanted substance is a classic battery. But if you possess a right to mental integrity, none of these actions is just a battery. Each is also an attempt at forcible mind control, which is a distinctive rights invasion.[85] It's this rights invasion that forms the gravamen of the wrong that the state perpetrates when it forces you to ingest or inhale something mind-altering—the physical battery being slight and potentially harmless. If the government could control your mind without battering you at all (say, by using light and sound to hypnotize you involuntarily), the intrusion still would wrong you, and it would wrong you because it would violate your right to mental integrity.

---

84. *See* United States v. Brandon, 158 F.3d 947, 953 (6th Cir. 1998) ("[T]he issue of forced medication implicates [the defendant's] Fifth Amendment liberty interest in being free from bodily intrusion."); Woodland v. Angus, 820 F. Supp. 1497, 1513 (D. Utah 1993) (finding "that the forcible administration of psychotropic drugs presents a substantial intrusion on plaintiff's liberty interest and an extensive encroachment on plaintiff's bodily integrity"); Khiem v. United States, 612 A.2d 160, 165 (D.C. 1992) (describing forced medication as an "intru[sion] upon [the defendant's] bodily integrity"); *see also* Washington v. Glucksberg, 521 U.S. 702, 720 (1997) (recognizing a right to "bodily integrity"); MARTHA C. NUSSBAUM, WOMEN AND HUMAN DEVELOPMENT: THE CAPABILITIES APPROACH 78 (2000) (listing bodily integrity as one of ten "central human functional capabilities"); LAURENCE TRIBE, AMERICAN CONSTITUTIONAL LAW 1329-62 (2d ed. 1988) (offering an overview of cases relating to government intrusion into the body). *See generally* THE RIGHT TO BODILY INTEGRITY (A.M. Viens ed., 2014) (collecting articles on topics relating to bodily integrity).

85. *See, e.g.*, Bee v. Greaves, 744 F.2d 1387, 1394 (10th Cir. 1984) ("Antipsychotic drugs have the capacity to severely and even permanently affect an individual's ability to think and communicate.").

The claim I'll defend over the next two Sections is that punishment for mere mental states is intrinsically unjust because it's a form of *indirect* mind control.[86]

Not only does this claim promise to give content to the picturesque but imprecise assertion that punishment for mere mental states transgresses the "inviolability of thoughts," but it also captures the essence of relevant American legal doctrine. Consider *Stanley v. Georgia*[87] and *Ashcroft v. Free Speech Coalition*,[88] two well-known cases in which the Supreme Court cited a constitutional prohibition on mind control to justify striking down statutes the enforcement of which had no direct effect on a person's mind. In *Stanley*, the Supreme Court struck down a state statute "forbidding mere private possession of [obscene] material."[89] The Court rejected the government's claim to a "right to control the moral content of a person's thoughts,"[90] noting that "[o]ur whole constitutional heritage rebels at the thought of giving government the power to control men's minds."[91] Decades later, in *Free Speech Coalition*,[92] the Court gave the same justification for striking down a federal statute prohibiting visual depictions of "an actor [who] 'appears to be' a minor engaging in 'actual or simulated . . . sexual intercourse.'"[93] The Court in *Free Speech Coalition* had to distinguish an earlier decision in which it had permitted the government to ban pornography involving real children on account of the harm done to the children depicted.[94] Unlike real child pornography, explained the Court in *Free Speech Coalition*, *simulated* child pornography is anathema for one reason alone: its effect on a viewer's mind. The Court deemed this reason an impermissible basis for criminal legislation. "The [g]overnment submits . . . that virtual child pornography whets the appetites of pedophiles and encourages them to engage in illegal conduct. This rationale cannot sustain the provision in question. The mere tendency of speech to encourage unlawful acts is not a sufficient reason for banning it."[95] Quoting *Stanley*, the Court concluded:

---

86. *Cf.* LAURENCE TRIBE, AMERICAN CONSTITUTIONAL LAW 899 (1st ed. 1978) ("In a society whose 'whole constitutional heritage rebels at the thought of giving government the power to control men's minds,' the governing institutions, and especially the courts, must not only reject direct attempts to exercise forbidden domination over mental processes; they must strictly examine as well oblique intrusions likely to produce, or designed to produce, the same result." (quoting Stanley v. Georgia, 394 U.S. 557, 565 (1969))).

87. 394 U.S. 557 (1969).

88. 535 U.S. 234 (2002).

89. 394 U.S. at 564.

90. *Id.* at 565.

91. *Id.*

92. 535 U.S. at 241.

93. *Id.* (quoting 18 U.S.C. § 2256(2) (2012)).

94. *Id.* at 240 (distinguishing New York v. Ferber, 458 U.S. 747 (1982)).

95. *Id.* at 253.

"The government 'cannot constitutionally premise legislation on the desirability of controlling a person's private thoughts.'"[96] In *Free Speech Coalition*, as in *Stanley*, the Court based its analysis on a constitutional prohibition on mind control even though the statute it found unconstitutional did not affect the mind directly: enforcing statutory bans on obscenity and simulated child pornography is a far cry from administering unwanted mind-altering drugs. The Court's position seems to have been that, because forcible mind control is impermissible, so too are certain governmental efforts designed to achieve the same end by indirect means.

The indirect method of mind control that the Court deemed impermissible in *Stanley* and *Free Speech Coalition* was the state's practice of punishing people for conduct believed likely to produce undesirable thoughts. A more blatant method of indirect mind control, which I presume the Court would disapprove of for the same reason, is the practice of punishing people for their undesirable thoughts themselves. The basic idea is easy to state: it's *because* the state mustn't control thoughts that the state mustn't punish them.

