



THINK AGAIN: THE THOUGHT CRIME DOCTRINE AND THE LIMITS OF CRIMINAL LAW

*Jordan Wallace-Wolf**

According to the thought crime doctrine, neither beliefs nor intentions may be subject to criminal punishment. The doctrine is widely endorsed, but puzzling in its scope. Beliefs have a free speech credential: they play a straightforward role in the sincere exchange of ideas. Moreover, they are harmless, in the specific sense that they do not aim at action and so not at lawbreaking. But intentions are otherwise. They do not necessarily further the exchange of ideas and they may aim at wrongful, illegal conduct.

So why should the thought crime doctrine categorically protect them in addition to beliefs? Why not allow the criminalization of at least some intentions, such as those that aim at criminal wrongdoing? The going answers to this question are unsatisfying because they do not identify any moral reason to protect intentions *per se*.

I argue that there is such a reason based in the importance of thinking. My argument has two premises. The first is that persons have a fundamental moral interest in thinking. The second is that thinking partly consists in believing and intending whatever one takes to be warranted, even if it's not. Hence, persons have a moral interest in their beliefs and intentions, whether they are true and laudable or false and odious. The thought crime doctrine is the sound legal recognition of this interest.

* UCLA School of Law (wallace-wolf@law.ucla.edu). I would like to thank Jane Bambauer, Ashutosh Bhagwat, Daniela Dover, Mark Greenberg, Jeff Helmreich, Barbara Herman, Pamela Hieronymi, Brian Hutler, A.J. Julius, Bill Kowalsky, Gavin Lawrence, Toni Massaro, Richard Re, Larry Sager, Eugene Volokh, participants in the UCLA Philosophy Department's Ethics Workshop, and participants in the UCLA Law School's Legal Writing Workshop. I would like to specially thank Seana Valentine Shiffrin for many helpful comments on this Article. Thanks also to Jamie Butts for her careful editing.

Introduction.....	6
I. Contemporary Thought Crime Scholarship	9
A. Alexander/Ferzan and Duff.....	9
B. Mendlow.....	12
1. The enforceability constraint does not describe current law.....	14
2. The enforceability constraint does not provide a justification for the thought crime doctrine	18
II. A Thinking-Based Argument Against Punishing Beliefs	20
A. Major Premise: Persons Have an Interest in Thinking	20
B. Minor Premise: Punishing Beliefs Interferes with the Interest that Persons Have in Thinking	26
III. Extending The Thinking-Based Argument to Intentions	33
A. Punishing Intentions Also Interferes with the Interest that Persons Have in Thinking.....	33
B. Addressing Objections.....	37
1. The argument proves too much: If intentions may not be punished then neither can actions, which is absurd.....	37
2. The argument proves too little: If actions can be punished then so can intentions	40
3. The argument does not justify the thought crime doctrine: Other interests could outweigh the interest in thinking	42
C. Consequences for Free Speech Doctrine.....	44
Conclusion	49

INTRODUCTION

Imagine that a man sits in a restaurant and writes out, on a paper napkin, a personal resolution of sorts. In it, he enumerates his reasons for, and his consequent harboring of, an intention to attack a particular person with brass knuckles. He leaves without taking the napkin with him, and the server cleaning his table picks it up, reads it, and gives it to the police. There is thus evidence that the author has at least two thoughts with criminal content: He *believes* that he should, all things

considered, use brass knuckles against another, and he *intends* to do so.¹

Nonetheless, he cannot be convicted of any crime, whether for his belief or his intention, because of the “centuries-old maxim” that “no one is punishable solely for his thoughts.”² This venerable doctrine, the *thought crime doctrine*, is widely accepted, but only partially understood.³

As applied to beliefs, the doctrine has been frequently defended in terms of First Amendment values.⁴ According to these arguments, the freedom to believe, free from punishment, is necessary to realize a number of individual and political values. Citation of these First Amendment values is often accompanied by the further observation that beliefs do not aim at changing the world, but only at representing it accurately. Consequently, beliefs are, in a limited but important sense, unthreatening to public order. People may believe what they will, and yet, respect the law.⁵

But the thought crime doctrine also encompasses intentions—understood here as commitments to act, either now or in the future—and their inclusion is both

¹ See Gerald Dworkin & David Blumenfeld, *Punishment for Intentions*, 75 MIND 396, 400 (1966) (imagining the discovery of a diary that reveals a person’s intention).

² *United States v. \$11,500.00 in United States Currency*, 869 F.3d 1062, 1075 (9th Cir. 2017) (collecting sources and discussing the thought crime doctrine at length, with emphasis on intentions). See also *Doe v. City of Lafayette*, 377 F.3d 757 (7th Cir. 2004) (discussing the thought crime doctrine in a situation that involves thoughts without action).

³ ALAN BRUDNER, PUNISHMENT AND FREEDOM: A LIBERAL THEORY OF PENAL JUSTICE 108 (2009) (noting wide agreement that thoughts may not be punished).

⁴ See, e.g., *Nat’l Inst. of Fam. & Life Advocs. v. Becerra*, 138 S. Ct. 2361, 2379 (2018) (Kennedy, J., concurring) (“Freedom of speech secures freedom of thought and belief.”); *Ashcroft v. Free Speech Coal.*, 535 U.S. 234, 253 (2002) (“First Amendment freedoms are most in danger when the government seeks to control thought or to justify its laws for that impermissible end. The right to think is the beginning of freedom, and speech must be protected from the government because speech is the beginning of thought.”); *Stanley v. Georgia*, 394 U.S. 557, 565–66 (1969) (discussing the right to receive various kinds of ideas, but also discussing the lack of a government interest in preventing “antisocial” thoughts); *W. Va. State Bd. of Educ. v. Barnette*, 319 U.S. 624, 641 (1943) (discussing freedom of belief as a necessary precursor for consent to be governed); *Whitney v. California*, 274 U.S. 357, 372–78 (1927) (Brandeis, J., concurring) (discussing belief in an ideology that favors lawbreaking).

⁵ First Amendment precedent draws this distinction in the course of protecting beliefs. See, e.g., *Brandenburg v. Ohio*, 395 U.S. 444, 448 (1969) (per curiam).

much more puzzling and much less discussed.⁶ It is more puzzling because intentions do not have any of the features that are supposed to make beliefs a good candidate for protection under the First Amendment. Intentions to commit crimes, for example, have no role in the exchange of ideas and they motivate criminal acts. Such intentions thus challenge public order to a greater extent than similar beliefs, and without any obvious compensating speech benefits.

Furthermore, intentions like those of the above resolution author seem relevantly similar to many actions that are, apparently, wrongful and justifiably criminalized. For instance, actions that correlate with and enable future wrongdoing may be punished as preparatory crimes, such as the possession of burglary tools or brass knuckles.⁷ Actions that advance or further a wrong are punishable as attempts, such as putting on brass knuckles in order to attack with them.⁸ And actions that aim to get another to commit a crime are punishable as solicitations.⁹ But the resolution author's intention is not dissimilar to these actions. It correlates with and enables an attack with brass knuckles. It also furthers such an attack. Moreover, if it is punishable to try to convince, suggest, or request that someone else intend to use brass knuckles to attack a third party, then why is it permissible for the resolution author to harbor such an intention himself?

These puzzles set the stage for the question addressed by this Article: Is there

⁶ See DOUGLAS HUSAK, *Does Criminal Liability Require an Act?*, in PHILOSOPHY OF CRIMINAL LAW: SELECTED ESSAYS 51 (2010) (“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.”); Frederick Schauer, *On the Distinction Between Speech and Action*, 65 EMORY L.J. 427 (2015) (questioning whether a normatively significant distinction can be made between thoughts, speech, and acts).

Only a few Supreme Court cases deal with the thought crime doctrine as applied to intentions, and those that do offer relatively sparse analysis. See, e.g., *United States v. Shabani*, 513 U.S. 10, 16 (1994) (“[T]he law does not punish criminal thoughts . . .”).

⁷ See, e.g., *Commonwealth v. Hardick*, 380 A.2d 1235 (Pa. 1977) (burglary tools); *Commonwealth v. Jones*, No. 3345OF2001, 2002 WL 34401140 (Pa. Ct. Comm. Pl. Oct. 8, 2002) (brass knuckles); HUSAK, *supra* note 6 (using brass knuckles as an example).

⁸ See 2 WAYNE R. LAFAVE, *SUBSTANTIVE CRIMINAL LAW* § 11.2 (3d ed. 2020) (summarizing attempt doctrine); *United States v. Soto-Barraza*, 947 F.3d 1111, 1114–15 (9th Cir. 2020) (discussing attempt liability principles).

⁹ LAFAVE, *supra* note 8, at § 11.1 (summarizing solicitation doctrine); see also *United States v. White*, 610 F.3d 956, 960 (7th Cir. 2010) (discussing solicitation principles).

any *principled* reason, i.e., one not grounded in doctrinal preference or pragmatism, for why the thought crime doctrine should categorically protect intentions in addition to beliefs, even when the former may have features that warrant criminalization? If there is, it must be because intentions have some redeeming value, just as thoughts. But what is that value? Thought crime scholarship has not squarely answered this question, and in some cases, has not even asked it.

My answer to it is based in the distinct value that thinking has for persons coupled with the distinct role that thoughts play in that activity. Thinking is valuable because it is how persons encounter questions for themselves and make themselves over in the image of their answers. This value is of moral consequence. It gives rise to a moral interest in thinking robustly and fully about such things as what is true, what is worth doing, and what is beautiful or moving. This moral interest protects what a person believes and intends because they are component activities of thinking, in the way that swimming and cycling are component activities of running a triathlon.

If we combine the value of thinking with the role of thoughts in carrying it out, we get the following justification for the thought crime doctrine: Persons have a moral interest in thinking, which they realize by, among other things, believing and intending things. Hence, punishing someone for a belief or intention amounts to punishing them for their thinking, which is objectionable given the latter's importance.

This Article organizes the above argument as follows. In Section I, I survey some of the contemporary scholarship on the thought crime doctrine to motivate my view and distinguish it from others. In Section II, I defend two premises: that thinking is valuable and that beliefs are a critical part of conducting it. In Section III, I argue that intentions are equally critical to thinking. I then defend the argument against objections and consider its implications for First Amendment doctrines. The Conclusion compares my view to others.

I. CONTEMPORARY THOUGHT CRIME SCHOLARSHIP

A. *Alexander/Ferzan and Duff*

A number of philosophers and jurists have offered justifications for the thought crime doctrine, but none of them rest on the moral significance that thoughts have

for persons.¹⁰ I think they suffer for that reason. In this Section, I illustrate this point with reference to some representative contemporary scholarship.

One strand of thought crime scholarship holds that thoughts are not punishable, not because their wrongful features are outweighed by some value that they have, but because they do not have any features that would make them wrongful.

Larry Alexander and Kimberly Ferzan (AF) have prominently defended an approach like this. They contend that punishment is only appropriate for culpable wrongs and that culpability requires “unleashing” an unreasonable risk to the legal interests of another, where an unleashed risk is one that cannot be recalled or averted solely by an act of will on the part of the agent.¹¹ They then contend that mere intentions to commit wrongs are not culpable, because an intention does not execute itself in a mechanical fashion, but only through the will of the one who harbors it. Thus, forming an intention does not unleash a risk that would not be completely extinguished by abandoning the intention.¹²

Both of these contentions are doubtful. The first has been heavily criticized on the grounds that it is culpably wrongful to commit oneself to violating the interests of another, even if taking on that commitment does not itself unleash any risks.¹³ I’m sympathetic to this point. After all, intending or attempting to commit a wrong both warrant blame and a demand for an apology, the same as any other culpable wrong.

The minor premise—that intentions don’t unleash a risk that can’t be controlled just by willing its control—has also been criticized, though usually on the

¹⁰ Some relevant scholarship I will not specifically address includes the following. Federico Picinali, *A Retributive Justification for Not Punishing Bare Intentions or: On the Moral Relevance of the ‘Now-Belief’*, 32 L. & PHIL. 385 (2013); HUSAK, *supra* note 6; 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 (2005).

