Law and Philosophy (2015) 34: 601–631
DOI 10.1007/s10982-015-9235-z

JOSHUA GLASGOW

THE EXPRESSIVIST THEORY OF PUNISHMENT DEFENDED*

(Accepted 20 June 2015)

ABSTRACT. Expressivist theories of punishment received largely favorable treatment in the 1980s and 1990s. Perhaps predictably, the 2000s saw a slew of critical rejections of the view. It is now becoming evident that, while several objections to expressivism have found their way into print, three concerns are proving particularly popular. So the time is right for a big picture assessment. What follows is an attempt to show that these three dominant objections are not decisive reasons to give up the most plausible forms of the view. Moreover, in addition to the three common objections, expressivism has an acknowledged question mark concerning whether the value of punitive expression outweighs its drawbacks. Here I also map out some promising avenues that the expressivist can take to answer this question.

Expressivist theories of punishment received largely favorable treatment in the 1980s and 1990s. Perhaps predictably, the 2000s saw a slew of critical rejections of the view. It is now becoming evident that, while several objections to expressivism have found their way into print, three concerns are proving particularly popular. So the time is right for a big picture assessment. What follows is an attempt to show that these three dominant objections are not decisive reasons to give up the most plausible forms of the view. Moreover, in addition to the three common objections, expressivism has an acknowledged question mark concerning whether the value of punitive expression outweighs its drawbacks. Here I also map out some promising avenues that the expressivist can take to answer this question.

Since expressive theories actually make up a family of views rather than any one view, Section 1 specifies what I take to be the

* For comments on previous versions of this article, I am very grateful to Thom Brooks, Victor Tadros, Bill Wringe, and two anonymous referees for *Law and Philosophy*.

most plausible form of expressivism before examining, in Section 2, the three standard criticisms. After dispatching those criticisms, Section 3 revisits the big question: whether expressivism can offer an overwhelming positive justification for punishing offenders. Although I concur with the consensus that fans of expressivism have some work to do on this front, I will offer a three-option sketch of how this work might be completed. Should the arguments below succeed, the overall takeaway is that the extant critical attitude toward expressivism is unwarranted, as the prevailing objections are at best undercooked. Expressivism is at worst merely viable, given the current state of the dialectic.

Before getting into that dialectic, though, a word of moderation is due. Like most or maybe even all contributors to the punishment debate, I think that our *current practices* of punishment are, as a whole, nowhere near justified. And more to the issue at hand, if expressivism is how we are to justify punishment, then we have to change a huge chunk of current punitive practice. (I have in mind in particular the regime of punishment in the United States, but the point is more general.) So the question on the table is not whether expressivism can justify *our current system* of punishment. Rather, the question is the bigger, more enduring philosophical question of whether punishment is *ever* justified. Skeptics say that punishment is never justified, and of course in mounting their defense, they target expressivism among other theories. The aim here is to show that, given the arguments so far presented, the skeptical claim is weak, since it is at least viable to say that *some* punishment is warranted on expressivist grounds. The anti-expressivist criticisms with the most currency simply do not prove decisive.

I. THE EXPRESSIVIST THEORY OF PUNISHMENT

The core position common to all expressivist theories of punishment is this: punishment is permissible at least in part because it is the only, or the best, way for society to express condemnation of the criminal offense. The ''in part'' part of this formulation allows that we might combine other elements with this element in the overall story about why punishment is permissible, something I will do below. However, the element featured here, the expressive element, is itself meant to have irreducible normative force. Expressing condemnation is warranted on this account not because it, in turn, leads

to good consequences like prospective offenders being deterred. Nor is expressing condemnation warranted on this account because it is what the offender *deserves*, where desert does the real justificatory work. The core position of the view is exhausted by the principle that punishment is permissible at least in part because that is the only, or the best, way for society to express condemnation of the offense. There is nothing here about deterrence or desert or fairness or any other justification for punishment.[1]

So the basic expressivist idea is that punishment can be appropriate when society has run out of non-punitive ways of expressing its disapproval of the criminal behavior. We wrote it into law. We posted signs. We discussed it publicly. We shouted and protested. Still we reserve one final, vivid way of expressing ourselves: punishing those who violate our laws. This is a necessary way of reaffirming our values and the dignity of the victim.[2] The criminal sentence, and the justice system's execution of that sentence, is one elaborate statement that we *really* don't like what the offender has done.

This way of putting it blurs an important line that separates the two main variants of expressivism: communicative and non-communicative expressivism. The communicative branch says that the point of punishment is for society to communicate with the offender (and perhaps others). Society may, on this account, transmit its disapproval with some further aim, such as that the offender be educated or find repentance, remorse, reconciliation, or reform.[3] Or there may be no further aim to the communication beyond the

[1] Many classify expressivism as a form of retributivism. See, e.g., Jean Hampton, 'The Retributive Idea', in Jeffrie Murphy and Jean Hampton, *Forgiveness and Mercy* (Cambridge: Cambridge University Press, 1988), 111–161; Thaddeus Metz, 'How to Reconcile Liberal Politics with Retributive Punishment', *Oxford Journal of Legal Studies* 27 (2007), 683–705; and Bill Wringe, 'Must Punishment be Intended to Cause Suffering?' *Ethical Theory and Moral Practice* 16 (2013), 863–877. My characterization of expressivism is consistent with that taxonomy, so long as it doesn't reduce the expressive element to some other element.

[2] Hampton, *Ibid*.

[3] R.A. Duff, *Trials and Punishments* (Cambridge: Cambridge University Press, 1986); Duff, *Punishment, Communication, and Community* (Oxford: Oxford University Press, 2001); W.A. Miller, 'Mr. Quinton on 'An Odd Sort of Right', *Philosophy* 41 (1966), 258–260; Miller, 'A Theory of Punishment', *Philosophy* 45 (1970), 307–316; Andrew Oldenquist, 'An Explanation of Retribution', *The Journal of Philosophy* 85 (1988), 464–478, at 468; John Tasioulas, 'Punishment and Repentance', *Philosophy* 81 (2006), 279–322, at 284.

offender, and broader society too, *getting the message*.[4] Either way, though, communicative theories say that punishment is justified as a method of communication.

Some consider the communicative version of expressivism, particularly Antony Duff's, the most important form. For example, David Boonin judges it the most ''prominent and thoughtful'' version of expressivism.[5] It certainly appears to be the more popular branch of the view and the branch taken most seriously by expressivism's opponents. Nevertheless, it doesn't work.

To see why, let's begin with a critique of the view that has only limited reach. Critics often point to cases of an unrepentant or defiant offender, one who won't accept the legitimacy of our condemnation, or of an already-repentant offender for whom communication of reason to repent seems like overkill.[6] Such cases speak against communication theories that say that punishment must *successfully achieve* repentance: in these cases communication via punishment makes no progress towards the goal of repentance. However, these cases do not impact other communicative theories. Some communicative views do not require the *achievement* of repentance, so long as the communication of punishment *aims at* repentance.[7] And again other communicative theories just require communication, having no further goal at all. For example, Primoratz positions his view as valuing communication intrinsically, rather than as a means to an end.[8]

Nevertheless, while these more relaxed communicative views might escape the problems of the already-repentant and unrepentant offenders, all communicative theories are still vulnerable to a more basic problem that is neglected in the literature. This is the problem of the *uncommunicative* offender. Communication is a multilateral relationship: all relevant interlocutors must participate for it to count as communication. If I call you on the phone and start talking, but a

[4] Uma Narayan, 'Appropriate Responses and Preventive Benefits: Justifying Censure and Hard Treatment in Legal Punishment', *Oxford Journal of Legal Studies* 13 (1993), 166–182; Igor Primoratz, 'Punishment as Language', *Philosophy* 64 (1989), 187–205, at 199–200; Andrew von Hirsch, *Censure and Sanctions* (Oxford: Oxford University Press, 1993), at 10.

[5] David Boonin, *The Problem of Punishment* (Cambridge: Cambridge University Press, 2008), at 172.

[6] Thom Brooks, *Punishment* (London: Routledge, 2012), at 119; Tasioulas *op cit.*, at 298; Michael Tonry, 'Obsolescence and Immanence in Penal Theory and Policy', *Columbia Law Review* 105 (2005), 1233–1275, at 1266; von Hirsch *op cit.*, at 10.

[7] Duff, *Punishment*, op cit.; Tasioulas *ibid*.

