

## Law and moral agency in *De libero arbitrio I*

IVAN WELTY

*Department of Philosophy, Willamette University, Salem, Oregon, 97301, USA*  
e-mail: [iwelty@willamette.edu](mailto:iwelty@willamette.edu)

**Abstract:** Augustine’s account of postlapsarian human moral agency in Book I of *De libero arbitrio* is analysed more fully than heretofore. Consideration is given to Augustine’s Stoic antecedents and, following a suggestion by R. J. O’Connell, a comparison with Kant’s moral philosophy is developed. The result is a more nuanced understanding of Augustine’s account of moral agency in the early period. Whether that account persists into Augustine’s later work is left an open question.

### Overview

In 1970, R. J. O’Connell observed a ‘strangely Kantian’ tone in the *lex aeterna* of Augustine’s *De libero arbitrio I*. Although nearly fifty years have passed, O’Connell’s tantalizing observation remains undeveloped. My purpose in this article is to bring out more fully the role played by the concept of law in *De libero arbitrio I*, including aspects that indeed might be called Kantian. My method is reconstructive. First, I examine a proper fragment of Augustine’s account of moral agency, expressly not including his concept of law, and I note some of the fragment’s implications. In brief, knotty problems emerge, including, rather startlingly, a puzzle familiar from Stoic philosophy. I then restore Augustine’s concept of law and observe the result – in brief, a neat resolution of all of the problems generated by removing it. The role of the concept of law in Augustine’s account of moral agency makes it clear that, far from being superficially Kantian, as one might surmise from O’Connell’s comment, Augustine’s complete account of moral agency in *De libero arbitrio I* is in fact deeply Kantian. There are, however, important differences between the two approaches, two of which I conclude by indicating.

## Background

We begin by setting the stage. Augustine reflected on the problem of postlapsarian human moral agency continuously over the course of his long career. Whether or not his commitments changed over that time, his emphasis certainly did: he began by insisting that postlapsarian humans are free and ended by insisting that humans are doomed to sin. This change is partly explained by the positions against which he was disputing. His first efforts were directed against the Manichaeans, who explained sin as the result of a power of darkness overcoming the soul. Against this, Augustine argued that postlapsarian humans are wholly responsible for the evil they do, a claim that presupposed, or so he took it, that humans are free. Augustine's later works were directed against the Pelagians, who held that humans can earn salvation by rightly exercising their free will. Against this, Augustine argued that postlapsarian humans are saved by divine grace alone, a claim that he took to presuppose an otherwise insuperable propensity to sin. The Pelagians were quick to note the apparent similarity of Augustine's later position to Manichaeism, and they enthusiastically cited Augustine's earlier work against him. Despite thus having ample opportunity to disavow his youthful commitments, Augustine never did. The absorption of those commitments into his mature philosophy is a subject of extraordinary and, one may still say, urgent interest, but I shall not examine it here.

The first book of *De libero arbitrio*, written in 387–389, soon after Augustine's baptism, belongs squarely to the early period. He is principally concerned in this book with finding agents responsible for the evil they do, but he makes it clear, as we will see, that postlapsarian human freedom extends also to the attainment of a narrow class of goods. The book takes the form of a dialogue between Augustine and his friend Evodius. As the book opens, Evodius is seeking to learn whether God is the ultimate cause of the evils committed by postlapsarian humans; Augustine directs the conversation as one who has already deeply considered the question. The two begin by specifying the class of evils concerned in their discussion. We take up their discussion at this point.

## Puzzles

Just punishment is evil in the *suffering*, Augustine and Evodius agree, but not in the *doing*. This acquits God of doing evil when he punishes sinners, but it leaves unclear what, if not suffering, makes evil *done* blameworthy. '[A]ll evildoings are evil precisely because they come about from lust, that is, from a blameworthy desire';<sup>1</sup> in turn, blameworthy desire is 'the love of things one can lose against one's will [*earum rerum amorem, quas potest quisque inultus amittere*]'.<sup>2</sup> Augustine calls those things that one can lose against one's will *temporal*, while those that cannot be so lost he calls *eternal*.<sup>3</sup> These definitions are rendered without qualification.

When one thinks of temporal as against eternal desires, what leap to mind are such morally dubious desires as those for physical pleasure and money. But Augustine's definition is radically more encompassing than the usual, with troubling implications for his account of moral agency. Three implications stand out.