¹¹ Larry Alexander & Kimberly Kessler Ferzan, *Risk and Inchoate Crimes: Retribution or Prevention?*, in SEEKING SECURITY: PRE-EMPTYING THE COMMISSION OF CRIMINAL HARMS 103–06 (G.R. Sullivan & Ian Dennis eds., 2012).

¹² *Id.* at 118–19.

¹³ See R.A. Duff, *Risks, Culpability, and Criminal Liability*, in SEEKING SECURITY: PRE-EMPTYING THE COMMISSION OF CRIMINAL HARMS 132–38 (G.R. Sullivan & Ian Dennis eds., 2012); Gabriel S. Mendlow, *Why Is it Wrong to Punish Thought?*, 127 YALE L.J. 2342, 2350–53 (2018); Gideon Yaffe, *On Ferzan’s “Beyond Crime and Commitment,”* APA NEWSL. ON PHIL. & L., Spring 2014, at 5.

grounds that intending to do wrong creates some rationally uncontrollable tendency or psychological inertia to follow through and commit the wrong.¹⁴ I'm sympathetic to this point as well, but I think the appeal to psychological inertia is not the only way to make it. This is because making something true unleashes the risk that others will discover that it is true, and that discovery may, by itself, constitute an injury to their legally protected interests. For example, threats are criminalizable because they make persons fear for their safety.

Similarly, if one forms an intention to do harm, there is a risk that the target will discover it and reasonably fear for their safety. Of course, the risk of an *unspoken* intention being discovered is small, but the size of the risk is not material on AF's approach so long as it is unreasonable.¹⁵ Persons sometimes inadvertently give away their poker hand through a twitch or a twinkle in their eye, and they might do the same with an intention. Alternatively, others may, in principle, use a mind-reading machine to find out their intentions. Privacy rules may be thought to bar such access, but again, the question will be why such wrongful intentions should be protected by privacy. After all, revealing them could aid in preventing crime too.

For these reasons, I don't think the thought crime doctrine is justified on the grounds that thoughts are innocent. Indeed, if thoughts were innocent, it would be strange that there would be a doctrine to protect them.¹⁶ Many activities are innocent but not doctrinally singled out for protection because their innocence is all the protection they need—no one has seriously considered criminalizing solitaire. It seems more likely that the thought crime doctrine is a response to the fact that *though* thoughts have some features that warrant their punishment, they are nevertheless privileged against punishment.

R.A. Duff and Gabriel Mendlow understand the thought crime doctrine in this latter way—as a privilege to do something that would otherwise be punishable. However, neither one convincingly identifies the reason it should be maintained.

¹⁴ Daniel Ohana, *Desert and Punishment for Acts Preparatory to the Commission of a Crime*, 20 CANADIAN J.L. & JURIS. 113, 122–23 (2007).

¹⁵ Alexander & Ferzan, *supra* note 11, at 107.

¹⁶ Beliefs are not innocent either. For instance, racist beliefs can be wrongful, as well as beliefs about the value of attacking others, such as the one had by the resolution author from the introduction (i.e., an all-things-considered belief that attacking is a good idea).

Duff contends that the thought crime doctrine is justified by the decisive interest that persons have in the opportunity to abandon a wrongful intention before executing it. Punishment cuts off this opportunity and so is impermissible until the aimed-at wrong is in progress and so, in effect, impossible to fully avoid executing.¹⁷

Both premises are suspect.¹⁸ First, the interest appealed to is too indulgent and not distinctively about thought.

On the former point: It seems that there is a basis to punish persons for a wrongful intention well before they are in the process of executing it (attempt liability ubiquitously assumes as much), especially since part of the intention may be to create a situation in which intervention just before the last act will be prohibitively difficult. Moreover, the victim may become reasonably fearful well before the last act.

On the latter point: For Duff the interest in not being punished for an intention is just a special case of not being punished until a wrongful act is in progress, but this does not answer the question of what the *object* of punishment may permissibly be—what someone may be punished *for*. If someone is in the middle of acting on a wrongful intention, Duff permits punishing him, but can he be punished for the intention, now that he is executing it, or only for the execution itself?

Duff's second premise, that punishment cuts off the opportunity to abandon a wrongful intention, is also doubtful. One may still harbor a wrongful intention even after being caught ("I'll escape and try again!"), and if one abandons the intention because apprehension makes it impossible to execute, one may still have the further realization that it would be wrong to execute it. In other words, being caught and punished for intending to commit a crime does preclude one from having a change of heart about that intention. It is even possible to have such a change of heart after *fully executing* a wrongful intention, such as when one regrets what one has done because it was wrong and apologizes for having done it.

B. Mendlow

Like Duff, Mendlow assumes that the thought crime doctrine is a privilege, and in an important recent article, he offers a deductive argument in support of it.¹⁹ The

¹⁷ Duff, *supra* note 13, at 139–40; *see also* R.A. DUFF, CRIMINAL ATTEMPTS 386–90 (1996).

¹⁸ *See* Mendlow, *supra* note 13, at 2361–66 (offering sustained criticism of Duff's position).

¹⁹ *Id.*

driving idea behind the argument is that criminal punishment and crime prevention go together, such that if an activity may be punished for being wrong, then it may also be prevented—or as Mendlow prefers to say, disrupted—for being wrong.²⁰ Mendlow calls this idea the *enforceability constraint*, and he thinks it is both a sound moral principle that all legal systems should comport with (like “punishment should not be excessive”),²¹ and an accurate statement of current United States law (i.e., that current doctrine adheres to the enforceability constraint, though obviously not under that name).²²

The enforceability constraint is plausible and certainly holds true for a number of activities. Take, for example, selling loose cigarettes.²³ Persons can be punished for selling loose cigarettes and police officers may disrupt their sale by seizing them from a seller’s hands. And it would be strange if they could not do so. After all, if selling loose cigarettes is sufficiently wrong to warrant punishment if it occurs, then it seems that police officers are warranted in keeping it from occurring in the first place, just for being wrong.²⁴

If Mendlow is right that the enforceability constraint is a sound, general legal principle, then it can anchor an argument against punishing thoughts. Remember, according to the enforceability constraint, if the government may punish an activity for being wrong, then it may also disrupt it for being wrong. So, if the government may punish (a person for having) a wrongful thought, then it must be true that it

²⁰ *Id.* at 2370 (“[I]t’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.”).

²¹ *Id.* at 2371–73.

²² *Id.* at 2367, 2373 (“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. . . . [N]o 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.”).

²³ *Id.* at 2374.

²⁴ The “just for being wrong” clause is needed to rule out other grounds on which a person’s thoughts could be extinguished or manipulated by the government, such as to prepare a potentially incompetent defendant for trial. The fact that such mental disruptions are allowed does not warrant punishment for thought, because they are not undertaken on the grounds that the target thoughts are morally wrongful or aim at illegality. Rather, the target thoughts are compulsive or otherwise deleterious to the person’s mental competency. *Id.* at 2376–79.

may disrupt that thought for being wrong. But, Mendlow argues, the government is absolutely barred from disrupting thoughts just because they are wrong, such as by administering thought-suppressing drugs or by employing a science-fictional thought scrambler.²⁵ Therefore, the government may not punish thoughts either.

Another way to put the argument is as follows. According to Mendlow, the government may punish the sale of loose cigarettes on the condition that it may seize them from one's hands. And that condition is satisfied: It may seize them, so punishment for their sale is, so far as the enforceability constraint goes, permissible. Similarly, the government may punish thoughts on the condition that it may "seize" them, i.e., erase them from one's mind for being wrongful. But this condition is not satisfied. The government may not seize thoughts, and so may not punish them.

Despite its plausibility in a number of cases, the enforceability constraint is flawed. I will argue, in the following two subsections, that it is false as a statement of current law, and that it is doubtful as a justification for the thought crime doctrine.

1. The enforceability constraint does not describe current law

Contrary to the enforceability constraint, current legal doctrine recognizes conduct that may be punished, and yet, not forcibly disrupted. Two examples are criminally punishable omissions, such as failing to file a tax return, and criminally punishable speech that is shielded from disruption by the prior restraint doctrine. I will go in order.

Some crimes can be accomplished by both commission and omission.²⁶ For instance, one can criminally trespass by going onto property or refusing to get off of it, because these are both ways that one's body can be in the wrong place at the wrong time.²⁷ For crimes like these that may, but need not be, accomplished through omissions, disruption in Mendlow's sense is possible because the state of affairs that results from omitting to act can be prevented through force: A police officer can disrupt a criminal trespasser who omits to leave another's property by

²⁵ *Id.* at 2368, 2376–83.

²⁶ See *Jones v. United States*, 308 F.2d 307, 310 (D.C. Cir. 1962) (explaining the possibility of and doctrine surrounding commission by omission).

²⁷ *E.g.*, *State v. Shelby*, 811 P.2d 682, 686 (Wash. Ct. App. 1991) (criminal trespass law "ensures that persons without permission do not enter or remain upon the property of others").

dragging him away.

However, some crimes are necessarily omissions. Such omissions may not be disrupted by doing something to the perpetrator's body (or property or possessions), but only by directly goading the will itself. And it seems that any such goading would be illegal. For instance, police could not hypnotize someone who refused to file their tax return. Nor could they use deception or intimidation to do the same. Failing to file a tax return is a duty that persons must undertake on pain of punishment, but not one they can be compelled to fulfill through disruptive force, because the only kind of compulsion that could be used would constitute the illegal manipulation of a person's will. So, the enforceability constraint seems to be falsified.

Now consider the prior restraint doctrine. According to this doctrine, some speech that may be criminally punished may not be disrupted ahead of time.²⁸ For instance, the government may not confiscate or destroy obscene materials in order to prevent their publication, even if they could be the basis for a criminal conviction after publication.²⁹ In these cases, punishment is permissible but disruption is illegal. Again, the enforceability constraint seems to be falsified.

I say that these cases only "seem" to be counterexamples because they are not decisive. Mendlow has a sophisticated response.³⁰ It consists of two claims, one legal and the other conceptual. The legal claim is that there is no categorical legal prohibition on enjoining criminal omissions or criminal speech before they occur, so long as the appropriate adjudication is made. Mendlow defends this point only with regard to speech, by citing *Balboa Island Village Inn v. Lemen*.³¹ In that case,

²⁸ See *Near v. Minnesota ex rel. Olson*, 283 U.S. 697, 713–14 (1931) (foundational prior restraint case in which an injunction against a newspaper was struck down); Martin H. Redish, *The Proper Role of the Prior Restraint Doctrine in First Amendment Theory*, 70 VA. L. REV. 53, 53 (1984) ("Under the prior restraint doctrine, the government may not restrain a particular expression prior to its dissemination even though the same expression could be constitutionally subjected to punishment after dissemination.").

²⁹ See *Quantity of Copies of Books v. Kansas*, 378 U.S. 205, 210–11 (1964) (holding that seizing books without a judicial determination that they were obscene was a prohibited prior restraint); *Channel 10, Inc. v. Gunnarson*, 337 F. Supp. 634, 637 (D. Minn. 1972) (determining that seizure of film was a prior restraint because it physically prevented plaintiffs from distributing it); see also *Gaetani v. Hadley*, No. 14-30057-MGM, 2016 WL 593496, at *3 (D. Mass. Feb. 12, 2016) (noting that prior restraints of a purely physical nature are rare).

³⁰ Mendlow, *supra* note 13, at 2374–75.

³¹ 156 P.3d 339 (Cal. 2007).

the defendant was enjoined from repeating speech that had been found to be defamatory.³² The conceptual claim is that an injunction that is imposed before wrongful conduct takes place is: (i) a disruption of the enjoined conduct (ii) for the reason that it (the conduct) is wrong.³³ If the legal and conceptual claim both hold up, then my examples only illustrate that the law sometimes limits the kind of disruption that is permissible, but never actually goes so far as to prohibit it altogether. The enforceability constraint is thus not contradicted.³⁴

Sophisticated as this reply is, I'm not convinced. The legal claim is subject to some complications,³⁵ but the main source of my hesitancy is the conceptual claim. No doubt injunctions are "forward looking"³⁶ in that they aim at conduct that has not taken place yet, but this is a far cry from showing that they have both of two critical features listed above (i and ii). A natural way to think about injunctions is that they are a way of unambiguously informing someone of what the law requires in their particular situation. Because injunctions are individually tailored and explicitly communicated clarifications of the law, violating them will be a correspondingly distinct kind of heedlessness to a legal regime. Someone who violates an injunction does not just break the law, but ignores a reasoned judgment about how it applies in her case. This kind of heedlessness to legal authority is rightly said to be contemptuous of it.

