

Why Should We Care What the Public Thinks? A Critical Assessment of the Claims of Popular Punishment¹

FREJ KLEM THOMSEN

I. INTRODUCTION

Sometimes the public gets it wrong, even on relatively simple questions of fact. For example, recent polls indicate that a majority of the US adult population continue to believe that a deity created humans in their present form (46%), while a large minority recognize that humans have evolved, but maintain that a deity “guided the process” (32%). Only 15% hold that humans evolved in a process devoid of supernatural interference. (Gallup Politics, 2012)

Presumably, few would dispute that public opinion can be and sometimes is mistaken, in the manner indicated by such examples. After all, the difficulty of forming an accurate opinion under the constraints we all face as epistemic agents is one of the reasons why most countries capable of doing so educate and employ highly specialized experts to research and assess complex issues. And yet, in public and political debates on punishment one is likely sooner or later to run into statements along the lines of: “The leniency with which the courts treat serious criminals offends the public’s sense of justice.” When such statements are offered, it is typically either implied or concluded that the discrepancy between popular sentiments and the sentences meted out by courts can justify revising penal codes to remove or diminish the discrepancy. Proponents of this view, penal populists, affirm what I shall call the claim of popular punishment:

The basic claim of popular punishment: We ought to punish in accord with popular sentiments about punishment.

When it comes to punishment, at least, public opinion is considered by penal populists to give us reason *in and of itself* to adjust policy. What are we to make of this? Is it a sensible view?

Despite its prevalence, the claim of popular punishment is a curiously underexplored topic. And to the extent that it has received attention from academics, the approach has often been the mirror opposite of the proponents’: since penal populism is (obviously!) evil and counter-productive what we need to do is explain why it is that it occurs and find a way to stop its spread of contagion to justice. (E.g. Pettit, 2002; Pratt, 2007; Lacey, 2008; for opposed arguments cf. Roberts, 2011; Dzur, 2012) Perhaps there is no great mystery here, and the normative issues are as clear-cut as assumed by (some) proponents and critics, but I suspect that this is not the case, and that it may be worth therefore to give the normative claim of popular punishment more of a run for its money. My main purpose in this article is to investigate the ways in which such a claim can be supported by an argument.

I proceed as follows: I first explore and attempt to clarify the claim of popular punishment, i.e. the idea that we ought to punish in accord with how

popular sentiments hold that we ought to punish. I then illustrate the general structure that an argument to support the claim must have, and distinguish between three ways of filling out the critical second premise. Finally, I investigate each of these three attempts to complete the argument in turn and demonstrate the difficulty of making the argument at once plausible and potent. I conclude that only one version of the argument can be successful, and that the limited circumstances under which it applies, as well as the demanding background conditions for its plausibility makes it far less interesting than proponents tend to assume. Thus, I shall reveal myself as a moderate critic of the claim rather than a proponent, but I believe that all parties can benefit from the increased clarity about the terms of the debate that it is my hope and principal purpose to provide in this paper.

II. STAKING THE CLAIM

In thinking about the claim of popular punishment we need to separate two issues: the question of whether and if so how to involve the public in the institutions that shape penal policy (e.g. Lovegrove, 1998; Morgan, 2002; Pettit, 2002; Dzur and Mirchandani, 2007; Pratt, 2007; Lacey, 2008; Dzur, 2012) and the question of what a good reason for shaping penal policy one way or the other would be. (e.g. Golash and Lynch, 1995; Bagaric and Edney, 2004; Robinson and Darley, 2007; Ryberg, 2010; Roberts, 2011) We might call them the participatory and the justificatory issues of the relation between popular opinion and punishment respectively. The two issues are at risk of being conflated because they both involve taking public opinion into consideration, but deciding whether and if so how to do so in the first manner – by allowing the public to participate in or influence the decision-making procedure – does not exhaust the scope of the issue.

Of the two issues, I shall be concerned only with the second and I intend to bracket the question of what the best arrangement of the relevant institutions would be; whatever your preferred model, feel free to assume that we are dealing with it. In a slogan, my concern is not who should deliberate, but which reasons should carry weight in the deliberations of those that do.

Having determined the scope of the discussion, it may be worth exploring the components of the claim itself in slightly greater detail. I suggested above that the fundamental idea of penal populism could be boiled down to the basic claim that “we ought to punish in accord with popular sentiments about punishment”. I shall discuss three points in turn regarding sentiments, information, and conflicts.

First, it is worth asking what exactly we are to understand by “popular sentiments”. I mean for the term to capture the foundation of this specific type of argument, which could be referred to also as e.g. “the public’s sense of justice”, “our common-sense feeling of right and wrong”, etc., and I suggest the following working definition:

Popular sentiments: A popular sentiment is a widely shared (within the society at stake) belief with normative content, e.g. “we ought to punish offence O with punishment P”.

Furthermore, I employ the term ‘sentiment’ in order to stress the particular character of the beliefs at stake in this discussion. These are, as I think penal populists will admit, arrived at through largely unconscious judgments, intuitive

and normally strongly emotionally laden. For comparison, we could imagine that public opinion held that crime O ought to be punished with P because the members had all individually applied a particular theory of criminal justice, such as communicative retributivism qua Antony Duff (Duff, 2001) or utilitarian deterrence-theory qua Jeremy Bentham (Bentham, 1996 [[1789]]), carefully and consciously calculated the weight of the moral factors emphasised by the theory given the relevant empirical conditions and concluded that P was the correct answer to what justice required. However, not only is it obviously not the case that the public sentiments which penal populists demand we show consideration are derived in this manner, but it also seems to me that if public sentiments *were* derived in this manner they would not be the type of beliefs that the argument is supposed to capture.² (cf. Robinson and Darley, 2007, p.4-8)

Second, a traditional distinction with regards to popular sentiments on punishment separates sentiments that are and are not informed. By uninformed sentiments we are to understand something like the average layperson's sentiment, based as this typically is on a limited and superficial knowledge of criminology, the law, the penal system and the crimes, criminals and victims. By informed sentiments, on the other hand, we are to understand the sentiment that the same person has once in possession of a reasonable amount of relevant information about these topics.³ Clearly, if interpreted as a binary concept this would be too crude, but however we graduate it, it raises an important problem: given that there will in at least some cases be a difference between a person's lesser and more informed sentiments, which of the two sentiments are we to take as the basis of the claim for popular punishment?⁴

In the following I shall assume that we can give lexical priority to more informed sentiments in intrapersonal comparisons, that is, we can assume that whenever a person's lesser and more informed sentiments diverge we can ignore the less informed and give full credit to the more informed sentiments.⁵ For simplicity's sake, I shall mostly refer to these simply as sentiments.

Third, the requirement that the sentiment be widely shared is slightly inaccurate. In practice it seems to me that any version of the argument will have to understand popular sentiments in a way that meets the following condition: The strength of the claim must correlate with both the strength of the sentiment and its prevalence. The more persons share the sentiment and/or the stronger the sentiment is, the stronger reason we have to punish in accordance with the sentiment. An implication of this condition is that to the extent that intractable disagreements exist in the relevant population, that is, when popular sentiments are genuinely divided, there will be conflicting reasons. The proponent of popular punishment can accept this implication, I believe, without undue concern. After all, there is no reason why the claim would have to be that there is an exclusive reason, and it seems perfectly legitimate to let the claim cash out in practice e.g. as one reason outweighing the others (because it is stronger held or more prevalent), its resulting strength being weakened by contrary sentiments, but so long as it outweighs them still counting as a reason in favour of adjusting penal practices in the appropriate direction.⁶ Given these clarifications, we might restate the claim of the argument to be:

The revised claim of popular punishment: We ought to punish in accord with informed popular sentiments about punishment, the strength of this claim varying in direct proportion to both the strength and the prevalence of the sentiment.