In what follows, I'll show how this idea follows from two interlocking propositions presupposed by our legal order—propositions that I won't be able to defend fully, but that I'll do my best to render plausible. The first proposition—the *Enforceability Constraint*—is that it's wrong for the state to punish offenses of a given type if it's always wrong in principle for the state to forcibly disrupt such offenses merely on the ground that they're censurable transgressions. The second proposition—grounded in the right of mental integrity—is that it's always wrong in principle for the state to forcibly disrupt a given mental state merely on the ground that it's a censurable transgression (although the state sometimes may disrupt a mental state on more exigent grounds). I'll defend these propositions in turn.

B. The Enforceability Constraint

In our system of criminal administration, the state may ensure compliance with penal norms not only indirectly through punishment, but also through direct compulsive force. When you're selling loose cigarettes, the police may take them from your hand. When you're making a bomb, the police may escort you from your laboratory. When you're absconding with stolen goods, the police may stop you and seize them.

An unexamined but signal feature of our system is that the direct and indirect enforcement authorities are linked in a particular way: in practice, and seemingly not by accident, the state may enforce a given penal norm indirectly only when

96. *Id.* at 253 (quoting 394 U.S. 557, 566 (1969)).

it also may enforce that norm directly. In other words, the state may punish someone for transgressions of a given type only when the state may in principle use reasonable force to thwart such transgressions merely on the ground that they're criminally wrongful, that is, without supplying any additional justification. Often the state will have some additional justification for thwarting a transgression—to protect the public, for example, or to arrest a suspect. But no such justification is required. For the state to be justified in disrupting a suspect's conduct, it's enough if the conduct is criminally wrongful. Inversely, if the state may not even in principle use force to thwart instances of a given transgression on the ground that they're criminally wrongful, then the state also may not make that type of transgression an object of punishment.

Why the state ever has the authority to ensure compliance with penal norms through the use of direct compulsive force is a deep and difficult question that I won't pretend to answer here. It's a question that strangely has received much less theoretical attention than the equally important question of why the state ever has the authority to punish. That the first of these authorities never exists without the second is a fascinating and striking aspect of our system of criminal administration—striking because, with respect to *nonpenal* norms, the two authorities frequently diverge. There are many nonpenal norms with which the state may ensure compliance only indirectly, through the imposition of sanctions.[97] Why, then, may the state ensure compliance with penal norms through the imposition of contemporaneous compulsive force? The explanation may stem from the extraordinary importance of the interests that penal norms serve: if the violation of a legitimate penal norm is by its nature a breach of the social compact so grievous that the state may subject the violator to criminal punishment—the severest form of sanction and censure—then perhaps it stands to reason that the state may disrupt such breaches as they occur.

The question I wish to address here is different: it's why the state may ensure compliance with a given legal norm through punishment only when the state may ensure contemporaneous compliance with that norm through direct compulsive force. My answer, in brief, is this: if ensuring compliance with a given norm through direct compulsive force would violate your rights, so too would ensuring compliance with that norm through the threat and imposition of the severest form of sanction and censure. I'll establish this proposition more firmly by means of an informal conditional proof, starting with the supposition that some supposed transgression is off limits to forcible disruption, and reasoning

97. *See, e.g.*, Guido Calabresi & A. Douglas Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089 (1972) (contrasting "property rules," which a court may enforce directly through injunctions, with "liability rules," which a court may enforce (only) indirectly, typically through awards of monetary damages).

from that supposition to the conclusion that the transgression is off limits to punishment.

Suppose, as our starting point, that the state would wrong you if it forcibly disrupted some supposed transgression of yours, T, merely on the ground that T is a censurable transgression. Suppose, further, that the wrong the state would perpetrate against you if it disrupted your T-ing is a wrong *intrinsic* to the disruption—a wrong that consists at least partly in the disruption of T itself, rather than consisting entirely in the fact (if it is one) that the method of disruption injures you in some other way.

Now, if it's the case that the state would wrong you intrinsically if it disrupted your T-ing merely on the ground that T is a censurable transgression, then there must be some reason *why* this is so. And the reason can't be that the method of disruption injures you in some other way, because we've supposed that the wrong is intrinsic—that it consists at least partly in the disruption of T itself. Why, then, does the state wrong you intrinsically when it disrupts your T-ing merely on the ground that T is a censurable transgression?

One possibility is that T is perfectly innocent and innocuous (like consensual sexual conduct between adults) or is at least less wrongful and less harmful than any censurable transgression that the state legitimately may criminalize. In either case, it follows straightforwardly that the state would wrong you if it punished you for T-ing.

But some transgressions may be immune from disruption on grounds of censurability even though they're wrongful and arguably dangerous. (Certain speech acts fall into this category, and so may certain thoughts, as I'll argue in the next Section. When the state prevents you from performing these speech acts or from thinking these thoughts, the state wrongs you. And it wrongs you intrinsically—which is to say, it wrongs you even if it uses means of prevention so delicate and precise that they cause you no injury.)

Suppose, then, that T is as wrongful and harmful as other censurable transgressions that the state may criminalize, yet the state nevertheless would wrong you intrinsically if it disrupted your T-ing merely on the ground that T is a censurable transgression.

If the state would wrong you intrinsically if it disrupted your T-ing on this ground alone, yet your T-ing is dangerous and wrongful, then a likely explanation—perhaps the only possible explanation—is that you've got a *right* to perform T, a right that the state would violate if it forcibly disrupted your T-ing merely on the ground that T is a censurable transgression.