This view of injunctions provides a basis on which to criticize both parts of Mendlow's conceptual claim. Against (i), notice that injunctions themselves do not

³² *Id.* at 343 ("[A]n injunction issued following a trial that determined that the defendant defamed the plaintiff that does no more than prohibit the defendant from repeating the defamation, is not a prior restraint and does not offend the First Amendment.").

³³ See *supra* note 24 (explaining why the enforceability constraint concerns disruption that is undertaken for the reason that the disrupted conduct is wrong).

³⁴ Mendlow, *supra* note 13, at 2375.

³⁵ See *Kinney v. Barnes*, 443 S.W.3d 87 (Tex. 2014) (representing the minority view that injunctions against future defamatory speech are not available, *contra Balboa*). Moreover, I do not think it is settled whether speech can be enjoined ahead of time anytime it is adjudged to be defamatory, as Mendlow thinks, or whether only the *repetition* of speech that has been adjudged to be defamatory can be enjoined ahead of time. The former view of the law is plausible, though as far as I can tell, the latter cannot be ruled out, and if it is true, then when criminal speech is uttered *for the first time*, it may be punished but not disrupted, falsifying the enforceability constraint.

³⁶ See *United States v. W. T. Grant Co.*, 345 U.S. 629, 633 (1953) ("The purpose of an injunction is to prevent future violations.").

apply force to prevent an ongoing or future activity. In this respect they are unlike Mendlow's main example of disruption: seizing cigarettes from someone's hand. Instead, injunctions are orders to act in a certain way. And while it is true that if the defendant does not obey the court's order, consequences will follow, this fact alone does not mean that injunctions disrupt wrongful conduct in Mendlow's sense. After all, criminal laws work the same way: they announce a requirement on conduct and then impose penalties for its violation. Yet criminal laws do not disrupt conduct; they punish it.

Against (ii), it seems that neither the reason for imposing injunctions, nor the punishment for their violation, is that the enjoined conduct is a criminal wrong. To be enjoined, conduct need not be criminally wrong, but only illegal in some way or another. For instance, failure to perform a contract can be enjoined. Likewise, punishment for violating an injunction is imposed for the same reason, whether or not the violation amounts to a criminal wrong in its own right; namely, that by violating the injunction, the defendant has ignored an authoritative, individually tailored, and explicitly-communicated statement of what she must do to act legally.³⁷ For example, if the defendant in *Balboa Island* were to violate the injunction against repeating the defamatory statements that she had made before, she would be punished with (criminal) contempt of court.³⁸ But, crucially, the basis of contempt would not be that she defamed the plaintiff (again), but rather that doing so constitutes contempt of the court's order telling her that she must stop. Consequently, her punishment would be for the wrong of disobedience to the judiciary, i.e., showing contempt, and not for defaming the plaintiff.³⁹

³⁷ An objector might interrupt at this point to stress that the punishment for the injunction might be for disobedience, and yet *also be* for the wrong itself, on the grounds that the wrong itself is the cause for the injunction being properly issued and so for its proper enforcement as well. Two possible responses are as follows. First, even preliminary injunctions are not justified simply because foreseen conduct is wrong, but, at least partially, to preserve the status quo for adjudication. See *Roe v. U.S. Dep't of Def.*, 947 F.3d 207, 231 (4th Cir. 2020). Second, and similarly, the collateral bar rule suggests that an injunction is to be obeyed not because it is correct in addressing a wrong but because it is a kind of judgment that must be given authoritative force in order for justice to be administered in an orderly fashion. See *In re Establishment Inspection of Hern Iron Works, Inc.*, 881 F.2d 722, 726 (9th Cir. 1989).

³⁸ *Balboa Island Vill. Inn, Inc. v. Lemen*, 156 P.3d 339, 353 (Cal. 2007) (Baxter, J., concurring) (explaining that violation of the injunction at issue would be remedied with criminal contempt).

³⁹ See Margit Livingston, *Disobedience and Contempt*, 75 WASH. L. REV. 345, 351–52 (2000)

Overall, I think that it is not correct to distinguish injunctions and criminal punishment on the basis of whether they address wrongdoing *ex post* or *ex ante*, remedially or prospectively. Instead, the relevant distinction between them is in terms of separation of powers and the specificity of law. The legislature imposes general obligations on citizens, some of which are criminal, and judges may make these obligations more explicit and determinate to particular citizens through injunctions. The latter do not seek to disrupt wrongs any more than the laws that are their basis. Instead, they make an obligation known, and because they do so in a highly explicit way, their violation is contemptuous of judicial authority, and may be punished for that reason. This is not disruption, but a distinct kind of remedy for a distinct kind of disregard for the law.

2. The enforceability constraint does not provide a justification for the thought crime doctrine

Mendlow not only argues that the enforceability constraint describes current legal doctrine, but that it proscribes the latter's correct form. That is, that the law should adhere to the thought crime doctrine because it is a sound moral generalization.⁴⁰

I am unsure that it is sound, partly because it seems to me that the prior restraint doctrine is not morally arbitrary. It seems to have something going for it as a moral principle—that persons may sometimes be entitled to “lay what sentiments he pleases before the public” though doing so may warrant punishment.⁴¹ For instance, persons may be punished for not telling the truth under oath, but I doubt they can be given a truth serum to disrupt their perjury by forcing them to speak sincerely. Persons must be given the chance to speak their truth, even if some will not take the opportunity and will choose to dissemble instead.

I'm also skeptical that the permissibility of punishment rests on the permissibility of disruption. Instead, it seems to me much more likely that punishment is the conceptually primary tool of a legal regime. I won't argue for this claim here, except to point out that punishment is addressed to a person, whereas disruption is not. It's possible to punish a person *for* something she did, but it's not possible to disrupt someone (for something she did). It's only possible to disrupt the *activity*

(discussing reasons for various kinds of contempt penalties).

⁴⁰ Mendlow, *supra* note 13, at 2371–73.

⁴¹ See 4 WILLIAM BLACKSTONE, COMMENTARIES ON THE LAWS OF ENGLAND 151–52.

she is engaged in. Put another way, disruption only accidentally involves a person but punishment necessarily does. I suspect this fact will play a key role in justifying punishment as a form of legal redress.⁴²

Given that I won't pursue the above arguments, let me stipulate that the enforceability constraint *is* a sound moral generalization in order to argue for a different point, which is that it still would not constitute a justification for the thought crime doctrine. To set up this objection, consider this true moral generalization: it is impermissible to levy a \$2 tax on some activity if it is impermissible to levy a \$1 tax on it. Still, the truth of this generalization does not constitute a justification for the immunity that the activity enjoys from being taxed. Why is it impermissible to tax it even \$1? The generalization just given assumes an answer to this question rather than giving it.

Likewise, the claim that it is impermissible to burden thoughts with punishment if it is impermissible to burden them with disruption could be true, but it is not a justification for the thought crime doctrine, because it does not tell us why thoughts enjoy immunity from being disrupted. Even accepting this generalization, we still want to know: Why is it impermissible to burden thoughts, even by disruption?

Mendlow does not focus on answering this question, but he does not ignore it. In addressing it briefly, he suggests that thoughts may not be burdened by disruption or punishment, because persons have the right to *mental integrity*, which is roughly, a right to retain whatever thoughts or mental states one has, free of brute psychological or physical interventions.⁴³ This right serves an interest in controlling one's identity. However, I doubt mental integrity is the base justification for the thought crime doctrine. My main reason for doubt is that it seems that there must be a (further) reason why persons have this right. Otherwise, we would be forced into a cramped libertarianism according to which control over ourselves is for the sake of having selves that answer to our control.⁴⁴

⁴² Michael L. Rich, *Limits on the Perfect Preventive State*, 46 CONN. L. REV. 883 (2014) (arguing that persons have a right to break the law and not simply be prevented from ever doing so).

⁴³ Mendlow, *supra* note 13, at 2376–83.

⁴⁴ Others have similar concerns about justifying a right to X by appeal to the good of being free to do X or to control X, since this is a description of what that right comes to. See Jed Rubenfeld, *The Right of Privacy*, 102 HARV. L. REV. 737, 750 (1989) (pointing out that invoking autonomy to justify the right to (constitutional, decisional) privacy is a restatement of what privacy protects rather than

To bolster this argument, consider taxing thoughts. I submit that it would be wrong to tax thoughts based on their content (e.g., a tax for every unpatriotic thought that one has), and it is plausible that the reason for this is, at root, the same reason they cannot be disrupted or punished. Yet taxing something is not premised, as disruption and punishment are, on the *per se* wrongfulness of the thing being taxed.⁴⁵ Rather, a tax assumes that the thing being taxed is permissible, conditional on the payment of the tax. If taxing thoughts does not violate mental integrity but nonetheless is an objectionable burden on thought, then it seems that the right to mental integrity is just one manifestation of some deeper reason against the legal burdening of thoughts. It is this deeper reason we should seek.

II. A THINKING-BASED ARGUMENT AGAINST PUNISHING BELIEFS

This Section develops an argument against criminalizing beliefs on the grounds that they are how a person thinks (about what is true).

A. Major Premise: Persons Have an Interest in Thinking

The major premise of my argument appears frequently in First Amendment jurisprudence and scholarship, which is that persons have a constitutionally protected interest in thinking.⁴⁶

For some, the interest in thinking is a derivative one. It is derived from some *other*, more fundamental interest that persons or societies have, such as sustaining democratic practices or knowing the truth. For example, according to some versions of the “marketplace of ideas” theory, the “freedom to think as you will and to

a justification for it); Robert Post, *Participatory Democracy and Free Speech*, 97 VA. L. REV. 477 (2011) (arguing that a flat appeal to “autonomy” in First Amendment law is suspect).

⁴⁵ This point is a conceptual one that also enjoys legal acceptance. A tax does not entail that the taxed activity is prohibited or may be prohibited, but rather entails only that it is conditionally permissible, i.e., permissible on the condition that the relevant tax is paid. Supreme Court opinions agree. See Nat’l Fed’n of Indep. Bus. v. Sebelius, 567 U.S. 519, 573 (2012) (“[T]he power to tax is not the power to destroy while this Court sits.” (quoting Okla. Tax Comm’n v. Tex. Co., 336 U.S. 342, 364 (1949)) (cleaned up)).

⁴⁶ See, e.g., SEANA VALENTINE SHIFFRIN, *SPEECH MATTERS* (2014); C. Edwin Baker, *Autonomy and Free Speech*, 27 CONST. COMMENT. 251, 259 (2011); Charles Fried, *The New First Amendment Jurisprudence: A Threat to Liberty*, 59 CHI. L. REV. 225 (1992); David A. Strauss, *Persuasion, Autonomy, and Freedom of Expression*, 91 COLUM. L. REV. 334 (1991).

speak as you think,”⁴⁷ is needed so that competition about what is true will be sufficiently fierce.⁴⁸ Fierce competition of this kind is not thought to be valuable in itself, but rather to be instrumentally valuable. That is, fierce competition about what is true is thought to be valuable because of its conduciveness to discovering and promulgating the truth, in the same way that fierce economic competition is thought to be valuable because of its conduciveness to the emergence of well-organized firms or superior products.

I do not deny that thinking is conducive to the truth, but I do not follow the marketplace approach in supposing that its conduciveness in this regard is the whole of its value. That is, I deny that thinking is worth fostering or protecting only because and to the degree that it produces good results, such as truth (or good art, or useful inventions, or a sound democracy and so on).