This is a workable definition, but I will suggest that we make two further small qualifications to the revised version of the claim. First, I have already discussed the claim by referring to competing ‘reasons’, and I believe that we can achieve greater clarity by rephrasing the normative claim in terms of a reason rather than the less precise “ought”. But the reason at stake is clearly a *pro tanto* reason, not a decisive (“all things considered”) reason. I believe proponents of popular punishment ought to accept that there can be other reasons for or against a particular punishment than those based on popular sentiments, and even that these other reasons can potentially outweigh the reason supplied by the claim of popular punishment.⁷ Even if some might wish to challenge this, I will stick to this assumption since it extends maximal charity to the proponent: the moderate claim of a *pro tanto* reason is less demanding, while conversely if the case for a *pro tanto* reason fails then the case for a decisive reason cannot succeed.

Secondly, we need to limit the acceptable type of reason. To illustrate why, consider the fact that there may be persons who hold what is intuitively the wrong kind of reason for adjusting penal practices to fit popular punishment. At the risk of being cynical, it seems that this is true of some politicians who are proponents of popular punishment, e.g. because they believe this will aid in their reelection. Clearly, they have a reason for adjusting penal practices, but their self-interested reason for the advancement of a political career is not the kind of reason that will satisfactorily support the claim of popular punishment. There may be different suggestions for how to properly constrain the range of appropriate reasons, but I believe that impartiality will do the trick. By an impartial reason I understand a reason that applies independently of the situation of the individual agent.⁸ This condition rules out the kind of reason available to the self-serving populist politician since that reason only applies given her individual situation as a politician seeking reelection.

We can introduce these two qualifications explicitly to more accurately specify the claim:

The minimal claim of popular punishment: There is at least one impartial *pro tanto* reason to punish in accord with informed popular sentiments about punishment, the strength of this reason varying in direct proportion to both the strength and the prevalence of the sentiment.

This is the claim that I shall take popular punishment to hold, and the minimal conclusion that an argument capable of supporting popular punishment must achieve. Let us turn therefore to how such an argument might look.

III. THE ARGUMENT(S)

The minimal claim of popular punishment is not self-evidently true, so we will want an argument to support it. Let me start by saying a little about how I

think we should understand an argument like this to work. In practice, the basic structure of the argument is frequently something like the following:

- i) Popular sentiments about punishment hold that offence
O ought to receive punishment P.
QED: Hence we ought to punish O with P.

As is apparent, the argument contains (or at least, must contain, to be valid) an implicit premise, connecting the first premise with the conclusion; something along the lines of:

- ii) If popular sentiments hold that O ought to be punished
with P, then we ought to punish O with P.

But although logically sufficient, as banal a premise as this would be unsatisfactory, because it is not apparent why or that we ought to accept it. Proponents of popular punishment must come up with a more plausible version of the second premise.

The argument for popular punishment, then, taking into account the minimal version of the claim I developed in the preceding section and the necessary elaboration mentioned above must look something like the following:

- i) Popular sentiments about punishment hold that offence
O ought to receive punishment P.
- ii) If popular sentiments hold that O ought to be punished
with P, then there is at least one impartial *pro tanto*
reason to punish O with P, the strength of this reason
varying in direct proportion to both the strength and the
prevalence of the sentiment, because (.....).
QED: There is at least one impartial *pro tanto* reason to
punish to punish O with P, the strength of this reason
varying in direct proportion to both the strength and the
prevalence of the sentiment.

In addition, whatever elaboration of the second premise the proponent offers, it must be capable of meeting what seems to me the most obvious challenge to the claim of popular punishment: the argument that popular sentiments should reflect, not constitute reasons for punishment. That is, if there is a right answer to the question of how to punish, then it will be determined by reasons for and against punishing, and if popular opinion is to be interesting or relevant it must be based on and subject to revision by these reasons, not the other way around. In Bagaric and Edney's much cited illustration: "Seeking public views on sentencing is analogous to doctors basing treatment decisions on what the community thinks is appropriate or engineers building cars, not in accordance with the rules of physics, but on the basis of what lay members of the community "reckon" seems about right." (Bagaric and Edney, 2004, p.129; cf. also Golash and Lynch, 1995, p.719) So while popular sentiments can be based on or even express the reasons we have for punishing (although, as Bagaric and Edney's examples suggest, we often suspect that they are not), they are at best superfluous and at worst mistaken. In a slightly more formal way, we may express this as *the dilemma of reasonableness*:

- 1) A belief about whether or not to ϕ is either
reasonable or unreasonable. It is unreasonable if there

is no reason for or against ϕ 'ing; it is reasonable if there is at least one reason for or against ϕ 'ing.⁹

- 2) In forming a belief about whether or not to ϕ , we should not let unreasonable beliefs affect our belief.¹⁰
- 3) In forming a belief about whether or not to ϕ , we should let the reasons for and against ϕ 'ing affect our belief.
- 4) We should not double-count reasons.
- 5) In forming a belief about whether or not to ϕ , we should let the reasons that support a reasonable belief about whether or not to ϕ affect our belief (from 1 and 3).
- 6) In forming a belief about whether or not to ϕ , letting a reasonable belief about whether or not to ϕ affect our beliefs is double-counting reasons, if we have already let the reasons that support this belief affect our belief.
- 7) In forming a belief about whether or not to ϕ , we should not let a reasonable belief about whether or not to ϕ affect our belief (from 4 and 6).

QED: Whether a belief is reasonable or unreasonable, we should not let it affect our belief about whether or not to ϕ (from 2 and 7).

To my mind, the dilemma is fatal to the vast majority of arguments for the claim of popular punishment encountered in public debates. And it serves the purpose of immediately refocusing the debate where many scholars probably feel it belongs: on the traditional issues of whether punishment should be based on just deserts, deterrence, some third factor or some mixture of factors, as well as how to determine the specific policies that would follow from whatever factor(s) we pick. Whether they are or are not based on reasons, it seems that we can skip sentiments entirely and just get to grips with the traditional debate about the problem of punishment. The result is a further condition on the type of argument available to the proponent. The argument must show both that popular sentiments do provide us with reasons, and that these reasons are different from and cannot be reduced to the traditional concerns of criminal justice ethics, such as just deserts, deterrence, reconciliation, etc.

What could such an argument look like? Overall, it seems to me that an argument for popular punishment must fall into one of three categories, the proper elaboration and individual appeal of which will be determined by background assumptions in ethical theory:

1. There can be an epistemic argument that popular sentiments are truth-tracking, and that we have a reason for bringing penal practice into alignment with them since they provide some reason to think that this level corresponds to moral reality.
2. There can be an institutional argument that popular sentiments are reasons-creating, and that we have a reason for bringing penal practice into

alignment with them since they influence (or even determine) what moral reality is.

3. There can be an instrumental argument that popular sentiments create costs and benefits for a system of justice, and that we have a reason of beneficence to bring penal practice into alignment with them since this will minimize the costs of divergence and maximise the benefits of correspondence.

Let me briefly flesh out each of them in turn before we examine them critically in detail in the second half of the paper.

The epistemic argument must involve something like the following version of the second premise:

- iia) informed popular sentiments are fallibly indicative of how we have reason to act; and iib) we have a reason to act in accordance with how we have reason to think we have reason to act.

