**Free Will Zombies**

**1.**

Libertarians were understandably jubilant when the invention of the I-detector was announced. This device detects in people the presence or absence of an indeterministic process that—as it turned out—always and only attaches to ordinary capacities for reason in any exercise of choice, and thus “effectively establishes a powerful correlation indicating the presence or absence of libertarian free will” (so said the Surgeon General of Bioethics on a label eventually affixed to the device). In other words when the detector shows that indeterminism is absent by producing no reading, effective reasoning capacity is absent too (as empirically demonstrated), so it maps the familiar territory of the presence or absence of basic-desert-bearing responsibility in libertarian terms.[[1]](#endnote-2) Compatibilists in particular grumbled that this demonstrated strict correlation of indeterminism with reasoned choice did not yield a plausible mechanism for how it works, but libertarians just snapped back that Humean causation and laws hadn’t troubled them much before—why should similar rigid correlations be worrisome now? Admittedly there were still many conferences headlining the back-and-forth between event-causal libertarians and their agent-causal cohorts because the I-detector data was epistemically compatible with both, but not many conferences were funded for hard incompatibilists or compatibilists or free will skeptics, though they still gathered together often for informal symposia (in the Platonic sense).

One product of this discovery was a re-energized emphasis on retribution in both personal and public policy matters. Critics complained that even the granted truth of libertarianism was conceptually insufficient to justify retributive practices, but advocates shoved I-data back in their faces as a plausible basis for a claim to moral ultimacy as a necessary condition for separating moral wheat from chaff.[[2]](#endnote-3)

**2.**

So things continued in this way, and the I-detector proved increasingly indispensible for matters of justice. Then something unexpected happened: evolution threw a wrench in the gears with a mutation. Over time a population appeared who were outwardly and (as far as self-reports can tell us anything) inwardly just like people with libertarian free will, but the I-detector yielded null results when trained on them. Of course initially these individuals were prejudicially thought to lack ordinary capacities for reasoning and treated accordingly. However, it eventually became apparent that that was not the case: double-blind studies proved that these individuals were indistinguishable in reasoning capacities from those with libertarian free will.[[3]](#endnote-4) They pondered choices in the same kinds of way, sometimes choosing well, sometimes ill and in comparable proportion to libertarians.[[4]](#endnote-5) Naturally a derisive term emerged to describe them. They were ‘free will zombies.’

Of course the I-detector became useless when attempting to assign libertarian-style ultimate responsibility in any practical sense to free will zombies, especially when it became a tactic for some more unsavory zombies to feign insanity in order to escape retributive blame with the assist of the I-detector.[[5]](#endnote-6) But this was also something of a philosophical windfall for those few struggling compatibilists and hard incompatibilists and free will skeptics left (usually employed as baristas rather than academics since such non-zombie ‘free will flat-earthers’ chose to deny obvious truth ironically of their own free will[[6]](#endnote-7)). They questioned whether in cases where libertarians and zombies behaved identically by all standards of assessment except for different I-detector readings, would it be the case that retributive treatment (based primarily on incompatibilist grounds) is fairly applied to zombies, since they lack what had become an accepted feature of praise- and blameworthiness? Would two systems of justice (perhaps modeled on the juvenile/adult systems) be necessary for equivalently rational defendants who were precisely distinguishable in just one way—the presence or absence of indeterministic I-properties? Or, were distinguishable zombies *really* *people* worthy of retributive concern? And perhaps more pertinently some whispers grew into a shout: was the system of retributive justice based on libertarianism somehow faulty from the get-go?

**3.**

Typically discussions of free will pose distinct accounts of human nature as inhabiting very different possible worlds—deterministic, indeterministic, fatalistic, mysterian, and so on. Various sets of intuitions and argumentative schema align with one world as favored and others thus decried, and dialectical stalemate among disputants is a familiar ellipsis.[[7]](#endnote-8) What gets lost in the fray is the prospect of finding some direct way to compare such accounts in more *practical* terms. My little story about zombies attempts to remedy that lack of comparison, at least in part.