Now, if the state would violate your right if it forcibly disrupted your T-ing merely on the ground that T is a censurable transgression, then I suggest that the state also would violate your right if it disrupted your T-ing in a particular

indirect fashion: by imposing terrible consequences on you for T-ing, merely on the ground that T is a censurable transgression.

But when the state *punishes* you for T-ing, it thereby imposes terrible consequences on you for T-ing, and it does so on no ground other than that T is a censurable transgression. (Ordinarily, to justify punishing someone, the state need only show that the person committed a criminal wrong.) So we may conclude that when the state punishes you for T-ing, it violates your rights. It wrongs you.

We've arrived at the following conditional claim: whether T is innocent and innocuous or wrongful and dangerous, if the state would wrong you if it forcibly disrupted your T-ing on the ground that T is a censurable transgression (our initial supposition), then so too would the state wrong you if it punished you for T-ing (our conclusion). This conditional claim is none other than the Enforceability Constraint.[98]

Justifying the Enforceability Constraint more fully is beyond the scope of this Essay. My present goal is more modest. It's to show how abnormal it would be to treat any type of transgression as an exception to the Enforceability Constraint. Deeming mental transgressions an exception would yield an anomaly: a type of crime that the state may punish but never forcibly disrupt on grounds of criminality alone.

No such type of crime exists, nor does any recognized limit to the state's enforcement power belie the gist of the Enforceability Constraint. In fact, no recognized limit on the state's enforcement power does more than restrict when, how, or pursuant to what procedures given instances of an offense may be forcibly disrupted.

The most salient limit on the state's enforcement power is the principle of reasonable force.[99] This principle governs *how much* force the state may deploy

---

98. Although I've presented these considerations as an argument for the Enforceability Constraint, they may in fact justify both more and less than the Enforceability Constraint. Insofar as certain forms of what we regard as punishment might fall short of imposing *terrible* consequences on an offender, the argument in the text won't establish that the state is always forbidden to punish what it may not disrupt directly merely on grounds of wrongfulness. Certain "lighter" forms of punishment might still be permissible—just as *nonpenal* sanctions are often permissible even when direct enforcement of the relevant (nonpenal) norm is forbidden, the way it's often permissible to award damages as a sanction for conduct that a court couldn't enjoin and that a plaintiff couldn't lawfully disrupt through self-defensive force. Furthermore, insofar as punishing someone for T-ing is but one way of indirectly violating his right to T, the argument in the text may in fact justify principles *beyond* the Enforceability Constraint, including a principle forbidding the state from preventively but nonpunitively detaining people for T-ing. I return to this possibility in the next Section.

99. *See* Graham v. Connor, 490 U.S. 386, 395 (1989) ("*[A]ll* claims that law enforcement officers have used excessive force—deadly or not—in the course of an arrest . . . should be analyzed under the Fourth Amendment and its 'reasonableness' standard . . . .").

to make someone comply with a given penal norm on a given occasion, not whether such force may be deployed at all. In the typical case, the state may deploy an amount of force sufficient but not greater than necessary to stop the relevant norm-violation. If you're selling loose cigarettes, the police may pull them from your hand, but they may not put you in a choke hold.[100]

Of course circumstances sometimes arise where the amount of force necessary and sufficient to stop a given transgression is unreasonably great. Suppose a narcochemist is manufacturing methamphetamine in a treehouse and the only way the police can stop him is by cutting the tree down, paralyzing him in the process. May the police cut down the tree? Clearly not, and the Enforceability Constraint agrees. What the state may punish, the state in principle may impede—but only with reasonable force. Unreasonable force wrongs the narcochemist.

It wrongs him because he has a right not to be paralyzed absent truly exigent circumstances—not because he has a right to make methamphetamine. And that's important. The Enforceability Constraint permits the state to subject the narcochemist to punishment, even as the principle of reasonable force forbids the state to thwart his meth-making. In a world where no single instance of a given offense is disruptable through reasonable force—a world where every narcochemist operates from a fortified treehouse—the Enforceability Constraint still permits offenders to be punished. The Enforceability Constraint says that an offense is unpunishable if it's always wrong *in principle* to disrupt instances of that offense merely on grounds of wrongfulness. In a world of fortified treehouse meth labs, it's always wrong to disrupt meth-making in practice, but it isn't always (or perhaps ever) wrong to do so in principle.

Other limits to the state's enforcement power concern *when* and pursuant to *what procedures* the state may use force to stop a given transgression.[101] Like the principle of reasonable force, these limits are fully consistent with the Enforceability Constraint.

Consider the First Amendment doctrine of prior restraint, which holds that certain expressive acts that are punishable after the fact may not be blocked in advance by a judicial order or administrative ruling.[102] The doctrine's primary rationales are evidentiary and institutional. "It is always difficult to know in ad-

100. Shakeer Rahman & Sam Barr, Opinion, *Eric Garner and the Legal Rules that Enable Police Violence*, N.Y. TIMES (Dec. 5, 2014), http://www.nytimes.com/2014/12/06/opinion/eric-garner-and-the-legal-rules-that-enable-police-violence.html [http://perma.cc/43AC-X75R].

101. *See, e.g.*, Terry v. Ohio, 392 U.S. 1, 16-31 (1968) (describing circumstances in which police may conduct brief investigative detentions and delimiting the scope of the resultant searches).