The second half of the premise is uncontroversial enough, I believe, that I shall set it aside. The first half implies that if popular sentiments hold that we ought to ϕ , then there is a reason to think that we ought *in fact* to ϕ . The premise supports the conclusion of the argument for popular punishment because, if true, it translates directly to the claim of popular punishment once we insert “punish O with P” for ϕ . It tackles the dilemma of reasonableness by challenging the first horn of the dilemma: a reasonable belief is interesting in itself because of its evidentiary value, where it might be taken to constitute evidence directly for the belief or for the soundness or importance of the reasons consistent with it.

Note that to be charitable once more, I am assuming the claim to be only that popular sentiments are *fallibly* indicative, which is to say that they can be wrong but that they are not mistaken so often as to be useless or misleading. We can leave open the question of how strong an indication they provide – the range is everything between only minimally better than chance and up to near certainty – because while this will affect how strong the resulting reason is, the claim we are examining is merely that such a reason exists, so any value in the range will do.

Similarly, the institutional argument will involve something like the following version of the second premise:

- iii) Informed popular sentiments express the values of members of a community; and iib) we have reason to treat members of a community in accordance with the values of members of the community.

Whereas the claim of the epistemic argument is that popular sentiments help us discover what we have reason to do, the claim here is that popular sentiments provide us with reasons for action that we would not otherwise have had. And such reasons emerge because those involved are part of a joint community wherein shared (but subjective) values can justifiably determine how persons ought to treat each other. It tackles the dilemma by insisting that normative beliefs are reasons-generating independently of their being based or not based on reasons, but owes an account, of course, of why that is the case; the critical part of the premise is likely to be the second half, therefore.

Note also that there is an important ambiguity concerning ‘members’. Depending on how we ultimately defend the premise, we may want to hold one of two views. Either that there can be competing sentiments, but that as I have previously suggested the strongest can win out, the force of the resulting reason being diminished proportionally to but not cancelled by disagreement. Or we may wish to hold that only unanimously held sentiments can justify treating members of the community in accord with the values they express, since anything else will be the imposition of the values of some members of the community on others who do not share them. To have any practical importance the argument must hold the former, since unanimously held popular sentiments are unlikely to ever obtain in modern societies, and I shall assume that this is the version of the argument at stake.

And finally, the instrumental argument involves something like the following version of the second premise:

- ii) When actual punishment P_a for O diverges from P this imposes costs C_{P_a-P} ; and iib) we have reasons of beneficence to avoid imposing costs.

Whereas the institutional claim above is that popular sentiments create reasons directly by virtue of the normative force of the values they embody, the instrumental claim is that popular sentiments provide us with reasons we would not otherwise have had only because the sentiments impose costs and benefits on our actions in terms of whatever values we take to apply independently. It dodges the dilemma by pointing out that even if not directly reasons-generating, beliefs have implications for how other reasons cash out.

The key component in this argument is likely to be the first premise, given that the second premise is both widely accepted and strongly opposed by those who dissent, so that we are unlikely to be capable of changing anyone’s mind about it. The central task of the argument, therefore, will be to make credible the idea that over- or underpunishment imposes, or at least is likely to impose, noteworthy costs.

IV. ASSESSING THE EPISTEMIC ARGUMENT

I sketched out above how both the general argument and the specific premise of the epistemic version must go. Putting the two together we get *the epistemic argument for popular punishment*:

- i) Informed popular sentiments about punishment hold that offence O ought to receive punishment P .
- ii) informed popular sentiments are fallibly indicative of how we have reason to act; and iib) we have a reason to act in accord with how we have reason to think we have reason to act.

The minimal claim of popular punishment: There is at least one impartial *pro tanto* reason to punish in accord with informed popular sentiments about punishment, the strength of this reason varying in direct proportion to both the strength and the prevalence of the sentiment.

As I noted above, the second half of the epistemic premise is less interesting because less controversial than the first. It is not, after all, apparent why or that we

should accept that popular sentiments are indicative of moral truth, even fallibly so. Indeed, we might think it quite likely that the public could get it as wrong on normative questions as on factual ones, if not more so, for at least two reasons: First, the methods of establishing moral facts are indisputably both more controversial and less precise than those of establishing non-normative facts of almost any kind. Second, almost everybody recognizes that the general public has been gravely mistaken about important moral issues in the (relatively) recent past – slavery, racism and sexism are widely accepted examples – and there is overwhelming reason to think that it remains mistaken about other crucial issues – the ethics of war, speciesism and poverty aid are prime examples that have strong support in the literature. (McMahan, 2009; Singer, 1995; Regan, 2004; Unger, 1996) Given these concerns, is there any reason to think that public sentiments are truth tracking?

One way of supporting the epistemic premise, which draws on lines of argument advanced in the so-called epistemic justification of democracy (Cohen, 1986; Estlund, 2009), would be to argue that the Condorcet jury theorem applies to sentiments about punishment. The Condorcet jury theorem as classically stated shows that in any group of people in which individuals are fallible but more likely to be right than wrong, the majority opinion in a binary choice situation is more likely to be true than false, the probability of its being true increasing with the size of the group. (Condorcet, 1785) For populations as big as those involved in modern societies the probability swiftly approaches certainty.

An obvious objection to the application of the classical theorem is that where “how we ought to punish?” means e.g. how long we ought to imprison the offender, we are not dealing with a binary choice situation, but with a choice between a range of options. However, there are formal solutions for applying the theorem to non-binary choice situations (List and Goodin, 2001), and it remains true that so long as the average person’s choice is better than chance (i.e. the probability of selecting the right option is greater than $1/n$, where n is the number of options) *and* not less preferred than a false option (i.e. no wrong option has a greater probability of being selected than the right option), the majority have an increasing probability of being right the greater the number of participants in the decision. And this fits our common-sense views in at least some scenarios. It would not be unreasonable to suppose e.g. that if we were to ask people to guess how many beans are in a large jar the average guess would become increasingly accurate the more people we ask.

The important question, however, is whether it is reasonable to suppose that the sentiments of the average person about the type of issue at stake is reliable enough to be at once better than chance and better than all false alternatives. Such sentiments are, I have suggested, the results of a largely intuitive judgement that involves weighing reasons with and/or against each other to arrive at an opinion, such as “offence O should be punished with P”. There are essentially two ways in which public sentiments (if reasonable beliefs of this type) could diverge from the correct conclusion: they could be based on a different set of reasons (fewer, more or other, including supposed but invalid reasons), and they could be based on an incorrect weighing of the reasons.¹¹ We ought, I suspect, to be sceptical of the epistemic quality of public opinion on both counts.