Thus libertarianism is planted squarely in just one world and given means (the I-detector) to rationally justify—at least by strict correlation—belief in its truth. Consequently, retributive practices based on libertarian intuitions are justified to that extent as well. But then the zombies come on the scene, who appear to be as much human as anyone else and in particular are reasonable beings indistinguishable from libertarians save having no discernible indeterministic processes involved in their rational choices. One might assume that this indicates that zombies are deterministic, but it is sufficient to say that whatever they are, they are not I-type libertarian beings.

So when we have a badly-behaving zombie on our hands, what is the most rational way to deal with that? Since zombies at best are only capable of having some form of compatibilist free will, or at worst none at all, then libertarian-based retributive practices appear to be off the table. Of course any forward-looking consequentialist-based deterrence or treatment used for libertarian criminals would be extended to criminal zombies on presumably equally just grounds, but not libertarian retribution. Still, if such retribution continued to be applied to I-verified libertarian criminals, then a two-tiered justice system would exist.

In that case, no doubt some libertarians (probably defense attorneys) would claim reverse-discrimination. After all, did they (or their libertarian clients) ask to be endowed with I-properties? Why should they, even as criminals, be treated differently—and frequently more harshly—for something they could not help but possess but otherwise made them no different from better-treated zombies? Of course they seldom stressed that praise heaped on them for trivial acts of kindness *might* mean so much more than that accorded Nobel Prize-winning zombies, but praise after all is so eminently ignorable compared with questions of doing hard time and capital punishment.[[8]](#endnote-9) Note, however, that such protests would be made not primarily on metaphysical grounds, but *practical* ones, because libertarians would thus claim that their I-properties were regarded as something meaningful for justice when it was really only an irrelevant point of reverse prejudice as compared with the more humane treatment of zombies, with whom they shared so much otherwise.[[9]](#endnote-10)

Other libertarians of course might wallow in that very difference and make I-properties a point of pride. ‘Take your stinking paws off me you damn dirty zombie!’ Nevertheless, it seems mind-boggling that one such small difference could legitimize I-elitism, especially since there had never been a conclusive account of *exactly* *how*—and now more strongly *if*—I-properties were significantly involved in making rational decisions.[[10]](#endnote-11)

I should hope that soon cooler heads would prevail, and everyone would realize that for all *practical* purposes the minor difference of those I-properties was irrelevant for being considered human and fully responsible, thus resulting in something like real moral progress.[[11]](#endnote-12) If free will zombies really possess all other salient features for full inclusion in the moral sphere, then it is mere prejudice to call them out as ‘zombies’ for differential—and what reverse-discrimination-zombie-envying libertarians would even call *preferential*—moral treatment. Zombies are rational decision-makers who deserve attributions of responsibility, even if less pointedly than in the sense of libertarian basic desert *ultimacy* riveted to I-properties, which palpably the existence of zombies demonstrates is only *circumstantially* allied to reason in I-verified instances. This makes justification for retributive responsibility *practically* incoherent, and favors by default consequentialist or compatibilist theories of responsibility and punishment for everyone based on the preeminent zombie example.[[12]](#endnote-13)

What is the moral of this tale? Just this: the free will problem has too long ignored the role that *moral* *pragmatism* might play in resolving it.[[13]](#endnote-14) Once brought squarely into the picture as with this one-world direct comparison of libertarians with free will zombies, we clearly see that it makes no practical moral difference whether we in our own world have I-properties or turn out to be just those I-less zombies, and thus emphasis on such a property that is taken by itself to entail retributivism is unjustified.14

**References**

[1] Campbell, J.K. and White, V. A. Humean Compatibilism, Dual Ability, and Simple Indeterminism. (in preparation)

[2] Fischer, J.M. TheMetaphysics of Free Will: An Essay on Control. Oxford: Blackwell Publishers. 1994.

[3] Hodgson, D. Rationality + Consciousness = Free Will. Oxford University Press. 2012.

[4] Pereboom, D. What Are We Fighting About? Flickers of Freedom (<http://agencyandresponsibility.typepad.com/flickers-of-freedom/2013/02/what-are-we-fighting-about.html>) accessed 2/15/13.

[5] Smilansky, S. Free Will and Illusion. Oxford University Press. 2000.

[6] Watson, G. Agency and Answerability. Oxford University Press. 2004. 184-196.

[7] Wolf, S. 1980. Asymmetrical Freedom. *Journal of Philosophy*. 77: 151-66.