Instead, I claim that thinking is valuable just as the activity it is, which is that of trying, seeking, or striving to understand something better, often, but not always, by trying to secure knowledge about it. Imagining something necessarily seeks to know what it would be like and considering a question necessarily seeks to know its answer. Some thinking is more open-ended in that it does not seek a particular piece of knowledge, but a general topic to think further about. So, one can broadly “consider the lobster,” or simply let one’s mind wander to whatever thoughts catch one’s attention.

The important point for my purposes is that attempting to understand something has a distinct kind of value. The reason is that understanding things is valuable, and attempting to do so is a distinct kind of commitment to that value. A different kind of example is helpful. If one sees another in danger, one may hope they get rescued. Such a thought is nice as far as it goes, but it’s cheap if one is in the position to pull off rescue oneself. By contrast, *trying* to rescue the person constitutes one’s commitment to the importance of their being rescued. The rescue attempt may fail due to unforeseen circumstances, but the attempt is no less good on that account.⁴⁹ The person was as committed as they could be to the right course of

⁴⁷ *Whitney v. California*, 274 U.S. 357, 375 (1927) (Brandeis, J., concurring).

⁴⁸ See C. EDWIN BAKER, *HUMAN LIBERTY AND FREEDOM OF SPEECH*, 6–24 (1992) (discussing the marketplace of ideas at length); Stanley Ingber, *The Marketplace of Ideas: A Legitimizing Myth*, 1984 DUKE L.J. 1, 2 n.2, 4.

⁴⁹ Gideon Yaffe has prominently argued for a parallel claim about wrong actions, in the context

action. Thinking is similar. By doing it, one is as committed as one can be to greater understanding.

By claiming that thinking is valuable as an attempt to understand, I am not suggesting that it is of no consequence whether one thinks correctly or not. Mistaken thinking is, just on that basis, flawed or deficient. It is not as good as it can be. However, and this is a reiteration of my key point: the fact that mistaken thinking is not as good as it can be does not mean that it's not good at all. It is still good in being the initiation of a relationship with its subject matter, in the pursuit of developing an understanding of it. Furthermore, a mistaken bit of thinking always lies ready to be redeemed, because mistakes pave the way, in principle, for a deeper understanding of what is correct. "Failure is success in progress."⁵⁰

Some examples will substantiate the foregoing arguments about the value of thinking as well as the corresponding moral interest that persons have in carrying it out.

Consider curiosity. Curiosity is a virtue because it is an attunement or sensitivity to the value of thinking. Someone who is curious does not simply want the correct answers to questions, in the way that a contestant on a quiz show might (to win!), but rather, is drawn to think about them and their significance. Someone who is curious enjoys forming and considering questions so as to reach their answers through engagement with the relevant evidence and arguments, and to thereby experience, explore, and ultimately understand them.⁵¹ Notice that curiosity does not guarantee accuracy. Someone who is curious may get things wrong, and yet their curiosity is still good, because it is a disposition to get caught up in thinking.⁵²

of criminal attempts; namely, that attempting something wrong is itself wrongful in virtue of being a commitment to the wrongful course of action. Gideon Yaffe, *Criminal Attempts*, 124 YALE L.J. 92, 101–03 (2014).

⁵⁰ This quote is frequently attributed to Albert Einstein. I return to this idea in the next section.

⁵¹ See Elias Baumgarten, *Curiosity as a Moral Virtue*, 15 INT'L J. APPLIED PHIL. 169 (2001).

⁵² A cousin virtue to curiosity is creativity. Curiosity is a desire to explore the reasons that bear on various truths, and creativity is a desire to explore various media, language, or bits, in an effort to find out what can be made of them. Here too, the exercise of creativity is valuable in itself, even if the final product ends up being banal or drab. See Marc Jonathan Blitz, *The Freedom of 3d Thought: The First Amendment in Virtual Reality*, 30 CARDOZO L. REV. 1141, 1172–73 (2008) (discussing the use of imagination); TIMOTHY MACKLEM, INDEPENDENCE OF MIND (2008) (focusing on languages of

The value of thinking can also be seen at work in the subtle norms that govern telling and being told. What I mean is that we frequently have reason to depend on the thinking of others instead of our own. We “take their word for it” or “take it on their authority.” We have access to the truth of what we are told (assuming it is true) only by crediting the thinking of the one who tells it. For example, non-specialists rely on physicists to tell them how various sub-atomic particles work, clients rely on lawyers to tell them how the law works, and juries rely on eyewitnesses to tell them what happened.⁵³

But such dependencies are a morally sensitive matter, precisely because they involve the dependent party not thinking and instead accepting authority. The one that is told something may gain knowledge about physics or the law, but because that knowledge is not gained through thinking, it is deficient as a bit of understanding. Knowing something about quarks on the authority of the physicist is not the same as arriving at that knowledge by thinking about how quarks work, and the difference is significant.

For example, it can be patronizing to tell someone something that they can be anticipated to already know or to disagree with, because it solicits their dependence on one’s thinking when they clearly have their own. Telling someone something may also displace or scramble a person’s thinking, as when a teacher is too heavy-handed about the right response to a complex issue (students may take their word for it and not think on their own), or when a respected adviser gives her opinion before the advisee has had a chance to organize or process his own thinking.⁵⁴

The pitfalls of testimony do not lie solely with the testifier. Persons may make themselves dependent on the thinking of others too willingly, by asking their opinion too quickly or by believing what they say uncritically. In law school, students complain that professors “hide the ball.” This criticism may be warranted in some cases, but it may sometimes be true that students wish to be told “the law,” when in reality, knowing the law is a matter of sustaining a critical judgment about the import of various legal materials. This is not just a matter of legal pedagogy. Courts

all kinds as providing a space for creative exploration).

⁵³ See Michael S. Pardo, *Self-Incrimination and the Epistemology of Testimony*, 30 *CARDOZO L. REV.* 1023 (2008) (discussing testimony).

⁵⁴ See George Tsai, *Rational Persuasion as Paternalism*, 42 *PHIL. & PUB. AFF.* 78, 93–94 (2014) (giving an example of a parent stating their (perhaps well-reasoned) judgment about where a child should go to college before the child has a chance to think about the question).

structure their opinions as opinions, viz., as reasoning that others may wish to follow and evaluate for themselves.⁵⁵

To sum: dependence on the thinking of others can be justified, but when it is, it is a second-best option that is forced by pragmatic constraints (persons can't know everything). When feasible, it is better for persons to think for themselves and so exit relationships of intellectual dependence.⁵⁶ This fact owes to the value of thinking for oneself.

This value, I will now argue, forms the basis of a moral interest. That is, it is not just valuable to think about things for oneself, but valuable in a way that generates moral requirements on others to avoid interfering with it. The main reason for this moral interest is that one maintains and develops one's particular self or personality through thinking. Rivers exist and evolve by flowing, and persons exist and evolve by thinking, which makes its continuation critically important to them. By finding one's own way through a particular set of reasons and claims, one also, thereby, lives out a life as the particular person one is.

This is not to say, with Mendlow, that one needs to control one's thoughts in order to control one's self. Rather, one needs to control one's thoughts in order to carry out the activity of thinking, where this activity necessarily includes, as a consequence, the creation and development of a self. For example, by imagining the characters in a book, one is related to them, and in being related to them, one is changed. By seeing reasons for some claim, one thereby comes to believe it, and this too is a change in oneself.

Concrete examples bolster these remarks. A stark one is mind control. It is wrong to implant a belief in someone's mind, even if true, because doing so circumvents the subject's own thinking. Moreover, insofar as the belief is not a product of the person's thinking, it is not a part of *her* at all, but a partial supplantation of her.⁵⁷ The same can be said of other, more subtle ways of mentally influencing

⁵⁵ For instance, courts try to state the question that a case poses. Even this is just a suggestion. Others may have a more defensible view of the true question that the case poses.

⁵⁶ Thinking for oneself does not require thinking alone. Many ways of thinking with others do not involve believing what they say outright, but only thinking about what they say, suggesting counterpoints, and so on. Much dialogue is like this.

⁵⁷ See Fried, *supra* note 46, at 233 ("No conviction forced on us can really be ours at all.").

persons, such as compelled speech or compelled listening.⁵⁸ They aim to carve away a part of the person and replace it with a “dead” psychological disposition.

In addition to the manipulation of someone’s thoughts, the interest that persons have in thinking can be violated by interfering with the conditions needed to realize it robustly. For example, it is wrong to deny someone an adequate education, and it is also wrong to limit their access to intellectual materials like books, conversations, and other media—so-called “food for thought.”⁵⁹ Without the opportunity to think robustly and fully, someone’s personality can wither, even if their biology is healthy.

A final example is paternalism, or “substituted judgment.” In some cases, persons not only have an interest in thinking for themselves, but also, for their thinking to be given effect in the world. For instance, I, along with my partner, have an interest in planning and carrying out our daughter’s birthday party. This is our business.

If someone else, say, a loving grandparent, planned and executed her party without our permission (suspecting, perhaps, that we would plan a bad party), we could reasonably object, even if the grandparent did a good job, or a better job than we would have done, along some metric of success. Our objection would be against the substitution of someone else’s judgment for our own, and it would be reasonable because we have a special relationship with our daughter that we deepen by thinking about who she is and what she would enjoy for a party, and by that thinking actually carrying the day and becoming reality (i.e., when our thinking about her birthday is not just an exercise or a game but the presumptive basis for what her birthday should be like). Likewise, persons have a special relationship to themselves such that their thinking about things should carry the day as to the person that they are.

In these examples then, we see that persons have a moral interest in thinking. The reason is that thinking, the ongoing doing of it, is how persons sustain and strengthen themselves, as the particular person that they are. A compact summary

⁵⁸ Seana Valentine Shiffrin, *What Is Really Wrong with Compelled Association?*, 99 NW. U. L. REV. 839, 852–60 (2005) (discussing compelled speech).

⁵⁹ SHIFFRIN, *supra* note 46, at 90–91 (discussing the effect of the lack of intellectual stimulation in solitary confinement); Rachel Aviv, *How Albert Woodfox Survived Solitary*, NEW YORKER (Jan. 8, 2017), <https://perma.cc/EL8E-F2TK> (same).

of this idea is given by the cliché: “it’s not the destination, it’s the journey.” One climbs a mountain, for instance, not just to get to the top, but because of the encounter with nature that one will have on the way to that goal. Moreover, only by being preceded by such a climb will one’s arrival at the top count as the accomplishment of summitting.

Thinking is similar. It often has a destination, such as knowledge of some kind, but part of its value lies in what happens on the way toward reaching it—namely, an encounter with ideas and the corresponding development of one’s position in response. Moreover, only by being preceded by such an encounter will one’s grasp of the truth count as the accomplishment of understanding.

B. Minor Premise: Punishing Beliefs Interferes with the Interest that Persons Have in Thinking

I will now defend the minor premise of my argument: that punishing beliefs necessarily interferes with a person’s interest in thinking.⁶⁰ The reason is that believing is a part of thinking. This is intuitive. What a person believes is, as we say colloquially, what she thinks. If this is right, then punishing a person for her beliefs amounts to punishing her for her thinking.

Intuitive as it may be, this claim—that believing is thinking—warrants some additional defense because it is subject to the challenge that a belief is not thinking but something else, such as a state or condition that thinking puts one in. This challenge has some heft, on account of the fact that beliefs seem to be things that one has *after* one has figured out what seems true regarding a particular subject, and so not a part of the figuring out.

The correct response to it is to question the assumption that thinking only takes place when one is trying to figure out what to think. But thinking is, plausibly, just as much a matter of maintaining or preserving the figuring out one has already done. Let me expand on this point.

Imagine that one starts to think about whether global warming is caused by humans. This is a question and thinking about it consists in trying to answer it correctly by adducing and weighing relevant reasons. If this bit of thinking yields a sufficiently persuasive answer, then one will form a corresponding belief, say, that global warming *is* caused by humans.

⁶⁰ My discussion below focuses exclusively on the punishment of false beliefs, because I think that it is easier to justify punishing false beliefs than true ones.