Starting with the weighing, the worry one might raise concerns the susceptibility of the type of judgement at stake to cognitive biases. An intuitive normative judgement of the sort involved in the claim of popular punishment is clearly composite: it must take into account a number of often conflicting reasons. The trouble is that as the number of moral factors involved increases, so do both the complexity, the possibility of mistakes and the influence of the cognitive heuristics necessary to lighten the cognitive load and make an intuitive judgement possible.¹² However, cognitive heuristics, while undoubtedly helpful and perhaps even necessary, unavoidably lead to cognitive biases. As behavioural economics has shown in fascinating detail, there are predictable and serious mistakes that agents will perform when making intuitive judgements, which are best explained by the agents' employment of generally useful but predictably flawed cognitive heuristics. (E.g. Kahneman and Tversky, 1979; Kahneman, Knetsch and Thaler, 2009; Kahneman and Tversky, 2009a; 2009b)

Let me illustrate with just one bias that we can expect to be prominent in the type of judgements we are discussing: anchoring effects. An anchoring effect occurs when a person adjusts subsequent estimates in the light of an initial factor, which either has no relevant relation to the issue or should not be given the influence it is granted.¹³ In a classic decision theoretical experiment Amos Tversky and Daniel Kahneman made two experimental groups observe the results of a spinning roulette-wheel, stopping in one case at the number 10 and in the other at the number 65. Immediately afterwards participants were asked to guess the percentage of African nations which had obtained membership of the United Nations, indicating first whether the percentage number was higher or lower than the number the roulette wheel had landed on, and second what the percentage was. The group who had witnessed the wheel stopping at the number 10 estimated on average 25%, while the group that had witnessed the wheel stopping at 65 estimated on average 45%, a difference that appears both irrational and incredible given that it was apparent to both groups that the results of the roulette wheel had no relation to the question being asked.¹⁴ (Kahneman and Tversky, 1974)

Obviously, whether or not anchoring effects occur in the context of intuitions about punishment is an empirical question, but it seems likely that they might. A prime candidate would be the perceived current punishment. If in effect, it would exert an intuitive pull towards what the person believes the current punishment to be. But of course there can be no guarantee that an anchoring effect pulls intuitions in the right direction, nor any reason to believe that the perceived level of current punishment would do so. The result would be that we would get convergence around the wrong estimate. Whether intuitions in such a case happen to have evidentiary value will be an effect of how the convergent level of intuitions, the actual level of punishment and the appropriate level of punishment happen to be related: if the convergent level is located between the actual and appropriate levels then intuitions will temporarily retain evidentiary value; although inaccurate, they will at least point us in what is initially the right direction (e.g. less severe or more severe punishment). If not, then they will be entirely misleading.¹⁵

The second and more fundamental problem for the epistemic argument is that there is good evidence that the intuitive judgements at stake fail to take into

account the relevant reasons *and only* the relevant reasons. (Singer, 2005; Greene, Sommerville et al., 2001) There is a blooming and extensive literature documenting the apparent influence of irrelevant factors on moral intuitions, but let me describe just one. In 2008 Simone Schnall, Jonathan Haidt, Gerald L. Clore and Alexander H. Jordan published the results of a series of experiments concerning the relation between the emotion of disgust and moral judgement. (Schnall, Haidt et al., 2008) Briefly, one of their experiments involved asking participants to intuitively assess the degree of moral wrongness, if any, in the action of the protagonist in each of three different detailed scenarios. All participants were seated in an empty office environment during off-hours, but the test group had their environment manipulated in that: “[the] workspace was set up to look rather disgusting: An old chair with a torn and dirty cushion was placed in front of a desk that had various stains, and was sticky. On the desk there was a transparent plastic cup with the dried up remnants of a smoothie, and a pen that was chewed up. Next to the desk was a trash can overflowing with garbage including greasy pizza boxes and dirty-looking tissues.” (Schnall, Haidt et al., 2008, p.7) The results were unequivocal: test group participants judged the behaviour of the protagonists in the scenarios much more harshly than the control group, i.e. were more likely to judge the action wrong and judged it as more wrong on average, although they were unaware of the influence of provoked disgust on their judgements.¹⁶ Admittedly there were variations within the test group correlating with self-reported susceptibility to disgust, but the results must be considered disconcerting by those who mean to give credit to intuitive judgements of moral scenarios nonetheless. If intuitions are to be considered evidence in favour of normative beliefs, then they should not vary with such factors as the presence of greasy pizza boxes.

One final challenge is worth noting. In the preceding I have discussed the epistemic qualities of public opinion in isolation. But to support the second premise, popular sentiments must meet a more demanding standard than this, for they must be not merely fallibly indicative in the sense that they do better than chance, but less fallible, i.e. better than the alternative procedure which they are meant to correct. Recall that popular sentiments are meant to give us a reason to think that we have reason to ϕ . But while this *may* be true in situations with no contrary epistemic background – although the reservations I have presented above suggest otherwise – clearly the claim of popular punishment is meant to apply even in the situation where, as today, there exist well-developed theories of criminal justice as competitors. And here it is no longer obvious that even popular sentiments that met the standard of better-than-chance could give us reasons to think that we have reason to ϕ .

Consider that given two different measuring instruments, one of which is both more accurate and precise than the other, adjusting our beliefs in the light of the results of the less accurate instrument is no longer desirable. Return for illustration to the bean-guessing example I mentioned earlier, and suppose that we now allow a second group of people to measure the jar, look up the average size of a bean and calculate an estimate on that basis. In effect this is to allow them to make a cumbersome, conscious, calculated assessment of essentially the same factors as the members of the first group base their guess on. If we accept reasonable assumptions about the advantages of this procedure compared to

guessing by sight alone, we would both expect the average of the estimates of this second group to be more accurate than the average of the first group, and we would expect including the guesses of members of the first group in an average of the estimates of the second group to *decrease* the accuracy. Here the power of the Condorcet jury theorem works in reverse: because the estimates of the first group are *less* accurate, as the number of misleading opinions included grows, so too does the probability of reaching the right answer rapidly deteriorate to zero. And the situation is analogous, because for all its defects and controversy, the characteristics of being a cumbersome, conscious, calculated assessment of essentially the same factors as popular sentiments are (ideally) based on seems to me an accurate (if incomplete) depiction of the work of criminal justice thinkers. The possibility remains that popular sentiments are better at tracking moral truth than criminal justice ethics, either because it better weighs the reasons or takes into account reasons that philosophers have missed. But barring a successful argument to that effect, and in light of the concerns I have sketched above, I see no grounds for assuming this to be the case.

In conclusion, although there are perhaps situations in which the average person's intuitive judgement is more likely to be better than chance, the assessment of concrete moral scenarios is not necessarily among them. As has been pointed out, the problems raised by awkward influences on intuitions may be difficult to contain. (Tersman, 2009; Sandberg and Juth, 2010) It may be hard to cordon off a set of intuitions that remain trustworthy, in which case the use of any intuitions is cast into doubt, and moral philosophy faces a severe methodological challenge. This remains a hotly contested subject in contemporary moral philosophy.¹⁷ But even if true this can hardly save the epistemic argument. It simply means that providing *any reason* for or against a particular type of punishment is going to be much more difficult than we have supposed, if not plain impossible.

V. ASSESSING THE INSTITUTIONAL ARGUMENT

Having assessed the possibilities of completing an epistemic argument for the claim, let us turn to the institutional version. Recall that the general premise I suggested was the following:

- ii) Informed popular sentiments express the values of members of a polity; and iib) we have reason to treat members of a polity in accordance with the values of members of the polity.

And as I pointed out previously, the critical part is the account of why we have reason to treat members in accord with their values.

One version of the argument might be based on the duties of democratic representatives, and allow for the intersection of the participatory and justificatory issues that I initially separated. Suppose that there are reasons to have a representative democratic institution which makes the decisions regarding punishment (e.g. a parliament which legislates for a criminal code containing sentencing guidelines), and furthermore that the representatives have reason (perhaps in the strong sense of having a moral duty) to represent the demos, i.e. to take decisions in accord with the values of the public that they represent. This, it seems, could provide the required reason.