[8] Primoratz *op cit*. And compare Narayan *op cit.*, at 174; von Hirsch *op cit.*, at 9–10.

system error has prevented us from hearing each other, then we are not communicating, despite our best intentions. The uncommunicative offender is a similarly incapacitated conversation partner. He cannot engage in communicative exchange with society. He is not receiving our call. He does not get the message. His remorse or even bare recognition of the message is not an achiev*able* goal for punishment, in which case communication is a non-starter.

Perhaps he is a psychopath. Perhaps he considers himself above social rules, maybe guided by the devil's hand. Or maybe God's hand: perhaps this offender only listens to people he believes are acting as the voice of God, and those in the justice system are not among the godly. Of course, he doesn't have to be such a fanatical character. Perhaps while undergoing his trial he underwent a horrible experience that led to post-traumatic stress of such severity that he cannot help but tune out authorities' attempts to communicate with him. Or perhaps he is just being boneheaded, not recognizing that we are attempting to communicate with him; some criminals simply don't know how to pay attention or listen, and some can't grasp what we're trying to say. Whatever the backstory, the feature common to all uncommunicative offenders is that they cannot receive the message being sent. Note that the point is not that they *will probably not* get the message; it's that they *cannot* get the message. The problem of the uncommunicative offender is that if the sole justification for punishment is to communicate with the offender, then there is no justification for punishing any of these uncommunicative offenders.[9]

Now communication theorists can argue that as long as there are some offenders who *can* communicate, communication theory has answered the challenge of showing that *some* punishment is justified, for at least those offenders. That would be correct, as far as it goes. But there is still the problem that communication theory gets highly counterintuitive results. Some of the *non*-communicative offenders will be murderers, rapists, thieves, and others who we intuitively

[9] Note that this argument is focused only on ability to communicate. It does not speak one way or another about what to do if an offender were responsible at the time of the offense but later suffered from mental illness that diminishes responsibility after the fact (except trivially, in those cases where mental illness impacts ability to communicate). Perhaps it would be fitting to punitively express ourselves nonetheless in such cases; perhaps not. My personal sympathies are with a more forgiving version of expressivism, but I do not make an argument for that position here. I only claim that inability to communicate does not itself immunize offenders from liability to punishment.

think are paradigm cases of people who should be punished, regardless of their ability to communicate.[10]

The problem is actually worse than that. Because being uncommunicative removes this justification for punishment, any offender who wishes not to be punished has a perverse incentive to *become* uncommunicative. Why should I listen or otherwise attend to you, if refusing to communicate with you releases me from punishment? In this way it turns out that the communication branch of expressivism incentivizes murderers to ignore society's moral messages!

The communication theorist could argue, of course, that strictly speaking there are no uncommunicative offenders. Recall that these are not offenders who are merely *unresponsive* when they hear our complaint, for as we have seen, some forms of communication theory, namely those where punishment *reasonably aims* at communication, can accommodate such cases. Nor are we talking about people who best receive messages in a certain, hard-to-achieve way – communication theory can say that we just have to take the difficult steps required to reach them. The proposed problem case for communication theory is the person who *cannot* communicate with us, who cannot even receive our message. That is a peculiar kind of case.[11]

Nevertheless, it is a legitimately worrisome kind of case. Of course we only need one such case for it to be a problem for communication theory. But surely many offenders cannot hear the message that society might try to send with punishment. The boneheaded offender is a clear candidate for futile communication, being simply incapable of understanding what we are trying to express or even *recognizing that* we are trying to communicate through punishment. The communication theorist might balk here that all such offenders are merely *unwilling* communicators, not *incapacitated* communicators. I think that such a reply would display a lack of imagination about the many

[10] Some communication theorists, like Miller, *op cit.*, welcome this implication. Miller focuses on "imbeciles" who we do not think should be punished anyway. I take it that Miller's "imbeciles" are supposed to be people we think are fit more for psychological intervention than for criminal punishment. But these are not the only offenders who are uncommunicative, and the *other* uncommunicative offenders still, intuitively, must be punished.

[11] An anonymous reviewer suggests that communication might be one-way in this fashion: if I email a student information about an assignment, and she fails to check her email to receive the message, I have nonetheless communicated the information to her. I believe that this puts considerable stress on the ordinary understanding of "communicate," but if you disagree, there is a more fundamental point that neutralizes this reading of "communicate." What we are interested in here are cases where the message's recipient is *incapable* of communicating, not merely cases where the recipient does not take up the message. In the former cases, I suggest, there is no communication, even if there is in the latter.

ways people can be dense. But even if lack of will is the source of the communication breakdown, that doesn't stop the case from being a counterexample to communication theory. If communication is known to be impossible due to the unwillingness of the offender, then it's hard to see any value, intrinsic or instrumental, in trying to communicate. A similar response would be appropriate if a communication theorist were to object that the unwilling communicator case is unlike the telephone example, because it is the offender's internal state rather than an external technology that is impeding the communication. This is immaterial: if the point of punishment is to communicate, then punishment has no point when communication is impossible, regardless of the cause of the impossibility. And willingness won't even be at issue in other cases, such as the person with PTSD or some psychopaths who are psychologically blocked from attending to messages at all, whether or not they want to communicate. But surely commonsense will demand that some such offenders nonetheless be punished.[12]

Another possible solution for communication theory is to claim that, since we cannot know in advance of our attempt at communication who is and who is not capable of communication, we must at least attempt (punitive) communication with all offenders.[13] Again, I think that human limitation comes in many forms, and we sometimes can be quite confident that conversation partners are not hearing us at all. So I believe it is likely that there will be some almost-certainly-uncommunicative offenders who we nonetheless think should be punished. And there is another troubling aspect to this approach: it treats the actually uncommunicative offenders as mere means to the end of communicating with communicative offenders. That is something that is often thought, by critics of punishment, to be a fatal weak spot in, say, consequentialist theories of punishment that require punishing the innocent.[14] Effectively, this reply concedes that there is no *objective* warrant for punishing uncommunicative offenders, but because we cannot know for certain who is and who is not

[12] An alternative to standard communication theory is the view that the point of punishment is to communicate to society at large rather than to the offender. See Bill Wringe, 'Collective Agents and Communicative Theories of Punishment', *Journal of Social Philosophy* 43 (2012), 436–456. Such a view has an analog to the problem of the uncommunicative offender: if society for some reason cannot receive the message—say, it is a battlefield punishment that will never see the light of day or society has undergone some sort of collective PTSD—then the offender cannot be punished, which is unacceptable in some cases.

[13] I owe this potential solution to an anonymous reviewer.

[14] E.g., Boonin, *op cit*.

communicative, we have *subjective* reason to punish uncommunicative and communicative offenders alike. Punishing the uncommunicative offenders without objective warrant is the only way to punish the communicative offenders with objective warrant. This seems like a tremendous cost for communication theory. This way of treating people as mere means to a good end might be – *might be* – an acceptable cost if there were no other viable theory of punishment out there, but it should also push us to see if we can find an alternative theory that doesn't have the cost.[15]

The good news is that there is an alternative theory without this cost: pure, non-communicative expressivism.[16] To be sure, an ideal penal system will communicate, deter, rehabilitate, facilitate victim-offender reconciliation, and perform other useful functions, but on pure expressivism, what justifies punishment is not that it achieves these noble goals. Instead, the justification is that punishment is our way of expressing ourselves.[17]

On this view, it doesn't matter whether the offender hears us. He can ignore us all he likes, though we wish he wouldn't. Unlike communication, expression is a one-way street: even if you don't

[15] Victor Tadros, *The Ends of Harm: The Moral Foundations of Criminal Law* (Oxford: Oxford University Press, 2011), at Ch. 12, has recently argued that we may punish culpable wrongdoers as a means to stave off threats to future victims from other wrongdoers. The argument I make here does not oppose (or support) that claim. The end to which using uncommunicative offenders as a means aims on the proposed solution is not avoiding serious threats to future victims, but rather communicating with communicative offenders. I think it is much less likely that this can be an acceptable means to *this* end, than that it could be an acceptable cost of avoiding harms to future victims. Note also that while "treating people as mere means" is a phrase that gets operationalized in many ways, some of which are inconsistent with that term as used here, I take it that the way it is used here is unsurprising, and that the main point that harming someone without objective warrant is a substantial theoretical cost, stands regardless of one's take on how to best understand "treating others as mere means".