First, on Augustine's unsupplemented definitions there can be no distinction between evils as to their severity. Temporal aims are temporal without degree, and so likewise are the evils they constitute. But surely negligent homicide, while bad, is less bad than premeditated murder.

Second, Augustine's definitions afford no moral distinction between temporal desires that are, on their face, some of them good and some of them bad. For example, one can no more save the life of a drowning child on merely willing it than one can drown a child on merely willing it, yet one should like to distinguish ethically between those willings. On Augustine's definition, they are both temporal, hence blameworthy.

Third, even where the account correctly identifies an evil, it does so on peculiar grounds: willing an innocent's drowning counts as evil, as one should expect, but the evil consists in willing something that might not come to pass despite one's willing it. Similarly, any good is good on account solely of the agent's ability to secure it at will.

To my knowledge, no philosopher before Augustine commits himself to the second and third implications, but the first implication is in essentials a well-known 'paradox' attributed by Cicero to the Stoics. As Epictetus puts it, more than a century after Cicero, a person should so direct himself that he 'may never fail to get what he desires, nor fall into what he avoids'.<sup>4</sup> The concept of eternity in *lib. arb.* has its root here: the will should be trained on those objects that it cannot fail to obtain. Now, the Stoics, as observed by Cicero, held the reward for so training the will to be maximal: 'No one who is totally self-reliant, and contains within himself all that he owns, can fail to be completely happy.'<sup>5</sup> By the same token, every failure of will, no matter how small, counts as evil: 'We must consider what custom forbids as a crime, but what law forbids as sacrilege. 'Even in little things?' Yes, since we may not of course be able to shape the way things are, but we can control our own minds.'<sup>6</sup> By avenues that Cicero leaves dark – Epictetus, too, for that matter – these considerations lead to the 'Stoic paradox' identified by Cicero: 'all evils are equal, and similarly all right actions [*Aequalia esse peccata et recte facta*].'<sup>7</sup> This is the first troubling implication mentioned above.<sup>8</sup>

Be Augustine's historical antecedents what they may, the first implication of Augustine's definitions is indeed troubling. An account of moral agency should, where possible, avoid flattening evil in just that way. Something similar goes for the second and third implications: an account of moral agency does well to avoid them. We will find it useful, in the discussion that follows, to regard these implications as *puzzles* raised by Augustine's account of moral agency, puzzles that his account should, if possible, supply the resources to solve.

### Solution

We can think of the puzzles identified in the previous section as laying down conditions on philosophically and exegetically satisfactory readings of *lib. arb.* So formulated, the conditions are these:

1. A reading of *lib. arb.* should furnish a distinction between worse and less bad blameworthy desires (first puzzle);
2. A reading of *lib. arb.* should furnish a distinction between, as it were, bad blameworthy desires (desire for a drowning) and good blame-worthy desires (desire for a child's physical well-being) (second puzzle);
3. A reading of *lib. arb.* should furnish plausible grounds for judging something evil or, likewise, good (third puzzle).

Together these conditions suggest another:

4. A reading of *lib. arb.* should identify, wherever there are, as it were, good blameworthy desires, blameless desires that conduce to the same ends. So, for example, if it is evil to desire the physical well-being of an innocent, there should be some blameless desire that motivates agents to preserve the physical well-being of innocents. In short, a reading of *lib. arb.* should supply at least the outlines of a plausible theory of practical reason.

The fifth and final condition on satisfactory readings of *lib. arb.* is the general one of fealty to the text:

5. A reading of *lib. arb.* must be not only consistent with Augustine's text but, to the degree possible, firmly rooted in his expressed views.

For a reading of *lib. arb.* satisfying these conditions, we turn to Augustine's conception of the law. First, however, it will pay to motivate that conception with an example.

Suppose that you are walking along a river when you see a drowning child. The child is near enough that you might save him, and you dive in. If we take as given the *lib. arb.* doctrines so far discussed, the physical well-being of the child is temporal and your desire for his well-being is evil. What possible justification could there be for this? The conclusion seems perverse, but a motivation is available. Suppose that, despite your best effort, you fall one foot short of saving the child's life. Clearly, there must have been something in the circumstance that it was right for you to do. But it could not have been actually *saving the child's life* – for, if it had been, you would have failed to do the right thing and, so, committed evil, your unsparing effort notwithstanding. Identifying the child's well-being with the good it was yours to achieve renders the rightness or wrongness of your action a matter of luck.