There might be good grounds for objecting to the assumptions of the argument, such as whether representatives genuinely have a reason to represent the current values of the demos in the direct way necessary, but let us grant the argument these assumptions. Even if we do so, this version of the institutional argument fails to provide the impartial reason required by the claim. To see this consider that it cannot provide a reason for those *represented*, only for those representing them. The situation is parallel to the situation where an agent has made a promise to punish in accord with popular sentiments. Now, moral philosophers disagree on whether and if so why and how making promises creates reasons for action, but let us suppose for a second that it does. Then, surely, the promiser now has a *pro tanto* reason to punish in accord with popular sentiments. “Certainly”, she might say, “I know that popular sentiments are misguided and unjust. Impartially, they give us no reason to do anything. But it just so happens that I have promised to act in accordance with them. Hence, I have a reason to do as they suggest.” Nobody, I suspect, would take this to be the kind of reason at stake in the claim of popular punishment, nor are we likely soon to encounter proponents of the claim lamenting the fact that they happen to be in a situation where they have reason to act in accord with popular sentiments that are otherwise unsupported.

A different variation might start with the widely accepted premise that there is reason to let the sentence vary with the severity of the crime, argue that the severity of the crime varies with cultural norms, that popular sentiments are, or at least can be, an expression of such cultural norms, and that therefore the sentence should vary with popular sentiments. Roberts sketches the contours of such an argument when he suggests that “the seriousness of any particular act is determined, to some degree, by the extent to which it offends community mores.”¹⁸ (cf. Roberts, 2011, p.115-116)

Now, there are two ways in which community views might affect the seriousness of offences. We can imagine that they directly determine, or at least affect, the seriousness, i.e. “serious is as the public says”. Alternately, we can imagine that sentiments indirectly affect the seriousness, e.g. because the factors that determine seriousness, such as the amount of harm caused, are influenced by community views. The trouble is that neither is plausibly capable of completing the institutional argument. The first reading of the argument implausibly holds that mere belief can affect seriousness independent of other factors (such as harm). This runs straight into the dilemma of reasonableness, and while it may be possible to extend the argument so as to avoid this I confess that I cannot see how. (cf. also Ryberg, 2010, p.161-165; Golash and Lynch, 1995, p.714) Meanwhile, the latter is really a variant of the instrumental argument. Thus, while it is undoubtedly true that sentiments can have this effect, e.g. because harm is at least partially a psychological phenomenon and as such is subject to the influence of cultural norms (and, we might add, individual psychological idiosyncracies), popular sentiments only enter the picture as part of the background conditions which together determine how much harm the action causes, and as a result thereof how serious it is. The severity remains a function of the harm, not the sentiment. (For a related point cf. Golash and Lynch, 1995, p.711-712)

Finally, an argument might be made on the basis of the voluntary relinquishing of individual rights. Thus, if individuals possess moral rights, e.g. a right to liberty based on autonomy, and subjecting them to punishment will constitute *prima facie* violations of these rights, it seems that a system of justice must require individuals to conditionally relinquish those rights. And it is arguable that at least citizens in a democratic society do so, but on the condition that they are capable of collectively deciding the system of justice which they are potentially subject to punishment by. Hence, the argument would go, if the system is not one that is endorsed (or, endorsable) by the persons subject to it, e.g. because actual punishment is disanalogous to the punishment that would be meted out by a system consented to, then the conditional relinquishment of rights has not been triggered and punishment is unjustified.¹⁹ An argument along these lines will mesh well with libertarian intuitions, which some may consider an advantage. As Dzur puts it, there is a “concern” “prominent in Anglo-American political thought” that “...citizens be held accountable only to rules they authorize and understand,” to which we might add that they can for the same reasons demand to be held accountable only in the way and to the extent that they authorize and understand.²⁰ (Dzur, 2012, p.117)

There are potential problems with this line of argument familiar within contractualist theory, e.g. about how to account for majority rule and the subjection of dissenting minorities to a general system of justice, as well as the necessary forms of consent, which will often in practice be implicit. The solutions to these problems are not uncontroversial, but I will suppose that contractualists can provide satisfactory answers to them. Even granting these assumptions, however, it seems that to work in the context of the claim of popular punishment the argument requires conflating the hypothetical and actual situations. Thus, traditional contractualism will impose a hypothetical contract situation on agents in which it might well require that their sentiments with respect to criminal justice count as reasons while constructing the social contract. But there is no reason to suppose that the actual (irrational, non-ideal) sentiments of a population at any given time have any relation to the sentiments individuals would have in such a hypothetical situation. In fact, the choice-situation for the contract is typically constructed exactly so as to avoid certain biases that afflict the actual choice-situation, Rawls’ imposition of the “veil of ignorance” being the classical example. (Rawls, 1999 [[1971]]) Furthermore, this version of the argument too seems better suited to the participatory than to the justificatory issue because, again, it seems it will be hard pressed to provide an impartial reason. It could, perhaps, be held to constitute an argument in favour of involving the public so as to provide them with the opportunity of con- or dissenting based on their actual, current preferences, rather than idealised or hypothetical preferences, but it could not, it seems, provide them with *a reason for or against consenting* if they were involved in such a manner. As such, it cannot support the claim of popular punishment.

VI. ASSESSING THE INSTRUMENTAL ARGUMENT

The last and perhaps most promising of the three arguments is the argument that the sentiments of the community can affect the way that other reasons cash out, and that in so far as we hold that these reasons influence the way we ought to

punish, the sentiments must therefore be taken into account. I formulated this as the following attempt at filling out the second premise:

- ii) When actual punishment P_a for O diverges from P
this imposes costs C_{P_a-P} ; and iib) we have reasons of
beneficence to avoid imposing costs.

The first step for the argument is to make credible the notion that divergence will impose costs. Costs of divergence could arise in the form of non-compliance, loss of legitimacy and cooperation, increased vigilantism and retaliation, loss of communicative meaning, including centrally censorious meaning, and sheer frustration, indignation and anxiety as a result of the perceived inadequacy of justice. (Golash and Lynch, 1995, p.708-710; Gardner, 2007; Robinson and Darley, 2007, p.18-31; Roberts, 2011, p.111; Ryberg, 2010, p.152-159) It may even be possible to cover a version of the argument for popular punishment according to which the cost imposed is disrespect towards the victim of the offence, because the punishment does not meet what popular sentiments hold to be the appropriate punishment. So it seems there are a range of potential costs, at least some of which should be acceptable to most moral theories. Does this mean that we should accept the instrumental argument?

At least some moral philosophers will be little impressed with the instrumental argument for the sole reason that they take it that costs and benefits cannot justify setting the level of punishment at any particular level. The severity of punishment, such absolutists could argue, must be determined solely by considerations of justice, and the relevant considerations of justice with regard to punishment is just deserts, i.e. what the offender *deserves* to suffer by reason of her offence. For those who hold such a *Fiat justitia et pereat mundus*-position, the instrumental argument is simply unavailable. But given the relative scarcity of absolutists in moral philosophy, it may be worth considering what those who are willing to grant at least some weight to considerations of beneficence in determining the justified level of punishment should think about it.²¹ In addition, it is worth noting that for some costs the instrumental argument can cut across the traditional divide between retributivist and consequentialist justifications of punishment. If, for example, the function of punishment is taken to be the fulfilment of our collective duty to express censure at the wrongdoing perpetrated by the offender, then a $P-P_a$ discrepancy that causes the punishment to lose censorious meaning imposes a cost, but one which is counted in the deontological currency of the loss of our ability to fulfil the duty of expressing censure at wrongdoing. (cf. Roberts, 2011, p.112-114)

Several further objections remain. First, it is worth noting that on the most obvious version of the argument it assumes that the costs and benefits emerge from the distance between actual popular sentiments and the *perceived* level of punishment, rather than the distance between sentiments and the *actual* level of punishment. Given that studies show that the difference between perceived and actual levels of punishment can be substantial, this is an important distinction.