However, this belief is not, I submit, an end to one's thinking about the original question, but rather is the stabilized or equilibrium form that one's thinking about that question currently takes.⁶¹ Put differently, believing something does not end one's thinking about the corresponding question, but rather, one continues thinking about it in the same way, in the sense of continuing to see the same conclusion as warranted or sensible.

A metaphor may clarify further. Consider jogging. One can jog as a way to get to a place, but one can also jog as a way of staying there (jogging in place). Thinking is similar in that one can think as a way of "getting" to a conclusion, but one can also think as a way of "staying" there. Believing is this latter kind of thinking. What one believes is what one thinks is true, currently, actively, and ongoingly.

This view of believing is supported by the fact that a belief is, owing to its nature as being one's current thinking, susceptible to being changed, just by doing more thinking. For example, if one believes that global warming is caused by humans, but then comes across what one thinks is a sufficiently powerful counter-consideration, then one ordinarily will change what one believes.⁶² This makes sense if believing just is the current form one's thinking takes. Changing the latter will necessarily change the former, because they are the same thing. Put slightly differently, a belief, by its nature as a bit of thinking, stands ready to be superseded by doing more thinking. It stands ready to be re-thought and revised when new evidence or arguments emerge.

This readiness to be rethought makes believing a fundamental means by which persons make progress toward a more defensible, reasonable, and responsible worldview. They are a stab at the truth that establishes a framework or position from which to encounter further evidence, engage in further discussion, and undertake further reflection.⁶³

⁶¹ See Matthew Boyle, *Two Kinds of Self-Knowledge*, 78 PHIL. & PHENOMENOLOGICAL RSCH. 133 (2009) (discussing believing as an ongoing process); Pamela Hieronymi, *Believing at Will*, 35 CANADIAN J. PHIL. 149, 175 (2009) ("My beliefs do not sit in my mind as last week's lecture sits on my hard-drive, recording what I once thought. My beliefs are rather my present, on-going take on what is so.").

⁶² I say ordinarily because there are many things that can interfere, but if things are going as they should, the transition from re-thinking to re-believing is automatic. Hieronymi, *supra* note 61, at 176 ("[A]s I think about what is so, my beliefs *therein* change.").

⁶³ There is a saying: "If you don't stand for something, you will fall for anything." Beliefs are

To draw out this claim, compare believing to taking a practice shot on the basketball court. A practice shot can succeed as a bit of practice, whether or not it goes in, because one can learn from either outcome. If one sinks the shot, one can note what worked, and if one misses, one can adjust for next time. Nothing guarantees that one will shoot better next time, or ever, but to have a chance at improving, one needs the opportunity to keep shooting.

A belief is similar in that it's a shot at the truth about some question, and it can succeed as a way of developing one's worldview, just in being such an attempt, whether or not it's true or false. The reason, again, is that just as a missed shot can set up a better next shot, so an incorrect belief can set up a better superseding belief. There is no guarantee that it actually will. A false belief may persist forever, but it is, in its nature, open to being corrected and therein contributing to a stronger overall understanding. To realize that possibility, one needs the opportunity to keep believing.

Beliefs, even false ones, set up improved thinking by allowing one to inhabit, or adopt, a perspective on the truth, so as to become especially sensitive to its strengths and weaknesses, and to see where it leads. For instance, a false belief may, if taken as the basis for further reasoning, lead one to a dead end, in a way that warrants circling back and correcting the belief. This activity of adopting beliefs and thereby becoming sensitized to their consequences is, more or less, the activity of theorizing. One sees an idea as sufficiently plausible to warrant believing it, and then follows out its implications, adjusting as necessary.

A second example of this same point is the process of revising. Things like books, articles, and emails are not written in finished form all at once. Drafts are necessary because they make vivid one way of thinking, in a way that lives on and informs later versions (as a missed shot informs later shots).⁶⁴ An empty page is intimidating because it contains nothing definite toward which to direct one's critical attention, whereas if one has written *something*, then one has a target to work

like this in that if you cannot take positions on what is true, you will have little ground from which to encounter new evidence and undertake new reflection.

⁶⁴ See Keith Oatley & Maja Djikic, *Writing as Thinking*, 12 REV. GEN. PSYCH. 9 (2008); Keith Oatley, *Thinking Deeply in Reading and Writing*, in *THE EDGE OF THE PRECIPICE* 175, 180 (Paul Socken ed., 2013) (quoting Richard Feynman as responding to an interviewer's statement that his notebook was a record of his work where he responded, "it's not a record, not really. It's *working*. You have to work on paper and this is the paper.").

with, or against. And as one accumulates discarded positions, one hones in on the right one.

Finally, consider activities like switch-side debating, the appointment of an *advocatus diaboli* in sainthood deliberations, and adversarial systems of justice like our own.⁶⁵ The common factor in these practices is that thinking gets better to the degree it must confront sincere adherence to a different view. In these cases, such a confrontation is organized interpersonally, by bringing in different people, but something similar can happen intrapersonally, within a single person over time. Someone who forms a belief but then is compelled to abandon it in the face of new arguments and experiences learns by making a mistake, and this is a path to a personally robust form of understanding that is sometimes termed wisdom.⁶⁶ John Stuart Mill famously formulated a version of this point when he pointed out that regular engagement with falsehoods and errors sustain a deeper grasp on the truth.⁶⁷ And there is no greater engagement than believing the falsehood oneself.

I have just defended the claim that believing is one way of thinking about what is true. A belief is just the form that the latter takes at one moment in time. With this claim, we can see why punishing a belief amounts to the punishment of a person's thinking. Take, as an example, someone who is curious about whether vaccines are dangerous. He is likely to defer to the experts, but he might not. He might decide to think about it himself, and on the basis of doing so, come to believe that vaccines are dangerous. Assume that this belief is false and that it is likely to cause him to act in ways that jeopardize public health.⁶⁸ Notwithstanding its falsity and dangerousness, his belief is a part of thinking, and so something he has an interest in forming and maintaining.

Of course, under different circumstances, the anti-vaxxer might believe the

⁶⁵ See Casey Harrigan, *Against Dogmatism: A Continued Defense of Switch-Side Debate*, 29 CONTEMP. ARGUMENTATION & DEBATE 37 (2008) (switch-side debating).

⁶⁶ Todd D. Rakoff, *The Implied Terms of Contracts: Of 'Default Rules' and 'Situation-Sense'*, in GOOD FAITH AND FAULT IN CONTRACT LAW 191 (Jack Beatson & Daniel Friedmann eds., 1995) (arguing that persons obtain wisdom by playing different relational roles, such as parent, friend, employer, trustee, etc.).

⁶⁷ JOHN STUART MILL, ON LIBERTY 34–35 (1895).

⁶⁸ See, e.g., Megan Elyse Waterman, Note, *Indorsing Infant Immunity: An Argument for Criminalizing Parents' Refusal to Immunize Their Children*, 51 TULSA L. REV. 153 (2015).

truth. Indeed, he might still do so in the future after he thinks more about the question or learns new information. As things stand now though, he cannot help believing that vaccines are dangerous. Crucially though, the fact that he cannot help believing it is not the reason that punishment is objectionable. Rather, such punishment is objectionable because what he believes now is his thinking on the matter.

At this point, I have offered a *prima facie* case for thinking that punishing beliefs is wrong. The reason is that their punishment is nothing less than the punishment of a part of thinking. Persons need to think, and they think by believing. Hence, punishing beliefs interferes with thinking. Put another way, persons have an interest in developing their own perspective and they cannot do that without thinking for themselves. Since the latter requires believing what one thinks is warranted, one must be permitted to believe on that basis. These points apply to the anti-vaxxer. He has an interest in knowing the truth about vaccines for himself, which requires believing what he takes to be true even if it's false. Hence, he must be permitted to believe vaccines are dangerous.

To bolster the *prima facie* case, I will now offer a second argument. Its conclusion is that punishing a specific belief will deter thinking about the topic of the belief more broadly. This is a chilling argument.

Some chilling arguments are based in the fact that innocent conduct may be easily mistaken by a judge or jury for conduct that is illegal.⁶⁹ Fear of these mistakes motivates a person to steer away from that conduct. In other cases though, a chilling effect may be due to the fact that one kind of activity leads, without volition, to another, such that imposing liability for the second activity provides a reason against engaging in the first. For example, imagine that some jurisdiction makes it a crime to use opioids to satisfy an addiction.⁷⁰ This law has a chilling effect because

⁶⁹ See Leslie Kendrick, *Speech, Intent, and the Chilling Effect*, 54 WM. & MARY L. REV. 1633 (2013) (pointing out the problems with some kinds of chilling arguments, including those that depend on legal bodies making mistakes).

⁷⁰ Such a law would likely be unconstitutional under current law, in virtue of coming close to criminalizing addiction. See *Robinson v. California*, 370 U.S. 660, 666 (1962) (the status of being addicted is not a permissible object of criminalization). Importantly, the *Robinson* court sees part of the problem as being that one could be convicted of violating the law “at any time before he reforms,” whereas the law I’m imagining does not have that consequence, but only permits criminalization of a particular opioid use that is motivated by one’s addiction. *Id.* 663.

those who wish to comply with it are well-advised not just to refrain from addictively taking opioids (the object of the law's prohibition), but to refrain from non-addictively taking them too.

The reason for this is straightforward: One non-volitionally becomes the other. If one uses opioids non-addictively, one risks crossing the invisible line into addiction, and if one becomes addicted, one will very likely violate the law, for two reasons. First, the addiction will motivate one to do what the law prohibits, and second, the addiction will make one unreasonably insensitive to the punishment one will suffer for doing what the law prohibits, which might otherwise keep one from doing it. In short then, if one wishes not to use opioids addictively, one will prefer not to use them non-addictively too, because the latter risks the former.

Thinking is, albeit for very different reasons, similarly vulnerable to being chilled. If one thinks, one may be persuaded to believe what it is illegal to believe. Nothing guarantees that one's thinking will yield the belief that the law requires. For example, thinking about vaccine side effects may lead to the belief that they are dangerous.

Moreover, if one will be persuaded to believe what is illegal to believe, the prospect of punishment will not deter one from believing it. This is because thinking about what is true properly ignores reasons that are practical. More carefully, one cannot think that something is true in order to secure the practical consequences of thinking that way. As many have put it, the good consequences that will come from believing that something is true is the "wrong kind of reason" to believe it.⁷¹

A stark example of this wrong-kind-of-reason claim is that one cannot believe that the earth is flat because an eccentric billionaire has offered to pay one handsomely for that belief. This is because though the billionaire's offer makes it good for one to have that belief, it does not provide any support for the truth of the belief's content, and so one cannot have the belief by thinking about its content.⁷² To believe that something is true by way of thinking, one must see a sufficient reason for the truth of its content, and monetary reward for harboring the belief itself is no such a reason.

⁷¹ See Pamela Hieronymi, *The Use of Reasons in Thought (and the Use of Earmarks in Arguments)*, 124 *ETHICS* 114 (2013).

⁷² Of course, one could take a drug that brainwashes oneself into believing the earth is flat to get the money.

We can now see why thinking about a topic unavoidably risks thinking about it in the way that could be punished. Take the vaccine example. If a citizen starts thinking about vaccines, then there is a risk that he will end up believing that they are dangerous, and this risk is not ameliorated by his desire to avoid punishment, since punishment will be the wrong kind of reason against having the belief. Thus, whenever a belief is punished, thinking about that topic is chilled in its entirety since the latter may risk producing the punished belief, whereas one is certain to avoid the punished belief by simply forming no beliefs on the subject at all.

An important feature of this chilling argument is that it is contingent. It only holds so long as the world doesn't change in certain ways. For example, punishing addictive opioid use chills non-addictive opioid use only contingently, as things are now. If technology advances sufficiently, it may become possible to ensure that one can avoid an addiction to opioids if one chooses. Were such a breakthrough to come about, then one would not risk addictively using opioids by using them non-addictively.

Similarly, we could suppose that pills were developed that prevented persons from forming certain beliefs, such as that vaccines are dangerous. In this case, it could be claimed, persons could think all they wanted, safe in the knowledge that their thinking would not progress to the point of having an illegal belief. Hence, thinking is chilled by punishing a specific belief only as things are now, contingently, and not necessarily.