Note that a retrospective strategy fails against this scenario. For example, let it be laid down after the fact that your falling one foot short was the right thing to do all along. Then doing right depended on your not suffering a stroke at two feet short. But surely your doing right could depend on no such thing. Since at no point in one's progress towards saving a drowning child is one immune to bad luck, no single step along the way can constitute the right thing to do. Indeed, a retrospective strategy must fail even leaving aside the possibility of bad luck: right action should reach as far as the child's actual rescue, circumstances permitting.<sup>9</sup>

What is the right thing to do in the case of the drowning child? Is an answer available that satisfies the five conditions? Here we take up Augustine's discussion of law. But Augustine prefigures that discussion in a short discourse on learning, which it will be useful to review first.

Evodius' initial thought, when embarking on the search for an 'author' of evil, is that postlapsarian humans learn to do evil, hence that evil can be traced to a teacher. Augustine rejects this, on the grounds that 'if [someone] is evil he is not a teacher, and if he is a teacher he is not evil [*Si enim malus est, doctor non est: si doctor est, malus non est*].'<sup>10</sup> The principal steps toward this conclusion are characterizations of *disciplina* (learning) and *intelligentia* (understanding) as success-terms: something counts as learning only if it improves understanding, and something counts as understanding only if it is good. Whatever goes on in a case that looks like evil instruction, its true description must be something else. This not only removes teachers as possible authors of evil, but supplies the outline of a negative criterion for evil: does *coaching* in the matter – to put it neutrally – count as genuine *instruction*? If so, it cannot be evil. This outline remains empty until grounds are supplied for telling instances of genuine instruction from counterfeits. Augustine's reflections on pedagogy – the principal text would be *De magistro*, contemporaneous with *lib. arb.* – are not by and large directed to this question, but the conception of learning he elaborates in them, on which learning includes not just knowing but also rational comprehension, supplies the raw materials for an answer.

We need enter no further on Augustine's theory of learning. The point is simply that Augustine's treatment of law has a parallel structure. 'Law' [*lex*] is a success-term: 'a law that is not just does not seem to me to be a law [*Nam mihi lex esse non videtur, quae iusta non fuerit*].'<sup>11</sup> Again this supplies the outline of a criterion of the good: something is good if it is mandated by a law. The question is what counts as a law. By a consideration of various cases, including some of apparent inconsistency – a people should govern itself, but a benevolent dictator should put order as needed – Augustine and Evodius arrive at an application to laws of the temporal/eternal distinction: 'let us call a law temporal if, although it is just, it can justly be changed in the course of time', while a law is eternal in so far as 'it should always be obeyed; through it good people deserve a happy life and evil people an unhappy one; and finally through it temporal law is both rightly enacted and rightly changed'.<sup>12</sup> All laws are eternal to some degree, otherwise

they would not count as laws. For example, a 'law' remaining in effect only while agents comply with it does not count as a law.<sup>13</sup> That law is most eternal 'according to which it is just for all things to be completely in order'.<sup>14</sup> Thus the question of the content of the legal criterion for goodness is reduced to the question 'What counts as order?'. We will return to this question later; here, we simply note that, for Augustine, something is ordered to the degree that it exists.

Now to the conditions. We must, according to (1), distinguish among blameworthy desires as to their severity; according to (2), distinguish, as it were, bad blameworthy desires from good blameworthy desires; according to (3), furnish plausible grounds for judgments of good and evil; according to (4), identify, for any given good blameworthy desire, a blameless desire that conduces to the same practical end; and, according to (5), remain faithful to Augustine's text. The conditions can be met. Consider your blameworthy desire for the well-being of the drowning child and a law to the effect that, say, strangers in peril should be helped.<sup>15</sup> Willing the law's fulfilment conduces to the same practical end as blameworthy desire for the child's physical well-being. Thus we meet the condition in (4). In turn, compliance with the law distinguishes your desire to save the child's life from a morbid desire to, say, witness a drowning, satisfying the condition in (2). As to (3), we required plausible grounds for judgements of good and evil, and now love of the law, implying willing submission to its obligations, furnishes those grounds, provided, as Augustine does provide, that laws are just.