It also opens the door to an objection based on the relative costs of changing actual penal practices and changing perceptions of penal practices. (cf. Ryberg, 2010, p.157-158) Presumably, if the reasons at stake are reasons of beneficence, they cannot support changing penal practice if there is a better (more beneficial,

e.g. because less costly) alternative available in altering perceptions of penal practices. In the most uncontroversial cases there will be no discrepancy between actual and desired punishment, only between actual and perceived punishment. Suppose e.g. that penal practices inflict punishment of five months prison for offence O, that public sentiments demand at least five months prison for O, but that public perceptions of penal practices are that O is typically punished with three months prison, and that this divergence has the costs that the argument assumes. We can then either change penal practices, i.e. increasing punishment until perceived punishment becomes five months, or we can change perceptions until perceived punishment becomes five months. If, plausibly, it will be easier and less costly to do the second, then clearly this is what we must be taken to have a reason of beneficence to do. Studies suggest both that the problem is real and that the solution of adjusting perceptions is feasible. (Mirrless-Black, 2002; Indermaur and Hough, 2002; Hough and Park, 2002; Bagaric and Edney, 2004, p.129-130)

However, since I have assumed that we are dealing with informed sentiments, this situation may be unlikely to arise. Presumably, an accurate perception of actual levels of punishment will be part of any realistically attainable maximum level of information. This leaves us with a further choice between changing actual punishments or changing popular sentiments. These may be harder to alter than the level of information, but there is no reason to assume that changing public attitudes in such a way is impossible. (Robinson and Darley, 2007, p.52-66) Suppose however, as we should, that, while possible, doing so is itself costly. And suppose that we limit the discussion to situations where the costs of changing sentiments outweigh the costs of changing practice. Will there not then be reason to align punishment with sentiments?

At this point, I am willing to concede. Once we have narrowed the field in this way, I do not think we can raise further objections. It remains true that, on the occasions when the empirical conditions are met, those who hold that costs and benefits play a part in setting the justified level of punishment are forced on pain of contradiction to accept that there is at least a(n impartial, *pro tanto*) reason for punishing in accord with popular sentiments. This is hardly surprising, and may leave all parties to the traditional debates feeling comfortable. Absolutist deontologists will perhaps take it as further proof that consequentialists and threshold-deontologists are willing to take all the wrong factors into account all the while studiously ignoring the moral trumps that *really* matter, while their opponents should be able to happily agree that in cases where it will have beneficial consequences to adjust punishment so as to better align with popular sentiments, clearly, this gives us reason to do so.

The impact of popular sentiments should not be overestimated, though. The reason is not decisive. The optimal punishment regime will depend on both normative and empirical facts: which costs go into the equation, and how they increase or diminish in a concrete social setting according to changes in punishment. It might be true for example, as Robinson and Darley argue, that the greatest benefits produced by the criminal justice system depend on alignment between (the ordinal ranking of) the punishments it metes out and popular sentiments, which amounts to the claim that the cost of not meeting popular

sentiments outweighs any individual competing cost. But it would remain an open question whether marginal deviations from perfect correlation between sentiments and punishments would be an all-things-considered improvement. (Robinson and Darley, 2007, p.21, cf. also p.45-48; Golash and Lynch, 1995, p.709-711)

VII. CONCLUSION

In the preceding I have attempted to clarify what we should understand the claim of popular punishment to (minimally) be, how an argument for the minimal claim of popular punishment must proceed, and how the difficult second premise of such an argument might be filled out. I examined three versions of such an argument. There may be others, but I suspect that most or all of them will turn out to be variations on the ones I have looked at, and will run into variations of the challenges I have described. I have concluded that these challenges are strong enough that the first two versions of the argument, the epistemic and institutional, are unpersuasive, while there are circumstances under which the third argument will give those who accept certain normative presuppositions the reason required by the claim of popular punishment.

Is this a triumph for the proponent of popular punishment then? I have argued that we cannot refute the claim, since there is at least one version where under certain circumstances at least some ethical principles will support it. And if this is enough to satisfy the proponent, then so be it. But it is a modest conclusion, and one that will be hard pressed to justify claims of popular punishment as they are typically made in practice, because it faces a heavy burden of proof in demonstrating that *this case*, in fact, happens to be one of those where it applies. There is a reason, then, why we should care what the public thinks. But for many proponents I suspect that this will be one of those cases where you get what you want, only find that it was not what you wanted after all.

¹ I am grateful to Kasper Lippert-Rasmussen, Martin Marchman Andersen, Thomas Søbirk Petersen, Sune Læggaard, Thom Brooks, Chris Bennett, Jesper Ryberg, Jakob Holtermann, Julian Roberts, Richard Lippke and Paul H. Robinson for helpful comments.

² An interesting question is whether it would make a difference to the argument for penal populism if it was based on the public holding such theoretically-based and reflective beliefs. I believe that the criticism I level at the claims of popular punishment based on public sentiments could be extended relatively easily to cover such cases, but since it seems to concern a claim that nobody in fact makes, nor is likely to make given the negligible probability of the general public attaining the requisite background knowledge of moral philosophy, I set the discussion and its attending complications aside.

³ Note that on the definition of a popular sentiment I have proposed, to retain the quality of an intuitively formed belief while being increasingly informed, the information must be about empirical rather than normative matters. If in addition to informing someone of criminology, the law, the penal system and the crimes, criminals and victims we informed them of the arguments, principles and positions current in criminal justice ethics, we would presumably eventually end up with a belief that was the result of a reflective process and more or less identical to one of the existing positions. I shall assume that this is not what we understand by being informed.

⁴ Studies show a clear tendency for opinions to become more lenient with respect to sentencing the better informed they are. Roberts labels this tendency “one of the most robust and often-replicated findings in the field”. (Roberts, 2011, p.105) Note also that Roberts advances a different argument for roughly the same position that I here adopt. (Roberts, 2011, p.106)

⁵ Once again, I think this is being maximally charitable to the proponent of the claim of popular punishment. In practice, this assumption is so difficult to work with, since counterfactual sentiments are well-nigh impossible to predict and the achievement of widespread maximal informedness is unfeasible, that most actual claims of popular punishment will have to find some way of working around it, i.e. by basing the claim on less informed sentiments. But while this represents a genuine challenge to putting the claim into practice, I want to focus the current discussion on the strength of its foundation.

⁶ Robinson and Darley suggest that the diversity of intuitions about punishment may be less dramatic than is sometimes assumed. Empirical evidence, they argue, supports the claim that there is relative convergence on comparative judgements of offence seriousness, i.e. “offence O_1 is less/more serious than O_2 ”. (Robinson and Darley, 2007, p.9-11; cf. also Robinson, Kurzban and Jones, 2007) Intractable differences of intuition will then rather concern the anchoring points of the penal scheme.

⁷ This seems to be the point Roberts has in mind when he rejects the “direct importation” model of popular punishment, because “one can easily imagine public sentiment turning against a particular category of offender or public hostility being aroused against a particular offense. If public views were followed, the result would be severity premiums for a particular kind of offender or some specific offense—which could not be justified by any retributive principles.” (Roberts, 2011, p.103) A different reading of Roberts’ argument would not require a balance of reasons, but set retributivist constraints on the influence of popular opinion. This reading is supported when Roberts emphasises that “a model that directly imports public views throughout the criminal justice system would lead to undesirable outcomes.” (Roberts, 2011, p.103) The desirability of such outcomes would, on the balance of reasons view, be a function of the relative strength of all the reasons at stake, and it would therefore be an open question whether direct implementation or maintaining status quo would better fit the outcome deemed desirable on a balance of reasons in each case. Only on a constraint view could we say beforehand with any degree of certainty that the results would be undesirable. Overall, I believe the balance of reasons position is the more plausible view.

