

TESTING THE INTUITIVE RETRIBUTIVISM DUAL-PROCESS MODEL: ADDITIONAL DISCUSSION

1. On Goodwin and Benforado (2015)

Like much other research on what motivates punishment, our study relies on the assumption that certain features of a crime or an offender are (more or less) uniquely relevant to retributivism or utilitarianism. This assumption has not gone unquestioned. In particular, Goodwin and Benforado (2015; also Nadelhoffer et al., 2013) have argued that many supposedly retributive features also matter from the perspective of deterrence and incapacitation (two main concerns of utilitarian theories of punishment). To make this point, Goodwin and Benforado asked participants to rate the importance of five features (magnitude of harm, offender motivation, detection rate, publicity of the offense and trial, frequency of the offense) for each of two theories of punishment: retributivism and deterrence. They find that participants rated magnitude of harm and offender motivation (two supposedly retributive factors) as highly and equally important for both theories.

We agree that Goodwin and Benforado raise an important conceptual worry. Nevertheless, we do not quite think that their results highlight an “insurmountable problem for all research in this area” (620). This is for two reasons. First, other research paints a more encouraging picture. Carlsmith, Darley and Robinson (2002), Carlsmith (2006) and Keller et al. (2010) all presented participants with lists of different features of crimes or offenders. They then asked participants to choose which theoretical perspective, retributivism or utilitarianism, each item fit best with. In contrast to Goodwin and Benforado, these authors generally report that participants’ classifications matched up well with what had been assumed by researchers.

Goodwin and Benforado criticize this forced-choice approach, arguing that it leaves open the possibility that “while a majority of participants classified both harm severity/magnitude and perpetrator motivation as more relevant to retribution than deterrence, they may only have regarded these factors as slightly more relevant” (623). If this were the case, however, then their results would actually show too much. This becomes particularly clear for magnitude of harm, which their participants rated as slightly more important for deterrence than for retributivism. Yet in the forced-choice studies of Carlsmith, Darley and Robinson, Carlsmith, and Keller et al., large majorities (upwards of 76%) rated magnitude of harm as most relevant to retributivism. In light of this mismatch, it is not clear to what extent (or if so, in what way) Goodwin and Benforado’s data bear on the interpretation of these studies.

In addition, there is reason to suspect that comprehension may have been an issue with some participants in Goodwin and Benforado. In particular, participants on average rated detection rate, publicity of the crime and its punishment and crime frequency above the importance scale midpoint for retributivism. Yet on any standard reading of that theory, this does not make a lot of

sense. The core tenet of retributivism is that “punishment is justified as an intrinsically appropriate, because deserved, response to wrongdoing” (Duff & Hoskins, 2019). It is hard to see how things like the rate at which crimes of the type in question are detected or how much publicity the punishment of the criminal would attract could have much bearing on the issue of punishment from this perspective. In light of these two points, we are not convinced that there is good reason to think that the evidence of Goodwin and Benforado (2015) is more informative or valid than that reported in the studies they criticize.

Second, to us, it is unclear that whether the folk categorizes different features as retributive or utilitarian really is the central issue here. As we see it, the goal of the research program Goodwin and Benforado (2015) are criticizing is to figure out the extent to which different philosophical theories are instantiated in people’s punitive reactions. But then, in order to assess which features of crimes and offenders are most relevant for retributivism and utilitarianism, it would seem that the most straightforward path is not to ask the folk, but to see what the experts, e.g. philosophers of punishment, have to say about the issue.