Still, the chilling argument is a justification for the thought crime doctrine in our world as it is, because mind altering pills are only science fiction. But even more important is that the pill scenario reinforces my first, non-contingent argument of this Section, which was that beliefs themselves are thinking. After all, it does not seem that a law that punishes a particular thought is rendered innocent by the widespread availability of a means to counter its chilling effect on nearby thoughts. Instead, it seems that the pill is just a guarantee of self-censorship. Put another way, the pill allows persons to more cleanly face the choice of either thinking and suffering punishment (no pill), or else avoiding punishment at the cost of chemically guaranteed self-censorship (the pill). But given the value of thinking, both options are objectionable, and so forcing a choice between them is objectionable as well.

Applying the lessons of this Section to the resolution author from the introduction is a good way to sum up. The resolution author, like our hypothetical anti-vaxxer, believes that something is true: that all-things-considered, he should attack

someone with brass knuckles. This belief is thoroughly wrong, but the harboring of it is, itself, thinking, and so important. Punishing it would punish the resolution author for trying to realize the interest he has in thinking. Moreover, punishing his belief in the wisdom of attacking will chill his thinking about what to do more generally, since he may worry that if he thinks about what to do, he will end up reaching a punished conclusion.

At this point, I conclude my argument against punishing beliefs. Succinctly summarized, it is that believing is thinking, so punishing the former punishes the latter and thereby interferes with the interest in doing it. In what follows, I extend the argument just given to intentions and then answer objections.

III. EXTENDING THE THINKING-BASED ARGUMENT TO INTENTIONS

A. *Punishing Intentions Also Interferes with the Interest that Persons Have in Thinking*

I argued just above that persons have an interest in thinking and that forming beliefs, even mistaken ones, is a crucial part of that enterprise. In this Section, I give a similar argument for intentions. They too are important for thinking.

It may seem that no argument to this effect is likely to succeed. After all, aren't intentions about doing things rather than thinking about them? Isn't someone who has an intention finished thinking and so resolved or committed to a course of action? The view of intentions that is implicit in these questions mirrors the one about beliefs that I tried to unseat in the previous Section. Recall that according to that view, beliefs are not a part of thinking, but only the products of it. Intentions could be conceived like that too—as products of thinking with motivational consequences. However, this view of intentions is as suspect as the parallel one about beliefs. A better view is that intentions constitute thinking about what to do, just as beliefs constitute thinking about what is true.

Before expounding on this parallel, it's important to note an important fact about intentions, which is that they have a special relationship to decisive, all-things-considered, or "all out" beliefs about what one should do.⁷³ If one has such

⁷³ An example of a belief that is not about what one should do would be a belief that the sky is blue. This belief is just about the sky and has nothing to do with how one should act. By contrast, the resolution author from the introduction has an all-things-considered belief about what to do—namely, an unconditional belief that attacking is what he should do.

a belief, one is rationally obligated to form a corresponding intention.⁷⁴ For example, if one believes, without any qualifications or caveats, that one should travel abroad next summer, then one must, rationally speaking, form an intention to do so.⁷⁵ Correlatively, the failure to intend to do what one decisively believes one should do is a rational error, often called weakness of the will.⁷⁶ An example of this kind of error is forming an all-out belief that one should start a project, but forming no intention to start (and perhaps continuing to scroll social media).

I can now state this Section's main claim with more precision: When one has an all-out belief about what to do, forming the corresponding intention continues the thinking begun by the belief, in a new form. To intend is to push one's thinking further, or, to think again, from a new angle. Put another way, persons are not just rationally obligated to intend to do what they believe they should do, they also have an interest in so intending, as that is how they think more thoroughly about what to do.

A good way to introduce my argument for this main claim is by drawing some parallels with what I have already argued. In the previous Section (II.B), I emphasized that for persons to think about *what is true*, they must do more than adduce and weigh reasons. They must be permitted to go further and believe what their adducing and weighing reveals to be true. For example, one thinks about the question of whether global warming is caused by human actions by trying to answer it, which is to say, by adducing and weighing up relevant reasons and then forming the corresponding belief. This attempted answer then sets up further thinking by

⁷⁴ This claim is endorsed by many philosophers, and often discussed under the heading of weakness of the will or akrasia. They call "judgment" what I have been referring to as a belief about what is worth doing. See, e.g., 1 DONALD DAVIDSON, *How is Weakness of the Will Possible?*, in *ESSAYS ON ACTIONS AND EVENTS* 41 (1st reprt. ed. 2001) ("What is wrong is that the incontinent man acts, and judges, irrationally, for this is surely what we must say of a man who goes against his own best judgement."); Sergio Tenenbaum, *Akrasia and Irrationality*, in *A COMPANION TO THE PHILOSOPHY OF ACTION* 274 (Timothy O'Connor & Constantine Sandis eds., 2010) (discussing Davidson). But see Nomy Arpaly, *On Acting Rationally Against One's Best Judgment*, 110 *ETHICS* 488 (2000) (acknowledging, but dissenting from, the "almost. . . universal assumption in contemporary philosophy . . . that acting against one's best judgment is never an instance of rational action").

⁷⁵ By contrast, one might believe something like "traveling next summer would be good" or "traveling in the summer is nice." Either of these beliefs is compatible with some other course of action being better, and so are not decisive or all out beliefs about what one should do.

⁷⁶ See *supra* note 73 (explaining further).

providing a framework or launching point for encountering new evidence and arguments.

Parallel claims are true for intentions. For persons to think about what to do, they must do more than just have beliefs about what to do. They must be permitted to go further and intend to do what they believe they should do. For example, one thinks about the question of whether to travel in the summer by trying to answer it, which is to say, by forming a belief about whether to do so followed by the corresponding intention. This attempted answer then sets up further thinking.

In this comparison, we see that the formation of beliefs is sufficient for thinking about what is true, because beliefs are how one answers questions of fact. However, thinking has to progress further when the question at issue is not just about the *truth* regarding what to do, but about what to actually *do*. This progression is accomplished by forming intentions because only they answer that latter question.

There are two specific ways in which intending to do what one believes one should do furthers thinking. One is that by intending to do what one believes one should do, one takes greater responsibility for one's belief, which attaches greater urgency to it. Imagine that two friends are watching a game and the first declares that one team will win. The other says, "talk is cheap, why don't you put your money where your mouth is?" Here, the request to back up her statement with a bet is a way of asking the first friend to take responsibility for what she says, by raising the stakes for being wrong.

Intending is to believing, what betting is to speech in the foregoing example. A belief that one should act a certain way is "cheap" because it does not require anything more.⁷⁷ However, by intending to act as one believes one should one commits to actually taking action, which increases the importance of being right, which in turn attracts more of one's cognitive resources to thinking about it. If one is asked to perform a calculation and is told that something important rides on its accuracy, one will likely calculate more carefully. The same is true of intentions. By intending to act as one believes one should, one takes responsibility for acting that way by committing to make it reality.

A second way that intentions further thinking about what to do is by changing its mode or nature. A belief relates a person to content in a descriptive mode. To

⁷⁷ Think of someone with a lot of resources who believes that helping others is important, but never helps.

believe X is to commit to X being a true description of how things are, and this is the case even if X is about an action. But intentions relate a person to content in a prescriptive way. To intend X is to commit to making X true. This change in mode as one goes from an all-out belief to intending to act provides insight in the form of information about what it will be like to act as the intention requires. Persons often think about what to do by imagining what it would be like to act in different ways, and intentions are a distinct way of doing that because they engage a person's agency. Intending to act is a way of simulating or previewing what it will be like to act by using one's agency to sustain a commitment to doing it.

By gathering up the preceding arguments, we get the following picture. Persons who believe they should do something have begun thinking about the question of what to do, but they have not yet answered it, because a belief, unlike an intention, does not commit one to acting. By intending, one gives and maintains an answer to the question of what to do. And by giving this answer, one is related to the question in new ways.

Of course, persons infrequently change their intentions once they form them, but that's also true of beliefs. Persons have many beliefs they rarely revise, and yet, beliefs are bits of thinking because they are susceptible to being re-thought. Intentions too can be re-thought and abandoned, unlike actions, which occur irrevocably. Recall that Duff thought this fact about intentions—that they can be rethought and so abandoned—was the reason that they could not be punished. I argued against that claim, but my argument in this Section illustrates why he might have thought the possibility of revising intentions was important.⁷⁸ As I see it, the importance of the fact that intentions can be revised is a *consequence* of the fact that they are bits of thinking, and it is this more fundamental fact that warrants exempting them from criminal punishment.

I will end this Section by returning again to the resolution author from the Introduction. In thinking about what to do, he adduces and weighs the relevant reasons, and believes that he should attack a particular person. This belief wears two hats. On one hand, it is a complete (but not unrevisable) answer to the question of what is true, albeit with regard to a practical subject, i.e., the wisdom of attacking. It thus earns protection as a bit thinking about what is true. Furthermore, the reso-

⁷⁸ See text accompanying *supra* note 17.

lution author's belief that he should attack is *part* of a complete answer to the question of what to do, in virtue of being an all-out belief that mandates the formation of a corresponding intention to attack. Hence, the belief also earns protection as a bit of thinking about what to do.

The intention to attack, however, is the resolution author's complete (but not unrevisable) answer to the question of what to do, and so earns protection for that reason. It's understandable that such an intention provokes concern, since it is a commitment to attack. However, the committal nature of an intention to action is part of what makes it an answer to the question of what to do, and so what makes it an important way of thinking about that question.

B. Addressing Objections

In this Section, I address three potential objections to the argument I have given. The first two deal with intentions specifically, and the last concerns my argument generally, as it applies to beliefs and intentions.

1. The argument proves too much: If intentions may not be punished then neither can actions, which is absurd

One might worry that my argument proves too much. Specifically, it may seem that if intentions are a part of thinking, then nothing prevents actions from being the subject of a similar argument. After all, the objection might go, persons think about what to do not just by intending to act, but by acting. Hence, actions continue the thinking accomplished by the intentions that they execute and are, therefore, a person's full answer to the question of what to do. If this argument is correct, and it turns out that actions play a critical role in thinking about what to do, then my position has the consequence that the thought crime doctrine should protect actions from being criminalized, and that consequence is absurd.

My response to this objection is to deny the key premise that persons necessarily think about what to do by acting.⁷⁹ To be sure, several nearby claims are true. Persons *learn* a great deal about what to do by reflecting on their actions. They also think about *how* to do things (know-how is gained by practice), *what it is like* to do them, and, in some cases, *who they are*,⁸⁰ by acting. Still, actions do not, generally

⁷⁹ Thanks to Eugene Volokh and Jane Bambauer for pressing me to clarify my position here.

⁸⁰ For instance, persons may think about their sexuality identity by having various kinds of intimate encounters with others.

speaking, constitute one's thinking about what to do.

A better view of actions is that they are ways of implementing or executing one's thinking about what to do and are not part of it, much as performing a contract is subsequent to setting its terms. Actions follow from one's thinking about what to do. This view of action is intuitive, and it has considerable support.

First, as I just mentioned, acting can help one learn how to do something. That's what practice is. But if action is the attainment and subsequent use of "know *how*," then it seems that it is not thinking about *what* to do. Instead, the thinking that happens in action is that of achieving a goal that one has set for oneself by intending. Second, actions do not necessarily involve a relationship between a person and an idea.⁸¹ Persons believe *that* something is true, and they intend *to* do something, but an action has no analogous subject matter. At least some actions are not about anything, they just get done. Third, it's often unproblematic for others to execute one's intentions. For example, persons in a position of authority may execute intentions by asking others to perform the relevant actions. However, it's not possible to do someone's intending for them. The reason for this difference is that what one intends is one's thinking about what to do whereas actions only implement that thinking.