The condition in (1) requires more careful handling. Augustine expressly recognizes a ranking of evils:

The law gives the people whom it governs permission to do lesser evils in order to avoid greater ones. It is much more civilized that someone who plots against another's life be killed rather than the one who is protecting his own life; it is much more barbarous that someone unwillingly endure a rape than that the assailant be slain by his intended victim.<sup>16</sup>

The question is whether rational grounds are available for the ranking. For present purposes, it suffices to observe that the temporality of a law mirrors the severity of the evil it proscribes: where two laws conflict, the less eternal will give.

The beginnings of a solution to the puzzle in (1) are now visible. Suppose that we have a law  $L$  prohibiting  $a$  and a law  $L'$  prohibiting  $a'$ , and suppose further that  $L$  is more eternal than  $L'$ . Nothing stands in the way of our counting desire for  $a$  as more blameworthy than desire for  $a'$ . Admittedly, not all laws are directly comparable, or if they are, it is not always clear that one is more eternal than the other. For example, is a law that prohibits larceny more eternal than a law that prohibits fraud? Perhaps we may lay it down that in such a case the laws are equally eternal. Another provision is required to compare laws formulated positively (mandates) and laws formulated negatively (prohibitions). Assuming such concerns can be met, however, it seems that we have at least begun to meet the condition in (1): in brief, a desire is more blameworthy the more eternal the law that we violate in acting on it.

We have met the conditions in (1)–(4), at least to a first approximation. Moreover, we have introduced nothing essential from outside of Augustine’s text, and so we have met the condition in (5), too.

### **Eternal law and the categorical imperative**

We have seen that omitting Augustine’s notion of law from his account of moral agency yields several puzzles, and that restoring his notion of law solves them. The notion of law that turns this trick is the one that long ago struck R. J. O’Connell as Kantian, and its Kantian aspect is no accident.<sup>17</sup> In this section, we see how deep the resemblance between Augustine and Kant goes.

Augustine recognizes, on one side, our absolute ethical responsibility – bad luck is never an excuse – and, on the other side, the brute fact of our limited power. He steers a course between them by admitting the will’s *complete* dominion over a *restricted* domain. A child might drown despite one’s willing the contrary, but there can be no such thing as failing against one’s will to will the fulfilment of the law.<sup>18</sup> So a good will is eternal and, indeed, the only good of which humans are capable (the case is different with God, for whom there is never a gap between willing and attaining what is willed). Precisely this conception of human will undergirds Kant’s ethics, on which, famously, only a good will is unqualifiedly good. Again, this similarity is no coincidence: Kant’s conception of free will rides between the same constraints as Augustine’s, namely, moral responsibility, on one side, and impinging circumstance, on the other.

If the notion of law is to mediate between moral responsibility and impinging circumstance, the notion must be, so to say, flexible on the side of circumstances and rigid on the side of the will. By this I mean that it must be possible for an agent to count as law-loving under a wide variety of circumstances and in view of any number of possible actions taken by the agent. Augustine and Kant achieve this two-sidedness in part by conceiving laws hierarchically, according to the range of circumstances they comprehend. So, for example, a law-loving agent might be released from the obligation to prevent a theft provided there is a sufficient risk of injury in acting to prevent it. For Augustine, release would proceed in accordance with the *lex aeterna*, by which ‘it is just for all things to be completely in order [*ordinatissima*]’.<sup>19</sup> According to (one formulation of) Kant’s categorical imperative, which governs the activation and deactivation of maxims in light of contingent circumstance, one should ‘act always in accordance with that maxim whose universality as law you can at the same time will’, for this is ‘the single condition under which a will can never be in conflict with itself’.<sup>20</sup> In both cases, a condition on lawfulness, preserved under all circumstances, is raised to the status of a super-law regulating lower-level laws. Augustine’s condition is order, Kant’s is consistency.<sup>21</sup>

We have seen that, for both Augustine and Kant, good faith failure is never grounds for blame, rather blameworthiness depends entirely on the will’s

disposition. We have seen, further, that the ultimate test of good will is not the same for Kant as for Augustine. Augustine asks whether one sincerely desires the maximization of order, while Kant asks whether what one desires is consistent with the idea of a universal law mandating it.

A.-H. Chroust once imputed to Augustine 'a formalistic and abstract definition of the nature of right and justice'.<sup>22</sup> Consideration of two further differences between Augustine's and Kant's conceptions of the law will show, however, that while Chroust's imputation might be true of Kant, it is certainly false of Augustine.