102. *See* Near v. Minnesota, 283 U.S. 697, 713-14 (1931).

vance what an individual will say," the Supreme Court notes, "and the line between legitimate and illegitimate speech is often so finely drawn that the risks of freewheeling censorship are formidable."[103] Moreover, as the Court observes elsewhere, "[a] criminal penalty . . . is subject to the whole panoply of protections afforded by deferring the impact of the judgment until all avenues of appellate review have been exhausted. . . . A prior restraint, by contrast and by definition, has an immediate and irreversible sanction."[104] If the Court is correct, these evidentiary and institutional considerations support the view that norms prohibiting certain types of speech may not be enforced at particular times (e.g., prior to a jury trial) or in particular ways (e.g., by a bureaucrat's edict).

What these considerations don't support (and have never been interpreted as entailing) is the view that certain penal norms may not be enforced at all except by criminal punishment. It's widely accepted, for example, that an expressive act immune from pretrial injunction may be blocked by a judicial order once the act has been formally adjudicated as unlawful. As the California Supreme Court explains, "[p]rohibiting a person from making a statement or publishing a writing *before* that statement is spoken or the writing is published is far different from prohibiting a defendant from *repeating* a statement or *republishing* a writing that has been determined at trial to be defamatory and, thus, unlawful."[105] The doctrine of prior restraint therefore isn't a counterexample to the Enforceability Constraint; to the contrary, it assumes the Constraint's soundness. The doctrine maintains only that criminal norms prohibiting speech acts are unenforceable at certain times and pursuant to certain procedures. The doctrine doesn't maintain that these norms are unenforceable in principle.

Now, what's enforceable in principle might not always be justifiably enforced in practice. It's conceivable that the above-mentioned limits on the state's enforcement power, if applied to penal norms prohibiting mere thought, would render such norms practically unenforceable except by retrospective criminal punishment. For one thing, it's possible that any direct effort by the state to disrupt the commission of a purely mental transgression would flout limits of timing and procedure. Given the relative inscrutability of the mind, in the absence of a judicial inquest the risks of erroneous intrusion might be too great to bear.[106] It's also possible that any amount of force would be excessive if deployed to disrupt a person's mere mental states. Given the crude technologies of mind control currently available, forcible intrusion into the mind might inevitably cause serious physical injuries or deleterious changes to a person's personality or mental

103. Se. Promotions, Ltd. v. Conrad, 420 U.S. 546, 558-59 (1975).

104. Neb. Press Ass'n v. Stuart, 427 U.S. 539, 559 (1976).

105. Balboa Island Vill. Inn, Inc. v. Lemen, 156 P.3d 339, 344-45 (Cal. 2007).

106. *See* MOORE, *supra* note 41, at 48.

well-being. Even if all these things are true, however, limits of timing, procedure, and proportionality still don't entail that mental intrusion is objectionable *in principle*. They don't entail that mental intrusion would be objectionable even if it could be carried out flawlessly: by a device that could detect malevolent intentions with high reliability and psycho-surgically remove them without doing other damage.

If such intrusion isn't objectionable in principle, then the Enforceability Constraint doesn't yield the conclusion that punishing thought is intrinsically unjust. So the question is whether psycho-surgical policing is actually objectionable in principle. May the state thwart your mental states merely on the ground that they're censurable transgressions?

C. The Right of Mental Integrity

My contention is that psycho-surgical policing is indeed objectionable in principle, and it's objectionable in principle because it violates the right to mental integrity, the right to be free from unwanted mental interference or manipulation of a direct and forcible sort.

A commitment to this right, like a commitment to the Enforceability Constraint, seems a basic feature of our system of criminal administration. The right to mental integrity figures not only in the reasoning of *Stanley* and *Free Speech Coalition* but also in the legal principles governing when the state may forcibly medicate a defendant to render him competent to stand trial[107] and when the state may forcibly medicate a mentally ill prisoner to ensure public safety.[108] The right to mental integrity also applies in a decidedly nonpenal context, undergirding a civilly-committed person's right to refuse involuntary psychiatric treatment. As one court explained, "[t]he [constitutional] right of privacy is broad enough to include the right to protect one's mental processes from governmental

107. *See* Sell v. United States, 539 U.S. 166, 179 (2003) (holding that "[the state may forcibly] administer antipsychotic drugs to a mentally ill defendant facing serious criminal charges in order to render that defendant competent to stand trial, but only if the treatment is medically appropriate, is substantially unlikely to have side effects that may undermine the fairness of the trial, and, taking account of less intrusive alternatives, is necessary significantly to further important governmental trial-related interests"); Riggins v. Nevada, 504 U.S. 127 (1992) (deeming it a violation of due process to forcibly administer antipsychotic drugs to a defendant during trial where the trial court hadn't found that less intrusive alternatives were unavailable and that the medication was both medically appropriate and necessary for the defendant's or others' safety).

108. *See* Washington v. Harper, 494 U.S. 210 (1990) (holding that the state may forcibly administer antipsychotic medication to a seriously mentally ill prisoner if the state first establishes that he is dangerous to himself or others and that such treatment is in his medical interest).

interference."[109] Legal principles aside, we generally blanch at the idea of brainwashing—the idea of one person controlling the thoughts of another through forcible conditioning—whether the controller is a cult leader or a totalitarian government.

The main obstacle to appreciating that our legal and moral order presupposes a right to mental integrity is the mistaken view that, if such a right existed, it would be unqualified or absolute. If the right to mental integrity were absolute, forcible manipulation of a person's mind would be absolutely forbidden. But forcible manipulation of a person's mind doesn't seem absolutely forbidden. For example, it might be permissible for the state to force a mentally ill prisoner to ingest psychiatric medication, as the Supreme Court recognized in *Washington v. Harper*.[110] If this sort of mental intrusion is justifiable, that might be thought to entail that there's no right to mental integrity after all—no right to be free from forcible mind control. But the justifiability of mental intrusion entails merely that the right to mental integrity, if it exists, is qualified or non-absolute, which is to say, the right can be invaded justifiably.[111]

In fact, the Court's willingness to permit forced medication in *Harper* actually seems to rest on an acknowledgment that people possess a *qualified* right to mental integrity rather than on a denial that any such right exists. In *Harper*, a mentally ill prisoner claimed that the state should be barred from forcing him to ingest antipsychotic drugs unless it could prove that he would consent to such

---

109. *See* Rennie v. Klein, 462 F. Supp. 1131, 1144 (D.N.J. 1978).