[16] Joel Feinberg, 'The Expressive Function of Punishment', in *Doing and Deserving: Essays in the Theory of Responsibility* (Princeton: Princeton University Press, 1970), 95–118; Hampton *op cit*.; Hampton, 'An Expressive Theory of Retribution', in W. Cragg (ed.), *Retributivism and Its Critics* (Stuttgart: Franz Steiner, 1992), 1–25; J. Kleinig, 'Punishment and Moral Seriousness', *Israeli Law Review* 25 (1991), 401–421, at 418; Metz, 'Censure Theory and Intuitions about Punishment', *Law and Philosophy* 19 (2000), 491–512.

[17] It is not a reason to reject pure expressivism that it fails to address offenders in "reciprocal" and "rational" ways, as Duff claims (*Punishment*, *op cit.*, at 79). We can require that punishment be addressed to the offender in a certain way, without saying that such communicative address *constitutes the justification* for punishment, just as we can require that punishment be public without saying that the publicity justifies the practice. Duff's other main reason for rejecting pure expressivism seems to be that if we are not engaging the offender with the aim of communication for her betterment, we are treating her as a mere means. As Hampton points out, if we understand "treating as mere means" as not respecting the value of the offender, then Duff's concerns are misplaced: we respect the offender's value, but we also reaffirm our moral commitments and the victim's value ('The Retributive Idea', *op cit*., at 144 n. 33). Duff also suggests that to not try to communicate with an offender who we *know*, with certainty, to be uncommunicative, is to treat him as "beyond moral redemption" (*Punishment,* at 123; cf. *Trials, op cit*., at 265–266). I think this is not quite right: it is to say that any such redemption isn't going to come from communicating with us. After all, such an offender is by hypothesis uncommunicative.

receive my call, I'm still talking. Because this is its distinguishing feature, pure expressivism avoids the problem of the uncommunicative offender. This makes it the superior expressivist view, the popularity of the communication wing of the theory notwithstanding. Accordingly, by ''expressivism'' I will refer to pure, non-communicative expressivism going forward.[18]

It is worth emphasizing that one-way expression of core commitments is quite common, which suggests that the value of pure expression cannot be dismissed out of hand. These one-way expressions cover the vast range of human values, from the profound to the mundane. We might spit on someone's grave[19] or address it solemnly, even when nobody is watching. We might in solitude approach a flag or memorial with reverence. An artist might create an expressive work just to destroy or hide it. We sing in the shower and hum in the car. We yell at the television to protest the referee's bad call during the big game. Some wear funky socks to express their quirky personality even when nobody sees them. Even the way we organize (or fail to organize) our desks can be a way of expressing ourselves. Again, then, given the thoroughgoing prevalence of pure expression, it's hard to simply dismiss it as unimportant.[20]

To turn more specifically to what *punishment* expresses, I have spoken of *disapproval* and *condemnation*, but expressivists could feature different attitudes. So while I will continue to use the language of disapproval and condemnation, note that if you find these attitudes problematic, you could substitute your preferred attitudes. That said, I use these broad categories precisely because they are so broad: they encompass moral rejection, dislike, disgust, resentment, repudiation, hatred, fear, and a host of other modes of moral disapprobation. They are also relatively uncommitted and conceptually shallow: they do not necessarily entail much else, such as an attribution of *blame*, for example. At the same time, the attitude of

[18] Reader, if you find experimental philosophy interesting (and if you don't, return your eyes to the main text now), it is also worth noting that recent research shows that in economic games, ordinary folk judge it best to dole out punishment even when it is not communicated to the offender. Here again pure expressivism does a better job of capturing commonsense intuitions than does communication theory. See Thomas Nadelhoffer et al., 'Folk Retributivism and the Communication Confound', *Economics and Philosophy* 29 (2013), 235–261.

[19] Metz *op cit*., at 495.

[20] These examples demonstrate that Hampton doesn't go far enough when she claims that expression of values only makes sense if there is someone else there to comprehend the message (*The Retributive Idea*, *op cit*., at 132). We make many choices that express our values even when nobody appreciates what we are ''saying.''

disapproval that is expressed in punishment does reflect two separate grounding judgments, on my construal. First, there is a negative judgment about the offender's action. Second and what is the basis for this negative judgment, there is the positive judgment that the victim has a significant moral status that the crime has violated. So on the view I'll be working with, the disapproval of the crime that is expressed in punishment is grounded in both moral rejection of the offense and reaffirmation of victims' rights. But whatever attitude we end up featuring, the key idea is that punishment is warranted because it is sometimes society's best or only mode of expressing certain important moral attitudes.

Finally, expressivists should not grant that every punishment that expresses how society feels is warranted, for there are many inappropriate feelings and many inappropriate expressions of appropriate feelings (quite independently of whether those expressions take a punitive form). A racist society is not, for example, justified in punishing interracial romance simply because that society mistakenly thinks that such romance is a crime against nature. Accordingly, for punishment to be justified two constraints must be satisfied: the attitude expressed must be a fitting reaction to the criminal behavior, and the punishment must be a fitting expression of that attitude.[21] These two fittingness standards will marshal reasons to have attitudes and to express those attitudes in certain ways. For example, they will include the boilerplate that expressivism only warrants proportionally punishing violations of just laws. How to cash out these reasons allows for various interpretive choices that we don't need to settle here. We might, for instance, say that the standards for fittingness are objective, holding for one perspective-independent standard; or we might go subjectivist, fixing the standards to some information available to the society at the time. (Relatedly, expressivists debate among themselves whether their claim that we may punish as a form of expression, is conventional or non-conventional.)[22] Similarly, we might say that a society is justified in its punitive expressions only if it has *met* the relevant standards; or we might say that it is justified so long as it is *making progress towards*

[21] Feinberg *op cit.*, at 118; Narayan *op cit.*, at §1.

[22] E.g., Hampton, 'An Expressive Theory', *op cit.*, at p. 3; Metz, 'Realism and the Censure Theory of Punishment', in Patricia Smith and Paolo Comanducci (ed.), *Legal Philosophy: General Aspects: Theoretical Examinations and Practical Implications* (Suttgart: Steiner, 2002), 117–129.

achieving adequate fit with proper standards. Because these variations do not impact the arguments that follow, I do not choose between them. I note them only to indicate that a plausible expressivism must tether legitimate punitive expression to some standards, and there are multiple ways of doing this.[23]

II. THE THREE OBJECTIONS

Now that we have a working conception of expressivism, we can turn to the three dominant objections to this view. Again, I will argue that although they keep getting recycled, they all fail to provide any compelling reason to reject the most plausible forms of expressivism.

A. The Excessive Expression Objection

Our first – and apparently most popular – objection says that punishment is *unnecessary* for expressing ourselves. Perhaps punishment would be justified if it were the only or best way for society to express its condemnation of the crime – perhaps this *conditional* is true. But punishment is not in actual fact the only or best way for us to express our condemnation – the antecedent is unsatisfied, according to the excessive expression objection. Instead of punishing, can't we also just *tell* offenders that we don't like their behavior, if the goal is simply to express ourselves?

Now Uma Narayan rightly observes that if we care about conveying different attitudes, mere verbiage is pretty weak.[24] It would be an inadequate expressive repertoire if all we could say was "We don't like what you did" for petty theft, "We *really* don't like what you did" for grand theft, and "We *really, really, really* don't like what you did…no, really, we mean it!" for murder. It is important, then, that the excessive expression critics call attention to a number of non-verbal but also non-punitive forms of expression. We can order the offender to engage in either mediation with his victims or some

[23] Importantly, all of these answers cash out fittingness in terms of either objective standards or subjective attitudes, but none of them reduce fittingness to any other retributivist idea, such as desert. Thus, this view is in two respects not redundant (see Brooks *op cit.*, Chapter 6).

[24] Narayan, *op cit.*, at 179.

sort of reconciliation project.[25] We can require the offender to get some moral education, not with the aim of harming him, but with the aim of communicating our moral views. And Nathan Hanna points out that we express our seriousness with a number of extra-punitive elements of the present criminal justice system, such as degree of investigative effort, procedural standards, and so on.[26] If we look outside of the formal criminal justice system, here too we find a range of non-punitive expressions, from hunger strikes to civil disobedience to violent protest to shunning to public mockery.

According to the excessive expression objection, these measures (and others) appear to jointly constitute condemnations that are at least as expressively adequate as punishment.[27] And since punishment by definition has the unique downside of doing real harm to the offender – beyond the harm of condemnation itself – and also is very costly to everyone else, alternatives to punishment are preferable to punitive expression.[28]

Again, this is a very popular objection, having been made by an impressive roster of commentators, including, in alphabetical order: Boonin, Thom Brooks, John Cottingham, Hanna, H.L.A. Hart, Narayan, T.M. Scanlon, Victor Tadros, and Bernard Williams.[29] But

[25] M.D. Adler, 'Expressive Theories of Law: A Skeptical Overview', *University of Pennsylvania Law Review* 148 (2000), 1363–1501, at 1424; Metz, 'Realism', *op cit.*, at 124.