⁸ A requirement along these lines fits with general principles of metaethics accepted by many moral thinkers, such as Gerald Cohen’s interpersonal test or Richard Hare’s universalisability condition, (Hare, 1972; Cohen, 1991) but I need not make the controversial claim that impartiality is a general condition for a reason’s being valid in moral or political reasoning. I shall stick to the simpler and less controversial claim that it suitably constrains the reasons at stake in the claim of popular punishment.

⁹ Some might argue that under this definition there can be no unreasonable beliefs about whether or not to perform an action, since there will always be at least one *pro tanto* reason for or against performing any conceivable action, or perhaps more modestly that this will as a minimum apply to any realistically relevant action. I do not have any firm belief about whether or not this is the case, but it seems to me that I could concede the point while

maintaining the thrust of the dilemma with respect to the argument for the claim of popular punishment.

¹⁰ Some may prefer to speak of “making a decision of whether or not to ϕ ” and letting beliefs affect our “decision” e.g. because the phrasing I have employed may seem to suggest moral internalism, where normative beliefs are themselves motivating or even determinate of an agent’s actions, which many thinkers would prefer not to assume. I believe that the dilemma holds on either version, and so do not mean to imply a particular stand with respect to the internalism/externalism-debate. Furthermore, and in a similar vein, some may wish to read “a belief about whether or not to ϕ ” as “a belief about whether or not we ought to ϕ ”, or, even more elaborately, “a belief about whether all things considered we have reason to ϕ ”. I take the three to be identical in the context of the dilemma, and so have used the simplest formulation.

¹¹ Some might object that I have defined such sentiments as intuitive, and that the process of weighing reasons is reflective. This is true on a more narrow understanding of intuition sometimes applied in moral philosophy, but not on the looser meaning I employ here. It seems to me that much – but not all – of our practical reasoning consists of weighing reasons in a way which is spontaneous, fast and unreflective, so that we are not conscious of the details of the process, and would struggle to put it into words.

¹² The alternative is to shift to a reflective process of deductive reasoning. But once we do that we are no longer dealing with the type of belief at stake in the claim of popular punishment.

¹³ Other notable candidates well described in the literature include e.g. availability and representativeness, the first of which concerns the tendency to overestimate the relative weight of prominent or emotionally charged factors, and the other the tendency to overestimate the relative weight of common or average factors. We might suppose e.g. that the first influences perceptions of offences or punishments based on the degree to which they can be visualised and become emotionally significant when so imagined, and that the second influences perceptions based on the degree to which the particular state of a moral factor is considered “normal”. Another prime candidate drawn from the philosophical literature is pure time discounting, which is highly intuitive to most persons, but extremely hard to justify. (Parfit, 1984; Cowen and Parfit, 1992; Broome, 1994; Cowen, 2001)

¹⁴ A somewhat similar objection concerns the difficulty of establishing empirically what the intuition responds to. (Golash and Lynch, 1995, p.720-725) The objection differs, however, in that it does not fundamentally challenge whether intuitions track the relevant features of the situation, but only our ability to identify which features of the situation are being picked out by any particular intuition, and hence our ability to formulate general principles from composite judgements of individual cases.

¹⁵ An objection at this point might be that I have already, for charitable reasons, assumed that a number of potential sources of error were not present in the sentiments at stake. I have assumed, recall, that these sentiments are informed, consistent, and weighable. Could we not simply adopt similar charitable assumptions at this point? I think not. While there is no sharp distinction between the sources of error at stake, it seems to me that we could narrow the field and educate the public in order to derive a set of actual sentiments that met the first three assumptions, while no such possibility exist for the problems sketched in this section. While it is perhaps theoretically possible to avoid cognitive biases and distorting influences (see below), it seems to me unrealistic to expect that we could achieve it with

public sentiments in practice. A different way of putting the point is that we *could* adopt such “charitable” assumptions for these problems too, but doing so would make an argument based on them vacuous since it effectively empties the field of candidate sentiments, and thus would hardly be a service to the proponent.

¹⁶ Some might object that this seems an example of disgust affecting the weighing of reasons rather than introducing a reason. There is some difficulty, I think, in distinguishing the two in cases like this. A related experiment by Haidt is perhaps more suggestive, since it introduces the factor of disgust directly into the scenarios persons are asked to assess, and shows equally strong intuitive responses, indeed responses so strong as to be maintained when test subjects are pushed to admit that they cannot explain why their disgust-response should constitute a reason. (Haidt, 2001)

¹⁷ Contemporary defenders of intuitionism include Robert Audi (Audi, 2005), Roger Crisp (Crisp, 2006), Michael Huemer (Huemer, 2007) and Derek Parfit (Parfit, 2011). Generalizing, one common line of defence is that there are certain fundamental intuitions which we can justify as reliable in the face of otherwise confusing factors, such as the intuition that welfare is good while harm to welfare is bad, or the intuition that we have reason to promote welfare. It seems safe to say at least that the position has not been abandoned yet.

¹⁸ Slightly earlier and in the same vein Roberts argues that: “Seriousness ratings and sentencing factors (for example) should reflect public opinion not for instrumental reasons—because the public will have more confidence in or respect for sentencing—but because community views constitute an inherent element of crime seriousness.” (Roberts, 2011, p.114) However, Roberts’ point may perhaps be better understood as concerned with the participatory issue, and advocating the use of popular sentiments through some form of public input – Roberts favours sociological research over democratic participation – to properly calibrate punishments so as to achieve the ends determined independently. This seems to me a more persuasive point. In any case it is not clear to me and I do not mean to suggest that he advocates the contractual argument in the shape I here present it.

¹⁹ Note that the claim need not be, and plausibly should not be, that punishment is thereby all things considered unjustified. It can be unjustified in the limited sense of there being a *pro tanto* reason against it, which is all that the minimal claim of popular punishment requires.

²⁰ Note that Dzur takes this to be an argument against institutional insulation. Whether it is so depends on whether one believes that it is explicit or potential consent that does the normative work. If the former, then institutional involvement is required to meet the condition; if the latter, then it suffices that the solution arrived at is one that citizens *would* consent to, i.e. that it aligns with their sentiments. Also note that the argument here is related to but different from the form of retributivism which takes punishment to be the rectification of a balance of benefits that has been disturbed by the unfair advantage the criminal has gained through enjoying the protection of the law while refusing the restrictions of liberty that granting that protection to other members of society would constitute. (Murphy, 1973) For criticism of using this as the basis of the claim of popular punishment cf. (Golash and Lynch, 1995, p.715-716)

²¹ I believe that the majority of retributivist thinkers will want to be non-absolutists and must therefore face the argument, especially of course those that accept Hart-style mixed theories, such that the institution of punishment can be fundamentally justified by its

beneficial consequences. But not much depends upon whether this is true: absolutists can reject the argument at the cost of certain other difficulties, while non-absolutists can avoid the difficulties absolutism raises at the cost of having to face the argument. Whether one or the other is overall the more defensible position is probably a matter on which thinkers can reasonably disagree.