110. 494 U.S. 210 (1990).

111. That rights can be invaded justifiably helps illuminate the otherwise perplexing moral structure of many ordinary transactions. If I moor my boat to your dock to save it from being destroyed by a storm, I unquestionably act permissibly, even if I act against your wishes. *See* Vincent v. Lake Erie Transp. Co., 124 N.W. 221 (Minn. 1910). A possible explanation of why my act is permissible is that necessity temporarily extinguishes your property right in the dock. But this explanation makes it mysterious why I owe you compensation for the damage your dock sustains after the storm throws my boat against it. (How could I be obliged to compensate you for the loss of something to which you had no right?) A better explanation is that your property right endures throughout my necessary trespass, the right making a persistent claim on me even as I justifiably infringe it. That would explain why I must take reasonable measures during the storm to minimize the damage, and why I must compensate you afterward for the damage I couldn't avoid. Although I acted permissibly in mooring my boat to your dock, by leaving it there I infringed your persisting right. My analysis here borrows from Joel Feinberg's discussion of the "backpacker" case. *See* Joel Feinberg, *Voluntary Euthanasia and the Inalienable Right to Life*, 7 PHIL. & PUB. AFF. 93, 102 (1978); *see also* Judith Jarvis Thomson, Self-Defense and Rights (Apr. 5, 1976), *in* THE LINDLEY LECTURE 3, 10 (1977) ("[W]e *violate* [a person's] right if and only if we do not merely infringe his right, but more, are acting wrongly, unjustly in doing so."). *But see* John Oberdiek, *Lost in Moral Space: On the Infringing/Violating Distinction and Its Place in the Theory of Rights*, 23 LAW & PHIL. 325, 327 (2004) (arguing "against incorporating the infringing/violating distinction into a theory of rights").

treatment if competent.[112] The Supreme Court denied the prisoner's claim, holding that the state may force a seriously mentally ill prisoner to ingest antipsychotic medication against his will as long as the state first establishes that he's "dangerous to himself or others" and that such treatment is in his "medical interest."[113] If this holding is correct—as a matter of political philosophy, whether or not as a matter of constitutional law—then the government doesn't *violate* (i.e., unjustifiably invade)[114] an inmate's right to mental integrity by interfering directly with his thoughts if doing so is practically necessary to ensure public safety and is in the person's "medical interest." It doesn't follow, however, that the proposed right of mental integrity is illusory. Nor does it follow that public necessity temporarily extinguishes the inmate's right to mental integrity, such that the right exerts no moral force in the covered circumstance. Rather, the best explanation of the Court's holding is that public necessity *overrides* the inmate's right without extinguishing it. If the right persists even when justifiably overridden, then the right continues to exert moral force. That explains why the unwanted psychiatric intervention must end as soon as possible, why the intervention must be no more intrusive than necessary to serve its purpose,[115] and why the very question of the intervention's permissibility is so momentous in the first place.

As my analysis of *Harper* shows, we can allow that the state may manipulate your mental states on grounds of public necessity without thereby denying the existence of a right to mental integrity. Just as important, we can allow that the state may manipulate your mental states on grounds of public necessity without thereby conceding that the state may infringe your right to mental integrity on grounds *other* than public necessity—such as the ground that the targeted mental state is a censurable transgression, a ground on which (per the Enforceability Constraint) the state would have to be allowed to invade the right if it were allowed to make mere thought an object of punishment.

Public necessity may justify many kinds of rights invasion that would be impermissible if undertaken on other grounds. For example, the state may subject you to excruciating pain as a way of preventing you from killing someone, but not as a way of punishing you for a criminal offense. Your right not to be subjected to excruciating pain prohibits the state from performing certain actions for certain reasons without forbidding the state from performing those actions altogether. Thus, your right not to be subjected to excruciating pain forbids the

---

112. *Harper*, 494 U.S. at 222.

113. *Id.* at 227.

114. *See* Thomson, *supra* note 111, at 10.

115. *See* Sell v. United States, 539 U.S. 166, 179 (2003) (holding that the state must prove forced medication to be necessary after "taking account of less intrusive alternatives").

state from causing you excruciating pain on the ground that doing so will serve as an unpleasant sanction that expresses the state's disapproval of your past wrongdoing (punishment)—but the state violates no right of yours when it subjects you to the exact same measure of excruciating pain on the ground that doing so will make you drop the gun you're threatening to fire at an innocent child (contemporaneous disruption).

Similarly, your right to mental integrity forbids the state from forcibly disrupting your mental states on the ground that they're censurable transgressions—but, if the holding of *Harper* is sound, the state doesn't violate your right to mental integrity when it forcibly disrupts your mental states on the ground that doing so is necessary to protect the public and is in your "medical interest" anyway. Indeed, mental intrusion on grounds of public necessity seems permissible even when it's not in your "medical interest." Imagine that a terrorist intends to detonate a bomb and the police have only three ways of stopping him: they can incapacitate him (e.g., shoot him), restrain him physically (e.g., handcuff him), or restrain him psychically (e.g., deploy a stun grenade). If the police aren't close enough to the terrorist to restrain him physically, they're left with two options: incapacitation and psychical restraint. Because the threat to public safety is grave—and because temporary psychical restraint is a mild invasion of a person's mental integrity, whereas permanent physical incapacitation is a grievous invasion of his bodily integrity—I presume that the government may forcibly disrupt the terrorist's intention (e.g., with a stun grenade) on the ground that doing so is necessary to prevent the terrorist from detonating the bomb (which, I hasten to add, isn't the same as disrupting the intention on the ground that it's a censurable transgression).