### **First difference: ethical bivalence**

In elaborating the example of the drowning child, we assumed without further ado several non-trivial ethical principles. The first such principle is that anything morally evaluable at all must have exactly one of two values, namely, good or evil. We took this principle for granted when we assumed that in failing to do good, you did evil. In semantics, the principle of bivalence holds that every proposition has exactly one of two truth-values, true or false; let us call the corresponding ethical principle *ethical bivalence*. Whatever might be said for or against ethical bivalence, it recurs throughout Augustine's corpus, if only implicitly, and to my knowledge he never evinces any interest in relaxing it.<sup>23</sup>

Now to understand the role played by ethical bivalence in Augustine's account of moral agency, we must see how ethical bivalence works together with a second principle unstated in the case of the drowning child. According to this principle, we always have it in our power to do the right thing. Only given this principle does it follow from the fact that your failure to save the child was in good faith that saving the child was not the right thing to do. This principle, call it P, is actually entailed by Augustine's basic definitions: after all, if the good were beyond one's reach then, absurdly, desire for it would be temporal and, so, evil. Note that the contrary of P, call it P', according to which it is always in our power to avoid evil, also follows from Augustine's definitions: if some evil turned out to be unavoidable then, absurdly, its avoidance would be temporal and, so, evil. From P' it follows that we are blameworthy only if we are responsible.<sup>24</sup> By ethical bivalence, anything that is ethically evaluable is either good or evil; by P, the good is always within reach.

We must now consider the relation between ethical bivalence and principle P. Principle P involves an uncertain level of modal commitment: if we say that it is always in our power to do the right thing, it may be asked what we mean by 'in our power'. Is it always *logically* possible? *Physically* possible? Possible *given the particular circumstances*? Or is it, perhaps, *attainable on merely willing it*? (The same holds, *mutatis mutandis*, for P': we must ask in what sense it is always in our power to avoid evil.) The scope for right action in a given case will depend, in part, on the degree of modal commitment in P. Recall your good faith failure to save the drowning child. If we *both* interpret P liberally, so that



saving the child might count as the good it was in your power to do, *and* maintain ethical bivalence, it is hard to see how we can avoid charging you with evil for failing to save the child. If, on the other hand, we relax ethical bivalence or interpret P more narrowly, the charge can be avoided. When Augustine trims the domain of the ethically evaluable so as to exclude such items as your failing to save the child, he is, more exactly stated, maintaining ethical bivalence while interpreting P narrowly.

But that is not quite the whole story. Augustine starts, as I have suggested, by admitting into the domain of the ethically evaluable all willings and assigning the value evil to every willing of something temporal. He thus undoubtedly advances what Mann has memorably called an ‘inner-life ethics’, that is, a view on which ‘moral appraisal should focus primarily – perhaps exclusively – on the agent’s mental states themselves’.<sup>25</sup> Having settled the core of the domain of the ethically evaluable, however, Augustine extends that domain to include the will to attain lawfully mandated temporal aims, then again to include those temporal aims themselves (the physical well-being of a child, for example), and finally to include the instruments necessary to attain those temporal aims – provided, always, that the aims and instruments are desired not for their own sake. Law leverages this domain-extension at every stage, by distinguishing the blameworthy willing of temporal aims *as ends in their own right* from the morally acceptable willing of temporal aims *for the sake of a law’s fulfilment*.<sup>26</sup> We may say, therefore, that for Augustine temporal aims are in fact morally evaluable, but only derivatively.

With Augustine’s account of moral agency now clearly in view, we turn to Kant. Any statement of Kant’s position on these matters must be, as so often with him, the tip of an iceberg. However, it can be said quite definitely that Kant accepts three categories of action, namely, prohibited actions, required actions, and indifferent actions (*adiaphora*), and hence that Kant rejects ethical bivalence.<sup>27</sup> Once ethical bivalence is rejected, it becomes possible, of course, to acquit you of committing evil in failing to save the drowning child while maintaining that saving the child was the good it was in your power to do, for it becomes available to say that failing to save him was merely indifferent. Kant, however, declines to count saving the child as the good it was in your power to do, instead maintaining, with Augustine, that whatever good it was in your power to do, it was in your power to do by merely willing it.

In sum, both Augustine and Kant maintain P, but only Augustine affirms the intricately related principle of ethical bivalence.