[26] Nathan Hanna, 'Say What? A Critique of Expressive Retributivism', *Law and Philosophy* 27 (2008), 123–150, at 138–139.

[27] As Victor Tadros has stressed in private communication with me, the speaker herself taking on burdens, such as in a hunger strike, might sometimes be *more* expressively adequate than imposing punitive burdens on the offender. For more on this, see Hampton, 'An Expressive Theory', *op cit.*, at 15–17.

[28] A debate has erupted as to whether punishment not only harms the offender essentially, but also *intends* to harm essentially. See Boonin *op cit.*; Hanna *op cit*; Hanna, 'The Passions of Punishment', *Pacific Philosophical Quarterly* 90 (2009), 232–250; and Wringe, 'Suffering', *op cit*. I am dubious about the intentional component, roughly for reasons articulated by Wringe (cf. Duff, *Punishment*, *op cit.*, at 96–99). However, my argument does not depend on rejecting this component. Whenever I talk about how our expressive resources require us to harm, fans of the intention thesis can replace this with talk about how our expressive resources require us to intentionally harm.

[29] Boonin, *op cit.*, at 176–179; Brooks *op cit.*, at 118; Cottingham, 'Varieties of retribution', *The Philosophical Quarterly* 29 (1979), 238–246 at 245; Hanna, 'Say What?' *op cit.*; Hanna, 'Passions', *op cit.*; Hart, *Law, Liberty, and Morality* (Palo Alto, CA: Stanford University Press, 1963), at 66; Narayan *op cit.*, at 171; Scanlon, 'The Significance of Choice', in S. McMurrin (ed.), *The Tanner Lectures on Human Values*, vol. viii (Salt Lake City: University of Utah Press, 1988), 151–216, at 214; Tadros *op cit.*, at 103 and 109; Williams, 'Moral Responsibility and Political Freedom', *Cambridge Law Journal* 56 (1997), 96–102, at 100. Boonin (*op cit.*, at 177 n. 15) also attributes this objection to Samuel Scheffler, as well. But while Scheffler notes in passing that there are multiple ways to express reactive attitudes, his point is the separate observation that justifying the punitive way of expressing ourselves requires a holistic assessment of its ''other features,'' beyond the fact that it allows us to express ourselves. This point generates the featured question of Section 3 below. See Scheffler, 'Distributive Justice and Economic Desert', in Serena Osaretti (ed.), *Desert and Justice* (Oxford: Clarendon Press, 2003), 69–91, at 76.

the criticism is misguided. It does work against theories that claim that punishment is the *only* way to condemn *any* crime. However, expressivism need not be framed so extremely. As noted above, the question is not whether our *current* system of punishment is required for us to express ourselves. We are not even asking whether for *many* crimes expression requires punishment. In fact, we can stipulate for our discussion that there are many crimes for which disapproval surely can be expressed non-punitively. Expressivism would be a ludicrous non-starter that never saw print if it denied that some minor first-time offenders must be punished rather than verbally condemned, for instance. In other words, the critics are correct that we can *often* best express our disapproval in non-punitive ways. Nonetheless, the remaining, enduring, big question is whether, even if some crimes can be condemned in non-punitive ways, there are still other crimes for which punishment is the only or best way to express our commitments. I think this is quite plausible. Let me elaborate.[30]

The proper subset of crimes for which punishment is the *only* way of expressing our disapproval is composed of those crimes for which nothing less than the ''hard treatment'' of punishment will truly express our outrage. Different ways of expressing the same proposition carry very different pragmatic and emotive contents, and some propositions can only be expressed in certain ways. As Justice Harlan noted in *Cohen v. California*, even limiting ourselves to *verbal* expressions, there is a lot of difference in how we express ourselves: Cohen's expression, ''Fuck the draft'', carried emotive content that is hard to capture with other words. And sometimes words alone will not suffice. If you are truly in love, you can't in normal circumstances adequately express it simply by sending your lover a greeting card. If you really appreciate another's significant sacrifice, you can't always adequately express that merely by giving her a ''thanks'' as you cross paths on the street. And deep desires to achieve career goals are not expressed simply by tweeting them in 140 characters. In all three cases, a complex and limited set of *actions* must be taken to fully express your attitudes. Some of those actions involve *saying*

[30] One striking feature of the literature here is that, though many cite Joel Feinberg's (*op cit.*) treatment of expressivism, many of the excessive expression critics neglect his arguments that run in the same direction I am about to head. I hope that my elaboration buttresses Feinberg's claims, which do sometimes move quickly.

more than a summary report of one's attitude. And some of those actions involve something other than speech acts. One expresses one's desire to achieve a career goal – as opposed to a mere career fantasy – by *working for* that career goal. And when it comes to love, famously, talk is cheap. Putting that love into action is where the action is.

Now when it comes to punishment, there is room for debate over what exactly it takes to truly express our profound disapproval of any given crime. But it is too facile to suggest, as Boonin does, that for all crimes we can simply "issue an official statement of denunciation."[31] And even recognizing that sometimes action is called for is insufficient, too. For instance, Scanlon rejects expressivism by saying this: "Insofar as expression is our aim, we could just as well 'say it with flowers' or, perhaps more appropriately, with weeds."[32] This recognizes that sometimes we must express ourselves with actions rather than words, but weeds are clearly insufficient to express fitting outrage in many cases. We simply have a limited range of expressive options for our highly complex attitudes.

Note that the claim here is not that we *could never* have a more expansive expressive range; rather the claim is that we *do not* have the ability to express everything we need to say with weeds and denunciations. I will elaborate on this point shortly, but to make the discussion more concrete, first consider the moral outrage we feel upon considering the horrific abuses committed by the violent Cleveland kidnapper, Ariel Castro. The sheer destructiveness of the things he did to the one young woman and two girls he kidnapped, along with the baby he coercively fathered with one, resist comprehension. It would not begin to convey the ordinary person's fitting disgust for a court to respond to his crimes by simply saying, "Clevelanders vehemently denounce your actions." Nor would a delivery of weeds suffice. In fact, very little can even *approach* an adequate expression of our disgust. Action must be taken, and the only action that begins to give voice to our revulsion is punitive hard treatment. Punishment is the *only* way to approach the *full* expression of this reaction, just like acts of kindness and romance are sometimes the only way to express certain forms of love.

[31] *Op cit.*, at 177.
[32] *Op cit.*, at 214.

Moreover, the character of our response is very precise. To express ourselves, weeds, denunciations, protests, and the like are simply inadequate. And we feel cheated if the offender dies of a heart attack before we can punish him. We don't want bad things to happen to just anybody, and we don't want them to just *happen* to Castro. Adequate expression requires us to impose bad things on him.[33]

A moderate opponent of expressivism might grant this and agree that it is too strong to say – as those more extreme opponents of expressivism cited above do – that expressing ourselves *never* requires punishment. She might grant that nothing short of punishment would allow us to adequately express sufficient moral outrage in the Ariel Castro case, and certainly agree that weeds and verbal censure are not enough to express ourselves. This, of course, would concede our debate to expressivism, for our question is whether punishment is ever the only adequate mode of expression, and here we'd have one case where punishment is granted to be necessary for expression. But it is still worth asking: what about more ordinary crimes, like cheating on taxes or petty theft or minor motor vehicle violations like parking on an expired meter? Surely we don't feel so much disgust at such actions that we have to express ourselves by harming offenders.

While it is true that these actions considered in a vacuum often do not disgust us, that is too narrow an evaluative scope. Such actions violate conventions involving private property and social cooperation, and to the extent that these conventions carry moral weight – they represent our commitment to sociality with one another and help us avoid harming each other, among other things – we do disapprove of violations such that, in many cases at least, it would be insufficient for the justice system to respond with a mere denunciation. It needs to respond with action that harms the offender, even if only minimally (such as a fine or compelled community service), if it is to truly express how seriously we take the

[33] Sure enough, soon after before Castro died in jail of either suicide or autoerotic asphyxiation (the facts are apparently elusive), reports emerged that this kind of death, by being an evasion of a socially-imposed punishment, shortchanged his victims. See, e.g., F. Brinley Bruton, 'Ariel Castro's Death Is 'Last Slap' to Victims' Faces, Psychologist Says,' *'U.S. News.' on NBC News*. http://usnews.nbcnews.com/_news/2013/09/04/20320005-ariel-castros-death-is-last-slap-to-victims-faces-psychologist-says?lite. Accessed Sept. 17, 2013. (Of course, such news reports also run together questions of psychological coping with justification, and, within that, different forms of both. They are of limited evidential value.)

cooperative arrangement for shared living that is our joint ongoing project.