In fact, I don't see any barrier in principle to the state *preventively detaining* people on the basis of their thoughts alone. But consider how heavy a burden the state would have to bear in practice if it sought to justify such a measure by appeal to the considerations generally thought necessary to justify direct mental intrusion. To forcibly medicate a prisoner, for example, the state must show that the prisoner is dangerous and that less intrusive alternatives to forced medication are unavailable. If the state could make a similar showing in regard to detaining a person on the basis of a given thought—if it could show that doing so were necessary to protect the public, less intrusive alternatives being unavailable—then I'd be willing to concede that it isn't always wrong to preventively detain people on the basis of that particular thought. I simply doubt whether the state could ever make the requisite showing. It isn't enough for the state to show that certain thoughts present an exceptional danger. It's also necessary for the state to show that the danger can be allayed in one way only: by preventively detaining people on the basis of those thoughts alone. No actual jurisdiction takes the possibility seriously. Several American states have laws permitting the preventive

detention of "sexually violent predators," but these laws require proof of previous violent conduct, rather than mere proclivity.[116]

Yet there's one strain of American law that might seem to lower the barrier to mental intrusion: the doctrine permitting the government to administer involuntary medication *without* a showing of public necessity when the purpose is to render a psychotic defendant fit for trial.[117] Under current Supreme Court precedent, the government may administer involuntary medication for this purpose if "the treatment is medically appropriate, is substantially unlikely to have side effects that may undermine the fairness of the trial, and, taking account of less intrusive alternatives, is necessary significantly to further important governmental trial-related interests."[118] Stephen Morse rationalizes this doctrine on the ground that the state's "interest in adjudicating guilt and innocence and achieving finality in the criminal process is . . . 'essential' or important,"[119] whereas the defendant's interest in freedom from unwanted mental intrusion is minimal under the circumstances. Forcibly medicating a psychotic defendant, Morse argues, "would appear to increase freedom of thought rather than to decrease it. . . .

116. *See, e.g.*, People v. Field, 204 Cal. Rptr. 3d 548, 553 (Ct. App. 2016) ("[T]he [California Sexually Violent Predator Act] provides for indefinite involuntary civil commitment of certain offenders who are found to be [sexually violent predators] following the completion of their prison terms . . . . To establish that a person is [a sexually violent predator], the prosecution is required to prove [beyond a reasonable doubt] the following: (1) *the offender has been convicted of a qualifying sexually violent offense against at least two victims*; (2) the offender has a diagnosed mental disorder; (3) the disorder makes it likely the offender would engage in sexually violent conduct if released; and (4) this sexually violent conduct will be predatory in nature." (emphasis added)).

117. *Sell*, 539 U.S. at 179 ("[T]he Constitution permits the [state] involuntarily to administer antipsychotic drugs to a mentally ill defendant facing serious criminal charges in order to render that defendant competent to stand trial."); *see also* People v. McDuffie, 50 Cal. Rptr. 3d 794, 798 (Ct. App. 2006) (holding that the involuntary administration of psychiatric medication to render a defendant competent to stand trial comports with California's right to privacy if the trial court finds that "[t]he [prosecutor has] charged the defendant with a serious crime against the person or property; involuntary administration of antipsychotic medication is *substantially likely* to render the defendant competent to stand trial; the medication is unlikely to have side effects that interfere with the defendant's ability to understand the nature of the criminal proceedings or to assist counsel in the conduct of a defense in a reasonable manner; less intrusive treatments are unlikely to have substantially the same results; and antipsychotic medication is in the patient's best medical interest in light of his or her medical condition." (quoting CAL. PENAL CODE § 1370, subdiv. (a)(2)(B)(ii)(III) (emphasis added))).

118. *Sell*, 539 U.S. at 179.

119. Stephen Morse, *Involuntary Competence in United States Criminal Law* 17 (U. of Pennsylvania Law Sch., Pub. Law & Legal Theory Research Paper No. 17-20, 2017), http://ssrn.com/abstract=2951966 [http://perma.cc/BUX3-642U].

[T]he 'freedom' to be psychotic does not seem to be a freedom worth having or freedom at all."[120]

If this reasoning and the doctrine it supports are sound, it's natural to ask whether the need to prevent people from having culpably wrongful thoughts couldn't sometimes be at least as pressing as the need to rid defendants of delusions pretrial. I'm not certain that the doctrine is sound, however, so I'm neutral between the following possibilities:

> (1) The need to rid defendants of delusions pretrial is more pressing than the need to prevent people from having culpably wrongful thoughts. Accordingly, although the state may forcibly medicate defendants pretrial, it may not punish people for their thoughts (thanks to the Enforceability Constraint).
>
> (2) The need to rid defendants of delusions pretrial *isn't* more pressing than the need to prevent people from having culpably wrongful thoughts, and each of these needs is insufficient to justify mental intrusion. Accordingly, the state may not forcibly medicate defendants pretrial, nor (thanks to the Enforceability Constraint) may the state punish people for their thoughts.