Suppose I am right that acting is not, itself, a part of thinking about what to do. Still, the objector could press on in a different way. She might concede that punishing action does not directly punish thinking about what to do, but insist that punishing action still heavily burdens the latter, since nearly every punishable action will be the direct result of thinking about what to do. For example, a law that punishes attacks with brass knuckles will make it burdensome to believe that one should attack or to intend to attack, since these thoughts will very likely lead to attacking and so, punishment. If this argument is right, then my view has the consequence that actions cannot be punished, on account of the chilling effect that such punishment would have on thinking. Again, this consequence is absurd, and so challenges me to show that punishing actions does not unduly burden thinking.

There are two responses to this objection. One is that actions are, as I just argued, the execution of one's thinking. Because they finalize one's thinking in an irrevocable form, actions are how persons signal its completion, so that it can be

⁸¹ Though speech does, because the act of saying something is constituted by the idea that is communicated.

evaluated by others. Hence, punishing the resolution author for attacking with brass knuckles does not punish him for what he thinks, but rather, for the fact that his thinking has come to rest in the form of action.

A second, related response makes use of the distinction between forming or changing a thought by thinking, and forming or changing a thought as the collateral consequences of doing something else. For example, one can change one's belief that vaccines are dangerous in two ways. One way is to rethink it, which one could do by reflecting on the evidence one has, reading a scientific article, or performing an experiment. A different way to change one's belief that vaccines are dangerous is to take a pill that overwrites it with a belief that vaccines are safe. The key point is that both methods change one's belief, but only the first way achieves that change *by thinking*.

This distinction applies to intentions too and it sharply distinguishes the punishment of actions and intentions. Punishment for an action provides a reason to rethink the corresponding intention. This is the premise of deterrence, which is that someone who intends to do something will think about the punishment that will come with it, and so "think again" about whether to do it, decide it's not worth it, and thereby lose the intention.

Punishing an intention does not work the same way, owing to the "wrong kind of reason" idea I discussed above.⁸² Recall that an offer of money is not a reason that bears on one's belief that the earth is spherical. It's the wrong kind of reason to change that belief, though it could be a reason to revisit it to see if one still holds it. This is because the money makes it good to harbor the belief that the earth is flat, but it provides no support at all for thinking that the earth is actually flat. Thus, the only way to benefit from the billionaire's offer is by overwriting one's belief that the earth is spherical with the belief that it is flat.

Intentions follow this pattern, if we think of punishment as an incentive like money.⁸³ Being punished for an intention to attack with brass knuckles provides no reason to rethink that course of action. This is because the prospect of punishment makes it good to not harbor the intention to attack, but it does not show its subject

⁸² See text accompanying *supra* note 71.

⁸³ Punishment is a negative incentive and money is a positive one, but the difference does not matter for the argument I'm giving.

matter—the attacking—to be bad or unreasonable. The only way to avoid punishment for an intention is to overwrite it, not rethink it (just as the only way to benefit from the billionaire’s offer is to overwrite, not rethink, one’s belief that the earth is spherical).

Let me clarify how these comments address the objection. The objection argued that punishing action chills thinking, so that my view must be, absurdly, committed to categorically shielding even actions from criminal punishment. My response is that punishing actions does not chill thinking at all—no more than scientific studies chill the belief that vaccines are dangerous. Punishing criminal actions deters them by giving persons a reason to reconsider their commission. Punishing intentions on the other hand provides no such reasons. It only provides reason to rid oneself of them.

2. The argument proves too little: If actions can be punished then so can intentions

A different objection picks up where my reply in the preceding Section left off. I just claimed that punishing actions does not interfere with thinking because it provides a reason to think differently (if not about the merits of the action apart from its legal consequences) about what to do, rather than providing a reason to overwrite one’s thinking about what to do. But, it may seem that this claim is false and that in fact punishing intentions is like punishing actions in giving reasons to think differently. If this is right, then it seems that punishing actions and intentions are similar not different, and hence that if punishing actions is permissible, then punishing intentions must be too. If this objection goes through, then I have not shown that intentions should be any more protected from criminalization than ordinary criminal actions.

This objection needs to be developed a bit. As we saw, one can only believe something for money by taking action that will overwrite one’s thinking, such as by taking a pill. Intentions can seem different. After all, one *can* intend something for the practical benefits of so intending, without taking a pill.⁸⁴

For example, one can intend to whistle for money. How? First, one recognizes

⁸⁴ Detailed discussion of this asymmetry can be found in the work of Pamela Hieronymi. See Pamela Hieronymi, *Controlling Attitudes*, 87 PAC. PHIL. Q. 45, 56–58 (2006); Hieronymi, *supra* note 61; Pamela Hieronymi, *Two Kinds of Agency*, in *MENTAL ACTIONS* 138 (Lucy O’Brien & Matthew Soteriou eds., 2009).

that if one sees sufficient reason to actually whistle, one will thereby form an intention to whistle. Then, one can see the money as a sufficient reason to whistle, thereby causing one to intend to whistle and earn the money. Notice though that one can only see the benefit of intending to X as a decisive reason to actually X if that benefit outweighs any downsides to X-ing. For example, it is unlikely that one would intend to drink a lethal poison for a few bucks, because the only way to form that intention and get the money would be to see the money as a decisive reason to actually drink the poison, and that's unreasonable.

If this is right—that one can take into account the consequences of intending to X in deciding whether to X—then it may seem that punishing intentions is really no different than punishing action. After all, if the punishment for attacking someone with brass knuckles were stringent enough, such that a would-be perpetrator would decide against it, then levying the same punishment against the intention to attack would have the same effect. The would-be perpetrator would see that deciding in favor of attacking would cause him to have the prohibited intention and so earn him the punishment, and since we are assuming that he does not see the crime as worth the punishment, he would not decide in favor of committing it.

But though the deterrent effect may be the same, and though no pill is needed, the route to reaching it is importantly different. I accept that one can intend in order to secure the consequences of so intending, but I contest the assumption that in doing so one is changing one's thinking in response to reasons as opposed to overwriting one's thinking through collateral means. Indeed, I think that the possibility of intending to get a practical benefit is like taking a pill to stifle the formation of criminalized beliefs that I discussed above.⁸⁵

One can take a pill to change what one will believe, but the change is not made by thinking about what is true, as it would be if one got new evidence and changed one's mind on that basis. Instead, the change comes from a source, the pill and the decision to take it, that is alien to thinking about what is true, but rather comes from a practical decision motivated by the value of not thinking. Intending so as to reap a benefit for that very intending is similar. Owing to how intentions work, one can pull it off without a pill, but in pulling it off, one is exploiting one's own agency to mimic the effects of the pill.

To elaborate, when one takes the pill to prevent one from believing something,

⁸⁵ See *supra* text accompanying note 72.

one has the end goal of halting one's thinking. The pill is a means to that end. As we have seen, one might take there to be a reason to adopt this end, say, some prize money for extinguishing the belief, but nonetheless, in adopting it, one adopts the end of paralyzing one's own thinking, through chemical means.

In a case where the intention to attack another with brass knuckles is punished, one may choose not to attack so that one will not intend to attack. But here, the decision not to attack is like the intention to take the pill: The reason for it is that doing so will be a means to the end of paralyzing one's own thinking. The fact that the "pill" is provided by the structure of one's own psychology is immaterial. That makes pulling things off convenient, but does not change the key fact, which is that it is a form of paralyzing one's own thinking. One's intention is the way it is because of one's decision to do something that will affect it (deciding not to attack so as not to have it) and not because of one's thinking about the relevant subject matter.

A second way of putting my argument is like this. Say that one was trying to consider whether to attend a party or not. If one weighs the reasons for attending, the result may be an intention to go or not, based on thinking about the merits of going. However, one could decide not to go on the grounds that deciding not to go would make it true that one did not have an intention to go. If one decides this way, then one has not made a decision about whether it would be worth going and so one's lack of an intention does not reflect a belief that it's not worth going. Instead, the lack of one's intention is what one values in this case, for itself, and not as a consequence or reflection of one's thinking that the party is not worth attending. This way of thinking about the party is distortive, since it elides the question of whether it's worth going to the party, and instead answers the question about going to the party on the basis of what that answer will cause in oneself.

3. The argument does not justify the thought crime doctrine: Other interests could outweigh the interest in thinking

This objection gives a chance to fill a gap in my argument. I have argued that persons have a moral interest in believing and intending as they see fit, because these are both ways of thinking. However, it does not follow that punishing thoughts should be categorically prohibited. It could be true that the interest that persons have in thinking should be outweighed, in certain cases, by other important moral interests, like that of preventing crime.

In response, I want to outline some reasons in favor of the current doctrine's categorical prohibition on punishing beliefs and intentions. One is that the interest

that persons have in thinking is a preeminent one, and another is that the interest in thinking plays a special role in political life.

First, persons enact their personality by thinking. The ability to do this according to one's view of the relevant reasons is a fundamental kind of freedom. Correspondingly, its abridgement is a sufficiently grave wrong to warrant its categorical prohibition. This is especially true because almost all of a person's important life choices flow from the thinking that they do about what is true and what to do.

Second, the permission to think as one will is a foundational part of a sound political order. Fully arguing for this would overwhelm the goals of this Article, but roughly, politics aims to impose sound constraints on citizens that flow from their collective thinking. Such collective thinking requires equality among citizens, which at least requires that each person be equally able to reach their own perspective about what is true and what to do.⁸⁶ Criminalizing thinking directly undermines this equality, because it makes the thinking of one group the basis of limiting the thinking of another. It amounts to nothing more than the legally backed assertion: "I'm right and you're wrong." Things are different with the criminalization of actions, because by leaving a person's thoughts alone, they give space for dissent.⁸⁷ Criminalizing action thus says something like "we're doing things this way for now," where the "for now" reflects the fact that in a society that permits thinking by all on equal terms, a criminal law (with a proper *actus reus*) is, like a belief, permanently susceptible to revision, even if in fact, it always earns sufficient support and so is never repealed.

Furthermore, one of the requisites for imposing constraints on persons in the name of collective thinking lies in the potential for those constraints to be, in principle, *thought of* as sound. But a law that punishes thinking about some topic X will be self-defeating on this front. It will preclude full thinking about X, and so full thinking about the justifiability of the law. Thus, laws that criminalize thought compromise their own status as possible objects of uncoerced acceptance. They are, in this respect, brute exercises of power. They cannot be expected to be understood by those whom they affect.

⁸⁶ For one sustained development of this idea, see Anne Barron, *Kant, Copyright and Communicative Freedom*, 31 LAW & PHIL. 1 (2011).

⁸⁷ See Fried, *supra* note 46, at 233 (pointing out that a government can regulate action but not thought).

Other more pragmatic considerations support the thought crime doctrine's categorical prohibition as well. For instance, categorically prohibiting the punishment of thought makes it clear to citizens and to prosecutors and police what is punishable and what is not. Such clarity reduces the chilling effect that would come from the comparatively uncertain application of a balancing approach by judge, juries, or both. Second, the burden that the thought crime doctrine places on crime prevention is not excessive. Requiring the government to withhold punishment until some action takes place gives the government a significant opportunity to prevent criminal wrongdoing from being completed, especially given that some enabling conditions for criminal wrongdoing can be prohibited altogether, like prohibitions on possessing things like burglary tools or brass knuckles.

C. *Consequences for Free Speech Doctrine*

My focus has been on why beliefs and intentions cannot be punished by themselves. This issue is of theoretical interest, but practically speaking, very few if any criminal laws attempt to punish nothing but a thought. They invariably include an actus reus. Does this mean that they invariably respect a person's interest in thinking? Not necessarily. In this Section, I will discuss how the interest in thinking is implicated by more than just the punishment of thoughts by themselves, but by laws that improperly incorporate thoughts as elements.

A focal example for this discussion is hate crime laws that add additional penalties for crimes that are committed out of racial bias. Some have supposed that such laws criminalize thought in virtue of increasing punishment due to the perpetrator's biased thoughts. Susan Gellman puts this point in representative form: "The only substantive element of most hate crime statutes is that the defendant had a bias motive for committing the base offense. As motive consists solely of the defendant's thoughts, the additional penalty for motive amounts to a thought crime which offends the First Amendment."⁸⁸ The problem with this argument is that motive "is by no means a unitary concept, and . . . the First Amendment has very different implications depending on what *aspect* of an actor's motive is being singled out for punishment."⁸⁹

⁸⁸ Susan Gellman, *Hate Crime Laws Are Thought Crime Laws*, 1992/1993 ANN. SURV. AM. L. 509, 514–15 (emphasis omitted).