**V. Alan White**

**University of Wisconsin—Manitowoc**

**alan.white@uwc.edu**

1. Throughout the rest of this paper it’s assumed that any putatively responsible party has in fact committed an act for which he or she is possibly responsible in general free will terms, whether parsed by libertarianism or compatibilism. Additionally, I’d note here that my thought experiment could very well run in the opposite direction: we could start with deterministic rational beings verified via a D-detector who were thus assessed by compatibilist responsibility criteria, and then later introduce comparably rational non-deterministic beings by mutation. There would still be no practical reason to assess the latter differently even if they came to be regarded as “angels” in miraculously choosing indiscernibly from D-beings (in fact given here the *base* view of compatibilism, no one would *pragmatically* argue the *superiority* and *more-deserving* nature of indeterministically choosing beings). But in any case the obvious disadvantage is that that story couldn’t invoke the cool contemporary reference to zombies! [↑](#endnote-ref-2)
2. I owe Derk Pereboom much credit for this point, and in fact for raising issues in (2013) that gave birth to this whole zombie tale, though he would likely reject any sense that he ultimately deserves it. [↑](#endnote-ref-3)
3. Note that I allow that zombies are *only* rational. This simplifies the narrative and is within bounds of a possible world posit of the nature of the mutation. [↑](#endnote-ref-4)
4. Thus there is no empirical basis for claiming that libertarians are *disadvantaged* by potentially luck-bearing I-properties over zombies. [↑](#endnote-ref-5)
5. Of course genetic tests could disclose that they were mutants, and decisively so. [↑](#endnote-ref-6)
6. The late David Hodgson details how libertarian choice may be tied to choosing beliefs in (2012). [↑](#endnote-ref-7)
7. John Fischer (1994) has written on the nature of some such stalemates. [↑](#endnote-ref-8)
8. The *locus classicus* of this point is of course Susan Wolf 1980. [↑](#endnote-ref-9)
9. One distinct metaphysical facet of this complaint is about being *meta-unlucky*: libertarians were meta-unlucky to inhabit a world where they would be treated differently as compared with their zombie companions. Joe Campbell and I set out the significance of metaluck in a forthcoming paper (Campbell and White). Still, our complaint is not pragmatically based as is one focused on a one-world comparative case as the zombie example here. [↑](#endnote-ref-10)
10. This is why the zombie-world account assumes the Humean ‘black-box’ problem of how we should understand that libertarian indeterminism is a claimed necessary condition of basic-desert responsibility given its prima facie pre-zombie correlation with reasoned choice, which then gives rise to the compatibility of the story with both event- and agent-causal forms of libertarianism since neither view has closed the deal (so far) on an account of that condition. [↑](#endnote-ref-11)
11. This is Gary Watson’s more eloquently expressed complaint about libertarianism in (2004). My zombies versus libertarians scenario just ups the morally pragmatic ante on the confrontation in one possible world. [↑](#endnote-ref-12)
12. What would turn the tide of argument back to the libertarian’s favor? Merely solve the ‘black-box’ problem of how indeterminism specifically contributes to the ultimacy of basic-desert responsibility. Appeal to the I-detector can’t do it, even with the pre-zombie strict correlation of reason with I-properties, since that is undermined with the later appearance of equally reasonable zombies. Good (meta-) luck libertarians! [↑](#endnote-ref-13)
13. Saul Smilansky (2000) might well put his stamp of approval on this point in a general way with his illusionism. The zombie world arguably dispenses with the need for global illusion due to the I-detector, so that at least Humean-type libertarians might genuinely exist. Smilansky might then wish to argue that illusions need only be required for the equal treatment of zombies, though the resultant asymmetry of illusionism would at the least make for a morally very awkward world.

    14 I thank Derk Pereboom, Troy Cross, and Neil Levy respectively for inspiration, clarification, and criticism. Levy notes in correspondence that this paper might be better presented as a straightforward argument for compatibilism, and I do not deny that. My point that moral pragmatism is best enlisted as arbiter for the equal treatment of co-existing Libertarians and zombies could well be extended as an *axiological* basis for the preeminence of pragmatic compatibilist moral values over more absolutist incompatibilist ones. This paper, preferring to focus on criticizing *moral* incompatibilism, is neutral about *metaphysical* incompatibilism. [↑](#endnote-ref-14)