My claim is simply that (1) is coherent. It nevertheless might be false. The better view might be (2): it might be that mental states are unpunishable only if forcibly medicating defendants pretrial is unjustifiable. This possibility doesn't seem a *reductio ad absurdum* of the proposition that mental states are unpunishable. We shouldn't unquestioningly accept that the government's trial-related interests truly justify infringing the mental autonomy of psychotic defendants.[121]

The one possibility I've rejected is this:

> (3) The need to rid defendants of delusions pretrial isn't more pressing than the need to prevent people from having culpably wrongful thoughts, yet each of these needs is *sufficient* to justify mental intrusion. Accordingly, the state not only may forcibly medicate defendants pretrial but it also may punish people for their thoughts.

I've rejected this possibility out of hand—precipitously, some might say. Although our legal order presupposes a right to mental integrity that applies across

---

120. *Id.* at 15.

121. *See* Aaron R. Dias, *Just Say Yes:* Sell v. United States *and Inadequate Limitations on the Forced Medication of Defendants in Order To Render Competence for Trial*, 55 S.C. L. REV. 517, 518 (2004) (arguing that *Sell* does not "sufficiently protect an individual's bodily integrity").

a range of penal and nonpenal contexts, in many of these contexts the right gives way to competing values. As conceived in law, the right to mental integrity clearly isn't absolute. This raises a basic question. If the right to mental integrity can be overridden on grounds of public necessity, and maybe also on grounds of judicial finality, why can't the right to mental integrity ever be overridden on the ground that it's being exercised wrongfully? If mental intrusion can be justified by the imperatives of public safety and criminal adjudication, why can't it also be justified by the imperative of law enforcement? Why can't the state at least sometimes manipulate a person's mind on the ground that his mental states are censurable transgressions?

I think this line of rhetorical questions gets things backward. Part of what it means to have a right is that any proposed invasion of the sphere that the right protects requires affirmative justification. Absent such justification, we can repel a proposed invasion just by asserting the right. Thus, if there's a right to mental integrity—as our legal order presupposes, and as intuitively seems to be the case—then the question we must ask of any proposed invasion of the right isn't "why *shouldn't* it be permitted?" but "why *should* it?". The burden is on the intruder to justify the intrusion, not on the right-bearer to defeat it.

Now, I don't mean to imply that such justification is unimaginable. We simply know too little about the foundations of either the state's enforcement power or the right to mental integrity to assert confidently that mental intrusion can never be justified merely on the ground that a person's mental states are censurable transgressions. Thus, we can't yet say whether the imperative of law enforcement is more or less compelling than the imperative of criminal adjudication—although I do think we can assume that the countervailing individual interests in the adjudication context are probably somewhat weaker. As Morse suggests, "the 'freedom' to be psychotic [may not] be a freedom worth having or freedom at all."[122]

I also think we can assume that the countervailing individual interests are weaker when the right in question is that of *bodily* integrity. I've assumed, as everyone does, that the right of bodily integrity routinely gives way to the imperative of law enforcement: that proposed invasions of the right to bodily integrity can be justified on the mere ground that the right-bearer is committing a censurable transgression. The police may take loose cigarettes from your hand, escort you from your bomb-making laboratory, and seize your stolen goods—all without violating your right to bodily integrity.

But why? If, as I've said, the burden is always on the potential right-intruder to justify an intrusion, not on the right-bearer to defeat it, then why does the imperative of law enforcement—the state's imperative to disrupt censurable

122. Morse, *supra* note 119, at 15.

transgressions merely on the ground that they're censurable transgressions — justify invading the body if it doesn't justify invading the mind?

To this important question I can offer only the beginning of an answer. My suspicion is that the right to mental integrity may derive (in a way that the right to bodily integrity does not) from the nature and moral significance of personhood. At the root of the normative asymmetry between mind and body may be the fact that one's mental states, far more than one's actions, determine who one is as a person. As Seana Shiffrin writes, "what makes one a distinctive individual *qua person* is *largely* a matter of the contents of one's mind."[123] Thus, if one has an interest in controlling one's identity as a "distinctive individual" — an interest in controlling who one is as a person — then one has an interest in controlling the contents of one's mind. I assume that this fundamental interest grounds the right to mental integrity, and that this right, unlike the right to bodily integrity, therefore serves as a decisive counterweight to the imperative of law enforcement.

In making these assumptions — in assuming that the state necessarily violates your right to mental integrity when it forcibly disrupts your thoughts on the ground that they're censurable transgressions — I've not simply assumed what I set out to prove: that thought is unpunishable. Grounding thought's immunity from punishment in its immunity from direct manipulation has required me to defend an unexamined but signal feature of our system of criminal administration: that the state's authority to punish transgressions of a given type extends no further than its authority to disrupt transgressions of that type using direct compulsive force. If sound, the Enforceability Constraint isn't a conceptual or semantic truth; it's a normative one. And it's a normative truth that doesn't hold for nonpenal law, where retrospective sanction is often permissible even when contemporaneous compulsion is not.

***

I've argued in this Part that the intrinsic injustice of punishment for thought has the following origins:

> (1) It's wrong for the state to punish you for your thoughts if it's always wrong in principle for the state to use force to thwart or disrupt your thoughts merely on the ground that they're censurable transgressions.
>
> (2) It's always wrong in principle for the state to use force to thwart or disrupt your thoughts merely on the ground that they're censurable transgressions.

---

123. Seana Valentine Shiffrin, *A Thinker-Based Approach to Freedom of Speech*, 27 CONST. COMMENT. 283, 291 (2011).

(3) Therefore, it's wrong for the state to punish you for your thoughts.