Of course, we often *don't* think that punishment is warranted for actions that technically break the law. This should ring true for everyone who has tried to talk their way out of a parking fine, but it also applies in more serious cases, such as desperately poor parents trying to feed their children through theft or someone driving a woman in labor to the hospital in an unregistered, but safe, car. Here the expressive theory of punishment is quite strong. In all of these cases, the expressive theory aligns with intuition: punishment is inappropriate precisely because it would fail to express fitting moral attitudes toward such actions. No *wrongful* crime was committed in such cases. We feel that an excusable violation was committed; therefore an exercise of prosecutorial discretion is warranted.[34]

This gets us to the other way that non-punitive expressions are insufficient: in many cases, even if punishment is not the *only* way that we can express ourselves, it is still the *best* way. This is particularly true for certain *criminals*, namely repeat offenders. We might let you off without punishment for your first or second minor crime, simply registering our disapproval with an explicit censure from the court. But when you repeatedly re-offend, explicit censure seems expressively inadequate. We rightly feel more frustration, resentment, and disappointment, and the only way to express our *heightened* disapproval will be some manner other than the statement that was used for the first offense. To express ourselves we must respond

[34] Boonin claims that in these cases, expressivism has a serious problem, for he thinks that "a successful solution to the problem of punishment must justify the right to punish in such cases" (*op cit.*, at 179–180). To fail to do this is to fail the project at hand, he claims, generating a novel objection to expressivism: we simply haven't justified punishment if we haven't fully justified punishment for *all* violations of the law. I cannot see why Boonin sets this strict standard, though. He is very deliberate in the early going of his book to lay out adequacy conditions for solving the "problem of punishment." However, in this deliberate portion of his argument he does not make the strong claim that a good solution to the problem of punishment must show that *every* violation of the law must be punished. He just sneaks the strict standard in later (page 54), in the process of discussing certain cases where not punishing an offender would be counterintuitive, which of course is a legitimate argumentative move but a different one than the one we are considering here. And he *shouldn't* commit to this strict standard, for the strict standard is false. Nobody thinks that every action that technically violates a just law should be met with punishment. That's one reason why we embrace prosecutorial discretion (Douglas N. Husak, 'Why punish the deserving', *Noûs* 26 (1992), 447–464, at 449–450). What we want is a theory that justifies punishing not *every* crime, but every *sufficiently wrongful* crime. Expressivism does this, since the very principle on which it is founded is that punishment is justified as a way of expressing our *moral disapproval*. (Relatedly, see Metz, 'Censure Theory', *op cit.*, for the argument that expressivism is uniquely strong in securing both the *pro tanto* reason to punish all the guilty and the proportionality of punishment).

to the repeat offense more ''loudly'' (proportional, of course, to the nature of the offense – the repeat petty thief does not need a louder response than the first-time kidnapper). Nothing short of action, in the form of punishment, will be optimal at that point, to express how seriously we feel about your multiple offenses. To be sure, this present point isn't meant to prove that punishment is *justified* for repeat offenders; rather, it is to claim that it is our only remaining expressive resource available.

To summarize, then, the excessive expression criticism misses the fact that our expressive resources are actually quite precise and limited. If it were simply a matter of vocalizing dissatisfaction then sure, mere denunciation would always do. But it is often more than that. We want to express severe moral disapproval. We want to reaffirm the sacred moral status of the victim. We want to register that our cherished moral and legal system is not to be flouted. And to express our profound dissatisfaction, we sometimes have no other means but to punish the offender. Maybe punishment is unjustified, but if so, it's not because punishment is never a uniquely apt way of expressing ourselves. It sometimes is.

What's curious about the excessive expression objection is that it makes a bold, universal claim that punishment is *never* needed or optimal for adequate expression. But of course proving a universal negative is pretty hard, and deploying a handful of examples of adequate non-punitive expression, as the critics often do, hardly gets us there. This is all the more surprising given that so many people clearly feel that they have no other choice but to express themselves punitively.

At the same time, the excessive expression criticism captures a deeper critical point that deserves recognition. Feinberg points out that even if punishment is an essential part of our expressive vocabulary, perhaps we could hypothetically add some arrows to our expressive quiver, devising some new, elaborate ritual to express our disdain without imposing so much hard treatment.[35] If adequate expression is sometimes limited to the punitive, this may be because our expressive conventions are not creative enough; we could, if we tried, come up with a better language than punishment.

[35] *Op cit.*, at 115–116. Cf. Tasioulas *op cit.*, at 289.

But while it is worth pursuing, this possible language is not a basis for a compelling objection to expressivism, for two reasons. First, as Feinberg observes, it is unclear whether this new ritual is just ''idle fantasy'' or if instead our expressive capacities actually have room to grow. But second, and more decisively, Feinberg's hypothetical doesn't really get us to the heart of the question. Our question is whether *we* are justified in punishing. It is not enough to show that some *other, expressively more advanced* society can do the job with non-punitive expressions of disapproval. As soon as the critic grants that it takes some *hypothetical* ritual to get us a non-punitive improvement over our current, merely punitive expressive resources, she acknowledges that our *current* expressive resources *are* limited to the punitive. And so this move effectively grants that punishment is not an excessive form of punishment for *us*, because we have no non-punitive alternatives. (Perhaps, then, punishment is permissible only so long as we simultaneously seek to improve our expressive repertoire.) In this way, the critic and the expressivist might agree that *if* we had some non-punitive alternative to punishment that was expressively adequate, it would be unjustified to engage in punishment.[36] But of course these expressively adequate fantasy alternatives are not in hand according to the hypothesis that the fully adequate non-punitive expression *is* fantasy.

This is why it is not enough for the critic to simply claim that, if conventional modes of expression limit us to expressing ourselves via punishment, we should just adopt a better convention.[37] In addition to not being grounds for rejecting expressivism for our kind of society, we also need a richly detailed picture of what that alternative mode of expression would look like. We are obviously far from perfect, but it's a pretty significant fact that we've been working on our modes of expression for millennia. If the critic concedes that we don't *yet* have an adequate non-punitive mode of expression for every possible crime, then given that history, the burden is on the critic to explain what our new-and-improved expressive apparatus will look like, how it would

[36] Similarly, as Jean Hampton allows, the expressivist can agree with the critic that *if* there were a non-painful punishment, that would be preferable to the painful method, too. 'The Retributive Idea', *op cit.*, at 126; 'An Expressive Theory', *op cit.*, at 16–17.

[37] E.g., Heather J. Gert, Linda Radzik, & Michael Hand, 'Hampton on the Expressive Power of Punishment', *Journal of Social Philosophy* 35 (2004), 79–90, at 86–87; Hanna, 'Say What?' *op cit*. Whether this should be framed in a conventionalist manner is an outstanding issue, but let's put that aside here. See Metz, 'Realism', *op cit*.

cover *every* crime, what the novel costs of it might be, and what the costs in transitioning to it might be. Moreover, that new release valve must be as psychologically compelling and expressively adequate as our current apparatus, and it is worth emphasizing, again, that the current apparatus expresses an enormously complex and wide-ranging set of attitudes: not merely that we don't like the offender's behavior, but that we reject the non-cooperativeness it represents, that our attitudes toward her have changed, that we feel affectively "hot" moral outrage the phenomenology of which is *directed at the offender* rather than, say, directed at ourselves, that our relationship with her has changed, and on and on and on. So it is not enough to simply say *that* there is a better way of expressing ourselves; it must be shown *what* that way is and whether it is available here and now.[38]

If these remarks are on target, the critics have fallen well short of making good on the excessive expression objection. Consider Bernard Williams' way of putting it: "'The idea that traditional, painful, punishments are simply denunciations is incoherent, because it does not explain, without begging the question, why denunciations have to take the form of what Nietzsche identified as the constant of punishment, 'the ceremony of pain.'"[39] That's it. That's all Williams says by way of dismissing expressivism, and it is representative in its assumption that expressivists shoulder the defensive burden of explaining why expressing some of our attitudes must happen via punishment. I hope to have demonstrated in this sub-section that the opposite is in fact the case. The critics are right that punitive expression is extraordinarily costly, not only to the offender but to the rest of society as well, and that should push us to seek less costly modes of expression. But then, because we haven't yet found a better, less costly way of expressing ourselves, the defensive burden is on the excessive expression critics to

[38] One anonymous reviewer worried that appealing to our current expressive limitations may be an illegitimate move on my part, since the question we are asking in this discussion is one of ideal theory rather than current practice, namely "Is punishment ever justified?" (rather than "Is our current regime of punishment justified?"). To alleviate this worry, we should distinguish between idealizing regimes of punishment and idealizing those who do the punishing. The question for our discussion is "Are we ever justified in punishing?" where "we" refers to people like us—this is one variation on the question "Is punishment *ever* justified?" As the main text hopefully indicates, I'm happy to concede that for some other creatures, or for humans in a different expressive environment, or under some other idealized conditions, punishment may not be justified. But the anti-expressivists claim more than this; they claim that *we* are not justified imposing any regime of punishment on expressive grounds. What our discussion is taking up, then, is the question of whether we—creatures like us, with our expressive limitations—can impose *some* punishment, on expressive grounds.