REFERENCES

- Audi, R. (2005). *The Good in the Right: A Theory of Intuition and Intrinsic Value*, Princeton University Press.
- Bagaric, M. and R. Edney (2004). "The Sentencing Advisory Commission and the Hope of Smarter Sentencing." *Current Issues in Criminal Justice* **16**(2): 125-139.
- Bentham, J. (1996 [1789]). *An Introduction to the Principles of Morals and Legislation*. Oxford, Clarendon Press.
- Broome, J. (1994). "Discounting the future." *Philosophy & Public Affairs* **23**(2): 128-156.
- Cohen, G. A. (1991). "Incentives, Inequality and Community". Tanner Lecture.
<http://www.tannerlectures.utah.edu/lectures/documents/cohen92.pdf>
- Cohen, J. (1986). "An Epistemic Conception of Democracy." *Ethics* **97**(1): 26-38.
- Condorcet, J.-A.-N. C. d. (1785). "Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix". Paris, l'Imprimerie Royale.
<http://gallica.bnf.fr/ark:/12148/bpt6k417181/f4.image>
- Cowen, T. (2001). "What is the Correct Intergenerational Discount Rate?". Fairfax, George Mason University, Department of Economics.
<http://www.gmu.edu/centers/publicchoice/faculty%20pages/Tyler/DISCOUNT.pdf>
- Cowen, T. and D. Parfit (1992). "Against the Social Discount Rate". *Justice Across the Generations: Philosophy, Politics, and Society*. P. Laslett and J. Fishkin, Eds. New Haven, Yale University Press: 144-161.
- Crisp, R. (2006). *Reasons & the Good*. Oxford, Oxford University Press.
- Duff, A. (2001). *Punishment, Communication, and Community*. Oxford, Oxford University Press.
- Dzur, A. W. (2012). "Participatory Democracy and Criminal Justice." *Criminal Law and Philosophy* **6**: 115-129.
- Dzur, A. W. and R. Mirchandani (2007). "Punishment and democracy: The role of public deliberation." *Punishment & Society* **9**: 151-175.
- Estlund, D. (2009). *Democratic Authority*. Princeton, Princeton University Press.
- Gallup Politics (2012) "In US., 46% Hold Creationist View of Human Origins." <http://www.gallup.com/poll/155003/hold-creationist-view-human-origins.aspx>
- Gardner, J. (2007). "Crime: In Proportion and in Perspective". *Offences and Defences: Selected Essays in the Philosophy of Criminal Law*. Oxford, Oxford University Press: 213-.
- Golash, D. and J. P. Lynch (1995). "Public Opinion, Crime Seriousness, and Sentencing Policy." *American Journal of Criminal Law* **22**: 703-732.
- Greene, J., R. B. Sommerville, et al. (2001). "An fMRI Investigation of Emotional Engagement in Moral Judgment." *Science* **293**: 2105-2108.

- Haidt, J. (2001). "The Emotional dog and its rational tail: A social intuitionist approach to moral judgment." *Psychological Review* **108**: 814-834.
- Hare, R. M. (1972). *The Language of Morals*. Oxford, Oxford Paperbacks.
- Hough, M. and A. Park (2002). "How malleable are attitudes to crime and punishment? Findings from a British deliberative poll". *Changing Attitudes to Punishment - Public opinion, crime and justice*. J. V. Roberts and M. Hough, Eds. Abingdon, Routledge: 163-183.
- Huemer, M. (2007). *Ethical Intuitionism*, Palgrave Macmillan.
- Indermaur, D. and M. Hough (2002). "Strategies for changing public attitudes to punishment". *Changing Attitudes to Punishment - Public opinion, crime and justice*. J. V. Roberts and M. Hough, Eds. Abingdon, Routledge: 198-214.
- Kahneman, D., J. L. Knetsch, et al. (2009). "Anomalies: The Endowment Effect, Loss Aversion, and Status-Quo Bias". *Choices, Values, and Frames*. D. Kahneman and A. Tversky, Eds. New York, Cambridge University Press: 159-170.
- Kahneman, D. and A. Tversky (1974). "Judgment under Uncertainty: Heuristics and Biases." *Science* **185**: 1124-1131.
- Kahneman, D. and A. Tversky (1979). "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica* **47**(2): 263-291.
- Kahneman, D. and A. Tversky (2009a). "Choices, Values, and Frames". *Choices, Values, and Frames*. D. Kahneman and A. Tversky, Eds. New York, Cambridge University Press: 1-16.
- Kahneman, D. and A. Tversky (2009b). "Rational Choice and the Framing of Decisions". *Choices, Values, and Frames*. D. Kahneman and A. Tversky, Eds. New York, Cambridge University Press: 209-223.
- Lacey, N. (2008). *The Prisoner's Dilemma: Political Economy and Punishment in Contemporary Democracies*. Cambridge, Cambridge University Press.
- List, C. and R. E. Goodin (2001). "Epistemic democracy: generalizing the Condorcet jury theorem." *Journal of Political Philosophy* **9**(3): 277-306.
- Lovegrove, A. (1998). "Judicial Sentencing Policy, Criminological Expertise and Public Opinion." *Australian & New Zealand Journal of Criminology* **31**: 287-313.
- McMahan, J. (2009). *Killing in War*. Oxford, Oxford University Press.
- Mirrless-Black, C. (2002). "Improving public knowledge about crime and punishment". *Changing Attitudes to Punishment - Public opinion, crime and justice*. J. V. Roberts and M. Hough, Eds. Abingdon, Routledge: 184-197.
- Morgan, R. (2002). "Privileging public attitudes to sentencing?". *Changing Attitudes to Punishment - Public opinion, crime and justice*. J. V. Roberts and M. Hough, Eds. Abingdon, Routledge: 215-228.
- Murphy, J. G. (1973). "Marxism and Retribution." *Philosophy & Public Affairs* **2**(3): 217-243.
- Parfit, D. (1984). *Reasons and Persons*. Oxford, Oxford University Press.
- Parfit, D. (2011). *On What Matters*. Oxford, Oxford University Press.

- Pettit, P. (2002). "Is Criminal Justice Politically Feasible?" *Buffalo Criminal Law Review* **5**: 427-450.
- Pratt, J. (2007). *Penal Populism*. Abingdon, Routledge.
- Rawls, J. (1999 [1971]). *A Theory of Justice*. Oxford, Oxford University Press.
- Regan, T. (2004). *The Case for Animal Rights*. Berkeley, University of California Press.
- Roberts, J. V. (2011). "The Future of State Punishment". *Retributivism Has a Past: Has It a Future?* M. Tonry, Ed. Oxford, Oxford University Press: 101-129.
- Robinson, P. H. and J. M. Darley (2007). "Intuitions of Justice: Implications for Criminal Law and Justice Policy." *Southern California Law Review* **81**(1): 1-68.
- Robinson, P. H., R. Kurzban, et al. (2007). "The Origins of Shared Intuitions of Justice." *Vanderbilt Law Review* **60**(6): 1633-1688.
- Ryberg, J. (2010). "Punishment and Public Opinion". *Punishment and Ethics*. J. Ryberg and J. A. Corlett, Eds. New York, Palgrave Macmillan: 149-168.
- Sandberg, J. and N. Juth (2010). "Ethics and Intuitions: Reply to Singer." *The Journal of Ethics* **15**(3): 209-226.
- Schnall, S., J. Haidt, et al. (2008). "Disgust as Embodied Moral Judgment." *Personality and Social Psychology Bulletin* **34**(8): 1096-1109.
- Singer, P. (1995). *Animal Liberation*. London, Pimlico.
- Singer, P. (2005). "Ethics and Intuitions." *The Journal of Ethics* **9**: 331-352.
- Tersman, F. (2009). *Moral Disagreement*. New York, Cambridge University Press.
- Unger, P. (1996). *Living High and Letting Die*. Oxford, Oxford University Press.