⁸⁹ Laurence H. Tribe, *The Mystery of Motive, Private and Public: Some Notes Inspired by the Problems of Hate Crime and Animal Sacrifice*, 1993 SUP. CT. REV. 1, 5.

For instance, if a motive for intending or acting a certain way is the *further intention* for which the action is performed, then punishing some motives will violate the First Amendment and others will not. For instance, taking property with the intention to permanently deprive the owner of it can be punished more severely (as larceny) than the taking itself.⁹⁰ But this does not amount to the punishment of the intention, but rather reflects the fact that the action of taking is of a different kind than one that was not performed to fulfill a further wrongful intention. In this case then, the best thing to say is that taking property with larcenous intent is punished more harshly than just the taking because it is a worse kind of taking, not because the intention is being assigned its own quantum of punishment.⁹¹

Based on the foregoing discussion we can see that a hate crime law that punished an assault more harshly if it was done with a further intent of “terrorizing a community” or provoking racial animosity would not thereby punish thought.⁹² Instead, it would punish assaults of an especially serious character. By contrast, a law that punished an assault more harshly if it was done with the further intent of expressing a racist message would punish thought. This is true because the further intention is protected by the First Amendment and so an illegitimate basis for treating the assault as worthy of greater punishment. For instance, some speech can be punished for violating legitimate time, place, and manner restrictions, but additional punishment cannot be heaped on such speech because of the message it sought to express. Another example is symbolic expression, which may be punished owing to nature of the symbol but not for the message the symbol is being used to convey. Flag burning can be punished if and because there are dangers posed by the act of burning something, but additional punishment is not warranted on the grounds that the burning expressed, or was done with the intention of expressing, criticism of the United States.⁹³

A hate crime law would also not necessarily punish thought if it operated in the manner of the law in *Wisconsin v. Mitchell*.⁹⁴ The law at issue in that case punished

⁹⁰ *E.g.*, *People v. Allen*, 19 N.Y.S.3d 592, 594 (App. Div. 2015).

⁹¹ This is the main theme of Mendlow in his paper subsequent to the one I discussed above. See Gabriel S. Mendlow, *Thoughts, Crimes, and Thought Crimes*, 118 MICH. L. REV. 841 (2020).

⁹² This is Tribe’s “Ordinance D.” See Tribe, *supra* note 89, at 2.

⁹³ See *Texas v. Johnson*, 491 U.S. 397 (1989).

⁹⁴ 508 U.S. 476 (1993). *Wisconsin v. Mitchell* is the basis for Tribe’s “Ordinance B.” See Tribe,

more severely those crimes that were motivated by the intention to victimize someone of a particular race.⁹⁵ Put another way, the law punished crimes more severely if the perpetrator had a *mens rea* of intent toward an additional circumstantial element: that of the victim being of a certain race. Here too, punishment is increased for the defendant's thoughts being a certain way, but it is plausible here too that the increased punishment is owing to the way that those thoughts alter the base crime and not because those thoughts are themselves worthy of punishment.⁹⁶

Finally, a hate crime law would punish thought if it more severely punished crimes that were committed for particular reasons, such as the belief that certain groups deserve to be mistreated.⁹⁷ This kind of law would not base increased punishment on what intention or intentions the defendant had, but on why she had those intentions, i.e., her beliefs about who is worth respecting.

So far, I have been trying to enumerate the correct doctrinal *conclusions* regarding hate crimes. The picture that emerges is that actions may be punished and that sufficiently related, non-speech intentions can be grounds to punish them more harshly, but may not be targets of punishment themselves.⁹⁸ On the other hand, a person's beliefs and her communication of them (including her intentions to communicate them) may not be punished and may not even be invoked as grounds to heighten the punishment of related actions. Burning a flag might violate a safety law forbidding open fires, but there cannot be the aggravated offense of "starting a fire with an intent to convey criticism of the U.S.," because its aggravative nature would lie in the expression of an idea.⁹⁹

supra note 89, at 2.

⁹⁵ See Yaffe, *supra* note 49, at 135 ("After all, it is common to believe a condition to be in place without it being something to which one is committed by one's intention. The runner believes that he lives in Los Angeles, but we cannot infer from this fact, together with the fact that he intends to go running, the further claim that he intends to go running in Los Angeles. That is something that he takes for granted, not something that he directs his will toward in the form of an intention.")

⁹⁶ Plausible, but not beyond debate. Gellman argues at length that crimes committed intentionally against victims of a certain race cannot be said to be worse than other crimes, except by making reference to the wrongfulness of bigotry. Gellman, *supra* note 88, at 518–20.

⁹⁷ This is Tribe's "Ordinance C." See Tribe, *supra* note 89, at 1–2.

⁹⁸ "Sufficiently related" is needed to rule the pure conjunction of a wrongful act and a wrongful intention, such as Mendlow's imagined crime of "littering with the intent to kill." See Mendlow, *supra* note 91, at 855.

⁹⁹ See e.g., *Texas v. Johnson*, 491 U.S. 397 (1989) (flag burning that expressed political criticism

This picture is, I submit, more or less what one would expect if my claims about beliefs and intentions are correct. I cannot demonstrate this point fully here, but some comments can support it. On the view of thinking I offered above, beliefs are fundamental. They are how one starts to think and how one comes to a conclusion about what would be good to do here and now. Intentions are ways of continuing to sit with that conclusion, and of encountering it in a new (agential) mode, but they also aim at action, and action is the execution and finishing of one's thinking. As a result, intentions occupy a middle space between thinking about what would be good to do (beliefs) and doing it (actions). They are one's conclusion about what to do that are poised to be done.

This middle space status is why intentions may not be punished themselves, but also why they are relevant to assessing punishment. This is clear in the ordinary *mens rea* situation. If one intended to do something, then it was not an accident. It is sufficiently connected to one's thinking to be an *intentional* action and so one that can be aptly evaluated and, possibly, the target of blame (where action that is unintentional is much less liable to blame). However, an intention is *what* one is committed to doing, but does not include *why* one saw fit to do it. The latter are one's beliefs about what is good and true, and as we saw, they are protected from punishment, and also from being used to calibrate punishment.

The special status of beliefs is mirrored by that of speech, which is how one makes one's beliefs known to others. Beliefs are thinking that is about something, and speech that communicates them is the same; it is about something in a way that permits others to think about it too. Hence, speech and intentions to speak are like beliefs in being about things, and as a result, they cannot be punished nor be the basis for increased punishment of admittedly wrongful actions.

This difference between intentions and (non-speech) actions on the one hand and beliefs and speech about what is true on the other finds tangled expression in cases of incitement and solicitation. In these kinds of speech, the speech is being used in lieu of one's own non-communicative actions, but in different ways. Take a request/demand. I can realize my intention of drinking by grabbing the cup myself or by requesting or commanding that you bring it to me. The latter seeks to

cannot be punished for expressing that criticism, though punishment on other grounds could be permissible).

realize my intention by seeing if you will realize it for me. For this reason, someone who starts to commit a crime and someone who advocates for the crime are similar in an important way, which is that they have acted on a wrongful intention in a way that seeks to bring it about. The former has actually initiated the crime under their own power and the other has spoken in a way that seeks to enlist another to the same purpose.

Incitement is related, but need not take the form of a request or demand. If one's intentions are to cause "imminent lawless" action with a sufficient chance of success, then one may incite others by saying all sorts of things.¹⁰⁰ One could incite another by exclaiming a stirring slogan,¹⁰¹ by making the case for acting, or by stating facts that are likely to spur action. Such talk is, judged in isolation, likely to have a substantial role in the advocacy of ideas. However, when it is aimed at producing a specific action with little time for further consideration on the part of the person being incited, even the offering of persuasive reasons becomes sufficiently close to a way of enacting or causing wrongdoing that it no longer earns protection as thinking. Here too, as with solicitation, the inciter must answer for the (illegal) object of her intention in virtue of how her words are connected to its enactment.

Admittedly, these arguments are highly sketched. There are details to consider and exceptions to contend with—can they be reconciled with my broader theoretical commitments? There are difficulties I have not grappled with here. My point has been to show that actions and intentions have the status they do in criminal law and in First Amendment jurisprudence because the former is the execution of thinking rather than its continuation, and because the latter aims at such an execution. Beliefs however are thinking without end. One holds them indefinitely and when one abandons them, thinking does not end but simply changes (I see that this is true where before I thought the other). Speech is an action, but distinctly related to thinking in being an attempt to do the indefinite thinking of a single mind with others, to reach common conclusions, and to do so, ideally, before acting. Criminal laws are a society's attempt to state the results of collective thinking about what conduct cannot be tolerated.

¹⁰⁰ *Brandenburg v. Ohio*, 395 U.S. 444, 447 (1969) (per curiam).

¹⁰¹ *Hess v. Indiana*, 414 U.S. 105 (1973) (discussing the exclamation "we'll take the fucking street later" but finding it was not uttered with the requisite intention).

CONCLUSION

I have addressed the question: What is the justification for the thought crime doctrine? My answer relies on claims about the value and structure of thinking.

I argued that thinking is valuable because it is how one builds up an understanding of what is true, right, and beautiful, on one's own terms. This activity, the constant doing of it, is how persons sustain and develop themselves, much as a river does by flowing, or plants by photosynthesizing. Due to its special connection to the formation of a person, the value of thinking gives rise to a moral interest in carrying it out.

I also argued that one crucial aspect of thinking is trying to carry it to completion, which requires that persons be permitted to give their own answers to the questions that they confront. Specifically, I argued that beliefs are a person's answers to questions about what is true, and that intentions are a person's answers to questions about what to do. Having these thoughts is not like having something in one's garage, where it lies idle until called upon. Rather, thoughts are actively maintained, and their active maintenance is itself, a way of giving them rational attention, i.e., thinking about them.

As a result, punishing a person's thoughts is like damming a river or putting a plant in the dark. It suppresses the activity by which persons preserve their personhood and preserve, in principle, their chance to see what is correct. No doubt, there are sobering consequences to giving persons permission to think for themselves—the resolution author from the introduction has reached objectionable conclusions—but this permission cannot be revoked, least of all by a political system. To do so would give up on persons altogether. Instead, a political community must have the courage to let its members realize the value of thinking, despite the risks. Courage in this regard is “the secret of liberty,”¹⁰² and the thought crime doctrine recognizes this secret.

This theory has some significant advantages. Mendlow argued that the thought crime doctrine is based in the interest that persons have in controlling their identity. On my view however, control is not important for its own sake, but only because it's inseparable from the activity of thinking. Thinking is how a person forms their own, particular relationship to ideas and reasons, and hence it requires respecting her view on them, i.e., permitting her to control the vehicles by which she

¹⁰² *Whitney v. California*, 274 U.S. 357, 375 (1927) (Brandeis, J., concurring).

develops her view. Hence, a person's control over their thoughts is necessary if they are to grapple with the subject matter toward which those thoughts are directed. In this way, my view does not celebrate control over one's identity for control's sake, but for the sake of being able to carry out thinking, which is a standard-governed and potentially social activity.

Duff argues that persons deserve an opportunity to abandon wrongful intentions, but we saw that this was too indulgent and, in any case, does not actually show that intentions should not be the basis for punishing someone, but only that punishment, for whatever reason, is impermissible so long as its target has not started on committing a crime.

On my view, both problems disappear because the role that thoughts play in thinking stems directly from their nature as thoughts. My argument thus permits even a "first act" theory of criminal attempts, since a first act is not a thought—though of course there may be other non-thinking-based reasons (or even penumbraic thinking-based reasons) against such a theory. On the other hand, my argument does prohibit punishing someone *for* their intention, even when they are in the middle of executing it or are about to take the last act needed to do so. Only the execution of the intention is the proper object of punishment. The intention itself is off-limits because it is a bit of thinking.

Given these advantages, a thinking-based theory is best positioned to support the full extent of the thought crime doctrine, and no more.