[39] *Op cit.*, at 100.

articulate a new, non-punitive language complex enough to capture everything currently captured in punishment and to show that it has been available to us all this time. Once we have an answer to this question, expressivists *then* can give up punishment. Until then, when it comes to the basic question of whether we should find some non-punitive way to express our outrage, it is fair for expressivists to simply observe that punishment is all there is in some cases.

B. The Promiscuous Expression Objection

Just as there are some violations of the law that warrant no disapproval, there are some *legal* actions that *do* merit moral disapproval. There is infidelity and duplicity, overt racism and sexism, and just plain jerky behavior. Much of this falls short of being illegal. But if we disapprove of such actions, and if punishment is required for us to express our disapproval, then it seems that expressivism must license punishment as a response to such actions, even though no law has been broken. Boonin uses this point to argue that expressivism is committed to punishing the legally innocent, an unacceptable implication.[40] Brooks takes the problem in another direction, arguing that whatever principle – retributivist desert, Brooks speculates – the expressivist uses to narrow the scope of relevant disapproval to cover only illegal actions, that principle will *itself* be the real justification of punishment, making expression irrelevant.[41]

One response to this objection is to say that criminal law is only concerned with a limited scope of wrong action, so-called "public wrongs."[42] For example, perhaps it is only concerned with those wrongs that injure the broader community or violate its fundamental values or simply are of concern to the broader public. This response, then, makes a broad claim about what justifies criminalization and combines it with the principle that the set of criminal acts overlaps with the set of actions that may permissibly be punished, in order to narrow the scope of when we may punitively express ourselves.

A more narrowly focused response to the promiscuous expression objection, and therefore perhaps a less onerous bit of theory, is that

[40] *Op cit.*, at 180.

[41] *Op cit.*, at 107–109.

[42] E.g., Husak, *Overcriminalization: The Limits of Criminal Law* (New York: Oxford University Press, 2008); Husak, 'Retributivism *in extremis*', *Law and Philosophy* 32 (2013), 3–31, at 26.

this objection conflates expressivism's theory of what *justifies* punishment with the theory of when someone is *liable* for punishment. That is, expressivism is only one part of a comprehensive theory of how punishment is *permissible*. Permissibility requires both that the harm involved in punishment have some justification, that is, some sort of value or point, and that the person harmed be liable for some reason. The expressive component provides the justificatory part: the point of punishment is to express moral disapproval. If expressivism adopts some theory of liability for intentional harm that makes criminality a condition of liability to harm, and then combines it with the expressivist principle for what justifies punishment, expressivism will only license punishing the legally guilty. It will not license punishing the innocent, and its justification for punishing the guilty will not collapse into some other theory's justification.

Any such distinction between liability and justification for punishment will do, but perhaps the point can be made perspicuous with an example.[43] In my estimation, the best liability criterion is that in the criminal act, the offender *forfeits her right* not to be punished. In taking on this theory of liability, I am of course taking on more theoretical baggage that exposes the overall account to more angles of attack. For instance, Boonin rejects forfeiture-based views of what justifies punishment, and we might expect him to think that the concerns he has about these theories might also extend to any theory that claims that forfeiture of rights renders us liable to punishment.[44] The details of this extension would need to be delivered, of course, but as it turns out we don't need to wait: Stephen Kershnar and Christopher Heath Wellman have given decisive replies to Boonin's objections to forfeiture-based justifications of punishment, so we can bracket this question.[45]

But another question is not so easily bracketed. By combining expression-as-justification and liability-by-rights-forfeiture, I deviate from those, such as Kershnar and Wellman, who think that forfeiture of rights is by itself sufficient for punishment's permissibility. So it is

[43] Expressivists generally focus on the justificatory power of expression and neglect the question of liability. For instance, in a whole chapter on the value of expression, Duff only spends one sentence to suggest that liability is separate from justification and happens in the criminal act itself. See *Trials*, *op cit*., at 255.

[44] *Op cit*., at 103–119.

[45] Stephen Kershnar, 'The Forfeiture Theory of Punishment: Surviving Boonin's Objections', *Public Affairs Quarterly* 24 (2010), 319–334; Christopher Heath Wellman, 'The Rights Forfeiture Theory of Punishment', *Ethics* 122 (2012), 371–393.

worth noting that this approach is also inadequate: simply because the offender has forfeited her right not to be jailed, this does not by itself render it permissible to jail her. A city's right of eminent domain means that it doesn't violate a resident's rights to put a road where her house is, but that does not render the action permissible. Exercising eminent domain permissibly also requires that there be some good reason for harming the resident in this way – the public good, say, rather than shortening the Governor's commute. Similarly, to impose the hard treatment of punishment on someone permissibly, we need to not only avoid violating her rights, but also to impose the harm for some good reason – we need both liability and justifiability, and rights-forfeiture only takes care of the first half of that equation. So utilizing the principle that we cannot harm someone for a bad reason or no reason, it is problematic in the case of punishment if the only point of punishing the offender is that the judge gets sadistic pleasure out of it or profits from her investment in private prison companies or simply tossed a coin that came up on the unfortunate side, even if the offender has forfeited her rights.

In short, a strong theory of punishment must provide both a theory of what the value is in punishment and a theory of who is liable for punishment. Once the liability side of the formula is taken care of, perhaps by the forfeiture-of-rights view, expressivism can give us the value of punishment. In so doing, it courts neither punishing the innocent nor irrelevance.

C. The Many Messages Objection

Expressivism claims that punishment is our way of collectively expressing ourselves. But what message, exactly, do we send? Hart, A. J. Skillen, and Brooks argue that pluralist societies will not embrace one single reaction to the same crime.[46] Factions will want to express multiple, jointly inconsistent messages, and it is unclear how one punishment can do this.[47] We don't have to look far to find concrete examples of criminal laws and sentencing guidelines that privilege one faction over others. In the United States, for instance, sentencing

[46] Hart, *Punishment and Responsibility* (New York: Oxford University Press, 1968), at 171; Skillen, 'How to Say Things with Walls', *Philosophy* 55 (1980), 509–523, at 520–521; Brooks *op cit.*, at 113–114.

[47] Sometimes this is framed as the offender not receiving one clear message or the listeners misinterpreting society's message (Brooks *op cit.*, at 109–110). These concerns only apply to expressivism's communicative wing, which we hived off in Section 1.

guidelines establishing punishments that are harsher for possessing rock cocaine than for possessing an equal amount of powder cocaine have been embraced by some (Congress) but roundly rejected by others (rational people). How can we say that society expresses itself in its controversial punishments of such crimes?

We have at our disposal at least three expressivist replies to this concern. Any one of them suffices, so I will not choose between them here.

The first solution is to let a certain kind of idealization constrain the actual society's expression. For example, we might say that the justified punitive expression is the one that would be arrived at under conditions of full information and rational deliberation. Brooks thinks that this reduces expressivism to retributivism, because it makes punishment a matter of what is "fittingly deserved."[48] But that reduction is not the most charitable construal of the idealization response. What this response does is make punishment a matter of legitimate expression, and it makes legitimate expression depend on a particular standard for arriving at reactive attitudes. There is nothing here about punishment being deserved. Rather this is a story about certain attitudes being relevant. That said, this view does have a limitation. To avoid the many messages problem, the idealized society must enjoy a lack of substantial disagreement in the relevant domain of judgments. It is not clear that in a pluralist society this will obtain; it *might*, but then again it might not.