So what do philosophers of punishment have to say about the issue? On one hand, most supposedly retributive features used in previous research have indeed been claimed by at least some utilitarians as relevant for their theories (e.g. Hare, 1986; Lyons, 1974). Nevertheless, in our view, the literature does support a differentiation of primarily retributive features from primarily utilitarian features which matches up quite closely with what was used in previous research on what motivates punishment. Retributivists across the board agree that offender intent, magnitude of harm, extenuating circumstances, etc. are directly relevant to retributive theories of punishment (for an overview, Walen, 2020). Similarly, classical utilitarian theories of punishment evaluate the relevance of any factor solely based on the principle of utility (e.g. Bentham, 1830/1998, Chapter 4), meaning that risk of offender recidivism, publicity of the offense and trial, and dangerousness of the offender, for example, are obvious candidates for factors that should be considered when determining punishment (Duff, 2003, pp. 3–7; Fletcher, 1982; Lee, 2007). In contrast, whether utilitarians are truly committed, or even coherently able to, incorporate considerations like intent, magnitude of harm and extenuating circumstances into their theories is a matter of controversy (for an overview, Boonin, 2008, Chapter 2). In fact, some utilitarians simply accept that they need to adopt some form of hybrid theory in which retributive concerns have to be introduced independently of the utilitarian core (e.g. Hart, 1959; Rawls, 1955). The fact that there is no consensus on whether paradigmatic retributive features should or can be built into utilitarian theories of punishment should, in our view, count against the claim that such features are equally a part of both theories.

In summary, we think that the worry raised by Goodwin and Benforado (2015) is an important one. Nevertheless, we have argued that they overstate their case. The assumption that certain features of crimes or offenders can be associated with either retributivism or utilitarianism is

often not implausible.

2. References

- Bentham, J. (1998). The Rationale of Punishment. In T. Vormbaum (Ed.), *Strafrechtsdenker der Neuzeit* (pp. 387–394). Berliner Wissenschafts-Verlag. (Original work published 1830)
- Boonin, D. (2008). *The Problem of Punishment*. Cambridge University Press.
- Carlsmith, K. M. (2006). The roles of retribution and utility in determining punishment. *Journal of Experimental Social Psychology*, 42(4), 437–451. <https://doi.org/10.1016/j.jesp.2005.06.007>
- Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish?: Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology*, 83(2), 284–299. <https://doi.org/10.1037/0022-3514.83.2.284>
- Duff, A. (2003). *Punishment, Communication, and Community*. Oxford University Press.
- Duff, A., & Hoskins, Z. (2019). Legal Punishment. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2019/entries/legal-punishment/>
- Fletcher, G. P. (1982). The recidivist premium. *Criminal Justice Ethics*, 1(2), 54–59. <https://doi.org/10.1080/0731129X.1982.9991706>
- Goodwin, G. P., & Benforado, A. (2015). Judging the Goring Ox: Retribution Directed Toward Animals. *Cognitive Science*, 39(3), 619–646. <https://doi.org/10.1111/cogs.12175>
- Hare, R. M. (1986). Punishment and Retributive Justice. *Philosophical Topics*, 14(2), 211–223. <https://www.jstor.org/stable/43153988>
- Hart, H. L. A. (1959). Prolegomenon to the Principles of Punishment. *Proceedings of the Aristotelian Society*, 60, 1–26. <https://www.jstor.org/stable/4544619>
- Keller, L. B., Oswald, M. E., Stucki, I., & Gollwitzer, M. (2010). A Closer Look at an Eye for an Eye: Laypersons' Punishment Decisions Are Primarily Driven by Retributive Motives. *Social Justice Research*, 23(2–3), 99–116. <https://doi.org/10.1007/s11211-010-0113-4>
- Lee, C. (2007). The judicial response to psychopathic criminals: Utilitarianism over retribution. *Law and Psychology Review*, 31, 125–137.
- Lyons, W. (1974). Deterrent Theory and Punishment of the Innocent. *Ethics*, 84(4), 346–348. <https://doi.org/10.1086/291932>
- Nadelhoffer, T., Heshmati, S., Kaplan, D., & Nichols, S. (2013). Folk Retributivism and the Communication Confound. *Economics and Philosophy*, 29(2), 235–261. <https://doi.org/10.1017/S0266267113000217>
- Rawls, J. (1955). Two Concepts of Rules. *The Philosophical Review*, 64(1), 3–32. <https://doi.org/10.2307/2182230>
- Walén, A. (2020). Retributive Justice. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020). Metaphysics Research Lab. <https://plato.stanford.edu/archives/fall2020/entries/justice-retributive/>