The first of these propositions draws support from the Enforceability Constraint, and the second from the right to mental integrity—two ideas to which our legal order seems resolutely committed. In explaining these commitments, I did my best to make both seem reasonable. I didn't pretend to offer a full justification of either. It's unlikely that any such justification would be beyond controversy, anyway. It would be surprising indeed if a somewhat controversial proposition—that there's an *intrinsic* injustice in punishment for mere mental states—followed straightforwardly from propositions that were themselves uncontentious.

CONCLUDING REMARKS

The state's enforcement power and the mind's inviolability are rich topics worthy of further inquiry. Especially ripe for study is their point of intersection. Positing a right to mental integrity raises difficult questions about the limits of the state's enforcement power, foremost among them the question of the right's precise scope vis-à-vis the state.

It can't be that the state violates your right to mental integrity every time it tries to influence your thoughts. The state violates no one's right to mental integrity when it pleads with a hostage taker, requires children to be educated, or simply attempts to communicate with its citizens. A police officer doesn't violate your right to mental integrity when she approaches you and begins talking, even though by doing so she causes you to experience certain perceptions and beliefs that you might not want to experience.

As these examples show, distinguishing between permissible and impermissible modes of interference with a person's mental life presents no small task. Why does the police officer's communicative act not violate your right to be free from unwanted mental intrusion? Is it because the means of interference (stimulating your perceptive faculties) isn't forcible? Is it because you implicitly consent to this type of mental intrusion just by going around in the world with open eyes and ears? Is it because your right to mental integrity simply doesn't cover perceptions and perceptual beliefs, the right being limited to other sorts of mental state? Or is it because of the purpose for which the intrusion is undertaken?

A complete theory of mental integrity would answer these questions by yielding an analytical framework for distinguishing in a principled way between modes of state interference that respect the right to mental integrity and modes that constitute impermissible mental intrusions. Like any moral or legal right, the right to mental integrity can be analyzed in terms of three aspects: (i) the

domain over which the right ranges; (ii) the type of mental intrusions that qualify as invasions of the right; and (iii) the kind of circumstances (including state motivations) that make an invasion a *violation*, an invasion that's impermissible.

By distinguishing these three aspects of the right to mental integrity, we might begin to make progress on questions like those above.

Why doesn't the state violate your right to mental integrity when a police officer accosts you and asks you questions? Plausibly, the perceptions and perceptual beliefs that the police officer causes you to experience don't fall within the domain over which the right ranges (see (i)).

Why doesn't a liberal state violate a child's right to mental integrity when it compels her to receive an education of one sort or another? A possible answer is that, even though the beliefs and dispositions that a liberal education instills all fall within the domain that the right protects (see (i)), a liberal education engages directly with a child's rational faculties, instead of bypassing those faculties in the fashion of brainwashing or indoctrination. Thus, compulsory education may not qualify as a rights invasion (see (ii)).

Why doesn't the state violate a mentally ill inmate's right to mental integrity when it forces her to ingest psychiatric medication as a means of ensuring community safety? Plausibly, the circumstances and intended effect of the intrusion render the rights invasion permissible (see (iii)).

Each of these tentative answers alludes to some general operating principle that differentiates impermissible mind control from softer modes of influence that leave people's mental integrity tolerably intact. Some such principles *must* exist, or else the state would be altogether forbidden from influencing people's beliefs and desires—an implausible position.

The operating principle that this Essay has aimed to vindicate is the age-old maxim of criminal jurisprudence *cogitationis poenam nemo patitur* ("no one may be punished merely for thinking"). But this operating principle is potentially just one among many.

Now, one absurd operating principle that the argument of this Essay might seem to entail—to its embarrassment—is a prohibition on punishing any crime that *involves* a mental state. If it's impermissible for the state to thwart an executed or partly executed mental state simply on the ground that the mental state is a censurable transgression, then, by virtue of the Enforceability Constraint, it's also impermissible to make someone's executed or partly executed mental state an object of punishment. But ordinary crimes of mens rea might seem to do exactly that: they might seem to make an offender's executed or partly executed mental state an object of punishment, punishing the offender for the combination of a bodily movement and an accompanying mental state. Therefore, it might seem to follow from my view that it's impermissible to punish someone for a crime of mens rea—which is absurd.

Does this chain of inferences demolish the Essay's positive argument? No, because there are at least two ways to block the absurd conclusion.

One is to reconsider the initial premise, that it's impermissible for the state to thwart an executed or partly executed mental state merely on the ground that it's wrongful. But solicitude for the right of mental integrity might make us reluctant to deem even *executed* mental states vulnerable to direct and forcible disruption on grounds of censurability alone. So I suggest instead that we reconsider an intermediate premise: that crimes of mens rea punish people in part for their executed or partly executed mental states.

Crimes of mens rea punish people for their *actions*. To be sure, actions involve mental states essentially. But we needn't assume that punishment for an action is, in part, punishment for the action's constituent mental state—any more than we must assume that punishment for an action is, in part, punishment for the action's constituent bodily movement. If criminal actions are properly conceived as unitary wholes rather than as mere aggregates of bodily movements and accompanying mental states, then punishment for the whole can't be decomposed into separate punishment for each of the whole's constituent parts. Thus, actions may be vulnerable to punishment even as their constituent mental states, taken in themselves, are immune.

As it happens, this conception of criminal actions yields important consequences for the apportionment of punishment, for the scope and limits of the criminal law, and for our understanding of the core requirements of actus reus, mens rea, and concurrence. But these are matters for another day.[124]

124. I develop these ideas and their implications for criminal jurisprudence in Gabriel S. Mendlow, The Unity of Action and the Action as Object (June 1, 2018) (unpublished manuscript) (on file with author).