A second solution is to return to a question left open above: what standards of fittingness decide what counts as a legitimate expression? If we are comfortable with the idea that there is some maximal set of objectively just laws out there, and some maximal set of fitting attitudes to have in response to violations of such laws, then we can say that expressions of only *those* attitudes may be justifiable.[49] Considered in that light, perhaps it is misleading for expressivists to say that we are justified in expressing our moral outrage via punishment. Rather, it would be more accurate to say that we are justified in expressing our moral outrage via punishment when that outrage is objectively correct.

The third response limits itself to a pragmatic division of expressive labor. On this view, our private substantive moral judgments are not what we are expressing when the state expresses itself on our behalf. In those

[48] *Op cit.*, at 111.

[49] Metz, 'Realism' *op cit*.

public expressions, we defer to certain authorities – the legislature, the judiciary, the penal system, the electorate, etc. – to express a collective judgment for us. These authorities are not *moral* authorities in the sense that they have more moral expertise than the rest of us. Nor do they have authority over our private reactions. Rather, they have *institutional* authority – we have arranged society so that they are the segment of our population that is allowed to speak for us on these matters even when, privately, we disagree. This institutional authority also helps expressivism avoid the problem that for many crimes – think of the enormously complex details of tax law, for example – society at large simply doesn't know what is illegal and therefore cannot have any view about violations to express.[50] This is avoided by delegating that knowledge and expressive responsibility to certain authorities, such as tax lawyers.

If this reply were to claim that our delegates must express our personal moral judgments, it would clearly be false. The U.S. justice system fails to express many citizens' commitments when it punishes rock cocaine more harshly than powder cocaine. So the deference view is to be rendered another way. We do defer to others in society to establish our collective judgment on what punishment is appropriate, just as we defer on the question of where to put a new school or highway. And we accept that a range of what we deem regrettable errors will be the price we pay for an overall *system* of collective action, coordination, and responsibility to each other. And this deference could be enough to establish an overlapping consensus of punitive expression.[51] It should go without saying that as we grudgingly accept the imperfections of the system, we should also push for progress, trying to change the imperfections. But if the system itself is sufficiently sound, we can have substance-based disagreement about what is ideally expressed but procedure-based agreement about what must be expressed.

Again, I believe that any of these three solutions is sufficient to answer the many messages objection.

III. THE BIG QUESTION

If what I have argued for so far is correct, then the three most popular objections to expressivism fail, at least when rendered as they have been rendered in the literature. However, that does not

[50] Brooks *op cit*, at 114.

[51] Cf. Primoratz, *op cit*., at 205; Brooks, *op cit.*, at 113.

mean that expressivism faces no obstacles. The biggest question is whether its positive justification for punishment is sufficiently strong.

In punishing, we knowingly harm people, a course of action that is clearly forbidden under normal circumstances. So even if there are no problematic downstream implications of punitive expression, such as punishing the innocent, why, we may ask, do we have the right to express ourselves at all, when the cost is so high? Even friends of expressivism have judged this to be a question that has not yet been satisfactorily answered.[52] Moreover, sometimes punitive expression looks like ''rubbing it in,'' as when we punish a grief-stricken parent who negligently let his child die because he did not secure the child's car seat.[53] Why punitively express what everybody, particularly the offender, already painfully knows? More basically, expressivism seems to violate a basic consequentialist constraint that the protection of citizens and more generally securing of serious benefits is a necessary condition on punishment being justifiable, in which case pure expressivism, in being loosened from any consequentialist constraints, looks unjustified when stacked up against the substantial psychological, economic, and social costs involved.[54]

There is an important truth in this concern. If simply giving honest voice to our feelings is the only value in punishment, that's a pretty weak justification for imposing such costs.[55] I suspect that this, rather than the three objections above, best explains why expressivism hasn't gotten more uptake: people just can't believe that we could do the terrible and costly things involved in punishment simply because it's our way of honestly expressing our feelings. If that's all there is to punishment, then maybe repression is the best we can do.

I don't here offer a decisive story about how expressivism can deal with these consequentialist concerns, but I do want to sketch three viable expressivist replies. This sketch is meant to point the way forward. Whether the claims will ultimately carry the argument depends on some questions left open here.

In order to give expressivism a fair hearing on this issue, recall first that the *present* system of punishment may well be much more

[52] Metz, 'Censure Theory', *op cit*., at 512.

[53] Cf. Boonin *op cit*., at 175.

[54] Husak, 'Retributivism', *op cit*.; Metz, 'How to Reconcile', *op cit*.; Phillip Montague, 'Recent Approaches to Justifying Punishment', *Philosophia* 29 (2002), 1–34, at 19; Narayan *op cit*., at 179–180; Scanlon, 'Giving Desert Its Due', *Philosophical Explorations* (online prepublication).

[55] E.g., Hart, *Law, Liberty, and Morality*, *op cit*., at 65; von Hirsch, *op cit*., at 12.

costly than would be warranted on the expressivist framework. The better way of asking the question that animates the punishment debate is, again, whether there is *any* crime, such as the most horrible kidnapping and murder, for which *some* punishment, even perhaps a moderate one, would be justified on expressivist grounds, even though that does incur some costs.

The most popular answer from expressivists has been that if we are going to meaningfully *outlaw* criminal wrongdoing, we must *express* our disfavor of that wrongdoing (and as we have seen, the way to do that is through punishment). But there are at least two ways of framing the claim that criminalization implies denunciation. The law-based version of this claim is that necessarily, criminalizing something entails censuring it.[56] On such an account, punitive censure signifies that our criminal laws are valid, such that without such validation, we literally *have no* crime.[57] The alternative, value-based version of this claim is that if we truly *care* about the law, we must express our disapproval of its violations; that expression is the only way to avoid *betraying our values*; and that punitive expression is necessary for even *having* or *honoring* values.[58] I'm sympathetic to a certain version of the value-based claim. And it gets us to the first of the three replies to the consequentialist concern.

That first reply is that it *is* important to express our core values, and the criticism at hand underestimates the value of this expression, at least if it is going to capture the entirety of ordinary moral thought. Expressing core values might not be the only thing that matters, but it is arguably one thing that matters, which the consequentialist objection simply denies without further argument. Obviously converting *arguably* into *decisively* is a task that exceeds the scope of our argument and that has been more ably discussed elsewhere.[59] But note that this is more than just being honest; this is publicly, officially, and explicitly recognizing and establishing principles for governing ourselves. In this way, we can conceive of punishment as a practice that constitutes an extended expression of

[56] Duff, *Punishment*, *op cit*., at 28.

[57] Feinberg, *op cit*., at 104; Primoratz *op cit*., at 196–197.

[58] For the first two claims, see Duff, *Trials op cit*., at 236; for the third see Oldenquist *op cit*., at 467 and 471.

[59] See Metz, 'How to reconcile' *op cit*., for one way in which this expressive approach can be expressed in terms consistent with liberalism, by acting *for the sake* of rights even when not acting to *protect* rights.

what we care most about. Failure to express ourselves on these matters is effectively a failure to publicly codify a collective moral identity. It arguably is a failure to *constitute* a collective moral identity. It is also failure to express the victim's value: punishment in part acts on behalf of the victim, reaffirming her full dignity in the light of a harsh and misguided attempt to dominate her.[60] So perhaps Duff goes too far to claim that failure to punish *betrays* our values. And it certainly seems false for Primoratz to say that "Where there is no punishment, there are no crimes, no criminal law."[61] After all, we can outlaw, try to prevent, and verbally denounce crimes without punishing them.[62] But it does not go too far to claim that failure to punitively express ourselves, when that is the only expression available, *neglects* our values pretty severely. And therefore it is safe to say that where there is no punitive expression, when that is the only expression available, our legal system fails to fully do justice to our core values.

To the extent that ordinary moral thought has some currency in this debate – and after all, where else could we start? – the burden is less on the expressivist to justify at least a partial place for expression than it is on the consequentialist to justify the *whole* devaluing of expression. We haven't seen that burden met, or even really addressed, in the anti-expressivist literature. The criticisms have simply assumed that non-consequential expression doesn't matter. But that, of course, is the whole question, and the expressivist side has the weight of ordinary moral thought on its side.[63] And the anti-expressivist cannot claim that expressivism has to justify *terribly costly*

[60] This position echoes Hampton's views in both 'The Retributive Idea' *op cit*., at 141–142 and 'An Expressive Theory.' See Gert, Radzik, and Hand *op cit*. for discussion of this element of Hampton's work. My own quasi-Hamptonian position, which does not run afoul of their concerns, is that the crime puts the offender in a socially dominant position over the victim, and that punishment expresses the view both that this dominance does not entail any sort of moral superiority enjoyed by the offender and that the offense and dominance themselves are actually morally repugnant and cancelled.

[61] *Op cit*., at 197.

[62] Adler *op cit*., at 1426.

[63] Lurking in the background of this discussion is, of course, a more basic question of what is of fundamental moral importance—recognizing the intrinsic value of persons or producing good outcomes. Tadros *op cit*., at 105–108, argues that the value of preventing harm surely trumps the value of preventing others from doing wrong, which he takes to be the point of communicative (though not necessarily expressive) theories of punishment, and he defers to Parfit to make this point. But there are powerful arguments in the Kantian tradition that suggest the independent importance of rational expression. So while the expressive theory has work to do here to vindicate itself, so does the opponent when the opponent denies the value of expression. (Tadros' arguments focus on preventing wrongs versus preventing harms. But that isn't our question, exactly. Our question is whether expressing moral commitments by punishing those liable can trump the importance of not harming those liable).

methods of punitive expression, because, again, the question is whether we should have *any* methods of punishment, even moderate fines. If expression of core values has substantial value, then such fines, and probably much more, may be warranted.

This first reply resets the terms of the debate: in response to the concern that the harm done in punishment obviously trumps any benefits of expressing ourselves, the first reply claims that more is at stake than harm and benefit. Harm and benefit are rivaled by the independent value of expressing moral commitments and establishing a collective moral identity. This captures the truest essence of expressivism as a non-consequentialist theory.[64]

The second reply, by contrast, works within the terms set by the criticism. If the harm done in punishment is supposed to be so weighty as to challenge the merits of collective expression, we must also ask how much harm is *averted* in collective expression. Expressivism, as I have framed it, takes none of its justificatory strength from consequentialist reasoning about deterrence and incapacitation. Nevertheless, it is plausible that when we so firmly and publicly express our core values through punishment, we do deter and prevent some crimes and therefore some harms. That deterrence may happen both through direct threat of punishment and through a more subtle process wherein our value system is more readily internalized the more socially dominant it becomes.

These are empirical questions, of course. My speculating as to the answers is not fully satisfying but is dialectically sufficient. For if what the opponent argues is that expressivism fails because it is too harmful, it has to be shown that it *is* too harmful. But in that case, it is not enough to show that punishment harms; it must also be shown that it does not *also* avert an even greater amount of harm. If expressive punishment averts more harm than it creates, then the creation of excess harm is not going to trump the value of expressing our moral commitments, since in that case there *is* nothing to trump that value. It would be like arguing that even if chocolate is pleasant,

[64] Though I present a non-consequentialist expressivism, others (e.g., Narayan *op cit.*; von Hirsch *op cit.*, at 12–14; Brooks *op cit.*, Chapter 6) go for a hybrid theory combining expressive and consequentialist elements. Wringe, 'Collective Agents', *op cit.*, at 444, thinks that for denunciatory communicative punishment to be justified, it would have to be shown to be beneficial. The view embedded in the first reply is that this concedes too much to benefit-focused reasoning (though Wringe himself classifies it as non-consequentialist): expressivists should insist that expression of core moral principles has *value* even if it brings no *benefit*.

it should be avoided insofar as its calories represent an indirect threat to heart health: whether the addition of calories is even a factor to consider all depends on whether the chocolate doesn't also, say, clear arteries, thereby improving heart health.

Again, this isn't to *justify* punishment on consequentialist grounds. It's not even to accept the consequentialist constraint that punishment must be beneficial to be justified. Rather, it's to say that consequentialist reasoning doesn't necessarily factor in as a potential trump over the non-consequentialist reasoning of expressivism in this instance. It's to say that the consequentialist objection might be a non-starter in this case, because expressive punishment might avert more harm than it causes. It might be the most harm-diminishing policy we could institute, for all that we have seen from the anti-expressivists.

(That said, those partial to the consequentialist constraint still might accept the constraint, argue that it is satisfied because of punishment's harm-reduction capacities, and then grant that the value of expression is still integral to the justification of punishment. Again, pure expressivism can be included in a hybrid justificatory system that incorporates both pure expressive and consequentialist considerations).

The third reply hijacks the consequentialist's concerns and incorporates them into the expressivist model of justification. As noted above, expressivist theories must limit what punitive expressions they consider appropriate. And any respectable set of fitting expressions surely must pay heed to the consequences of our actions. If an action is extraordinarily costly, it might well be unfitting. The expressivist in this way can "expressivize" any consequentialist concerns, including the consequentialist constraint under consideration here. On this way of thinking, whenever the consequentialist maintains that certain benefits must accrue for punishment to be permissible, the expressivist can maintain that those benefits must accrue for punishment to fittingly express a fitting attitude, leaving the justifiability of punishment solely a function of fitting expression.

On this version of expressivism, what we will ultimately permissibly express is not just our attitude toward the offender considered in a vacuum, but rather our attitude toward the offender considered in a network of immensely consequential social practices. That attitude has the nuance to incorporate concerns about punishment's costs to

society and to the offender, alongside concerns about the need to express our moral outrage at the offender's behavior.

At the same time, just as expressivists can expressivize any consequentialist concerns, consequentialists can ''consequentialize'' any non-consequentialist concerns, in a way that makes the consequentialist approach much more appealing than is traditionally recognized. It does this by operationalizing what counts as a good or bad state of affairs to be produced in terms of whatever non-consequentialist concerns we might have. For example, it can allow for proportionality constraints on punishment, if such a constraint is built into what counts as a high-ranking state of affairs to be produced by our actions and policies. Accordingly, much of our ability to flesh out a sensible regime of justifiable punishment will hang on how we operationalize ''good state of affairs'' and ''fitting expression.'' But that is how it should be. Our project here is not to articulate an entire framework of punishment, but rather to investigate whether the expressivist approach can be the foundation of such a framework. This move to expressivize consequentialist principles shows that expressivism can supply that foundation, at least as far as the consequentialist is concerned.[65]

As I said, these three answers to expressivism's most difficult obstacle are incomplete sketches. The third is, in my mind, a decisive maneuver, but it will only be available to those who are comfortable incorporating cost/benefit reasoning within the expressive framework. (Relatedly, it assumes certain analyses of what constitutes expressivism and consequentialism – see footnote 65.) Making good on the second requires answering thorny empirical questions, to which we do not yet have adequate answers. And the first hangs upon a broader moral and political theory that the value in

[65] This reply co-opts parallel strategies from the recent movement to ''consequentialize'' non-consequentialist moral theories and ''deontologize'' consequentialist theories. See, for example, Paul Hurley, 'Consequentializing and Deontologizing: Clogging the Consequentialist Vacuum', in Mark Timmons (ed.), *Oxford Studies in Normative Ethics*, vol. 3 (Oxford: Oxford University Press, 2013), 123–153 and Douglas W. Portmore, 'Consequentializing Moral Theories', *Pacific Philosophical Quarterly* 88 (2007), 39–73. There is some debate as to whether this kind of move empties out consequentialism and deontology (and now expressivism) by making them all equivalent, on the premise that deontic equivalence—equivalence in verdicts about what is to be done—is total equivalence. I agree with Portmore and others who think that there is still a real difference between consequentialism and non-consequentialism, even when they are made deontically equivalent, namely, they have different answers to the question, ''What makes Φ right''? On the consequentialist view, what makes punishing justified (or not) is that it brings about (or fails to bring about) the best consequences. On the expressivist view, what makes punishing justified (or not) is that it fittingly expresses (or fails to express) fitting attitudes. This difference remains even if the views are deontically equivalent.

expressing core moral principles is enough to rival the value of consequences. So there are some outstanding questions. Yet, while some take the consequentialist constraint to mean that we should be skeptical of expressivism,[66] I hope to have shown that all three avenues are least promising. In that case, the appropriate reaction to expressivism is not skepticism but tentative curiosity and credence.

IV. CONCLUSION

Given the questions I have left outstanding, the foregoing has not demonstrated conclusively that expressivism is the winner of the punishment debates. However, I hope to have shown it to be more attractive than one might think from a survey of the critical literature. The three dominant objections to it are not compelling, and there are encouraging reasons to think that the biggest question is answerable. There might be other problems coming, of course, but until then the view looks like a real contender.

Philosophy Department, Center for Ethics,
Law, and Society,
Sonoma State University, 1801 E. Cotati Ave.,
Rohnert Park, CA, 94928, USA
E-mail: joshuamglasgow@gmail.com

[66] E.g., Husak, 'Why Punish', *op cit.*, at 457.