### **Second difference: grounds of law**

As mentioned earlier, Augustine holds that all genuine laws are just. But how are we to tell genuine laws from counterfeits, or, as we might be inclined to call them, ‘bad laws’? It cannot be according to whether the ends they mandate are good, given that we are making out goods as ends mandated by law.

To set out the problem clearly, suppose that a 'law' were passed mandating that blue-eyed children be drowned. We might like to argue that this 'law' is unjust, hence a counterfeit, on the grounds that it is evil; but we have specified that the grounds for judging something evil are proscription by law, in fact, that an evil is worse the more eternal the law that proscribes it. If there are no grounds for judging something good or evil independently of the law, there is little we can bring against this abject 'law' beyond brute insistence that it is a counterfeit.

Augustine does, however, offer independent grounds for ruling out this 'law', when he identifies goodness with being and being with order.<sup>28</sup> This twofold identification makes available a conception of law as an existence-promoting norm. If evil is the privation of being, drowning blue-eyed children represents an obvious and great evil, for rather than promote existence it reduces it. Indeed, it reduces being of a particularly high order, for Augustine takes it, on scriptural grounds, that human beings are the image of God. This answers the further question, Why can we not freely will evil, or again, why is evil itself not eternal? Augustine's answer, detailed after *lib. arb.*, most forcefully in *Confessions*, is that an evil will undoes *itself*, as it declines into habit and, eventually, degeneration so complete that only divine grace can restore it.

We have asked what test Augustine offers to distinguish genuine laws from counterfeits, and we have answered that among principles of action, laws and only laws – that is, just and only just principles of action – promote order, hence being, hence the good. With Kant, the question must be different. He already gives us a partial test for morally binding principles of action in the categorical imperative: if a principle of action is moral, it can be consistently universalized. Augustine would readily agree to this; after all, a principle of action that would obligate one to both follow it and also not follow it hardly merits recognition as a law. But how do we tell morally binding universalizable principles of action from universalizable principles of action that are not morally binding? Kant holds that no further test is needed: consistent universalizability in a principle of action is not only necessary for its morality but also sufficient.<sup>29</sup> Whether Kant is right in holding that consistent universalizability is sufficient for morality is a deep question; for our purposes, it is enough to observe that there is no *obvious* reason why there could not be a consistently universalizable but *immoral* principle of action, for example, one for whose sake a person might passively observe a drowning child.

It is now clear why Chroust rightly could have attributed a 'formalistic and abstract definition of the nature of right and justice' to Kant but not to Augustine. The difference between them on this score is, in fact, deep, and not without consequences for their other commitments. I conclude this section by considering one such consequence.

Why should an agent *want* to be good? Augustine has a ready answer: to be good is simply to be, and surely being is preferable to non-being. This answer turns on the interpretation of evil as the privation of being, a non-obvious interpretation, to be sure, but one that I will suppose can be carried through. If the question is

pressed why being should be loved, Augustine will appeal, as he does especially in his later writings, to the divine grace whereby humans come to know love at all.<sup>30</sup> Kant, for his part, roots moral agency so firmly in pure reason that the question of agent interest emerges only as a devilishly slippery loose end. Here is Kant taking up the thread at the end of the *Groundwork*:

In order for a sensibly affected rational being to will that which reason alone prescribes the 'ought', there obviously must belong to it a faculty of reason to instill a feeling of pleasure or satisfaction in the fulfillment of duty, hence a causality of reason to determine sensibility in accordance with its principles. It is entirely impossible, however, to gain insight, i.e., to make comprehensible a priori, how a mere thought that contains nothing sensible in it would produce a sensation of pleasure or displeasure; for that is a particular kind of causality, of which, as of all causality, we can determine nothing at all a priori, but rather we have to ask experience alone about it. But since experience can provide no relation of cause to effect except that between two objects of experience, but here pure reason, through mere ideas (which yield no object at all for experience), ought to be the cause of an effect which obviously lies in experience, it is entirely impossible for us human beings to have an explanation how and why the universality of the maxim as a law, hence morality, should interest us.<sup>31</sup>

Despite its appearance as a loose end, Kant allows that the problem of 'moral feeling' brings us to 'the supreme boundary of all moral inquiry', a horizon at which 'all knowledge has an end' but on the other side of which pure reason must be arousing ardour by some means akin to causation.<sup>32</sup> Augustine, too, recognizes a sharp distinction between the hidden and the visible: 'From a depth which you cannot see rises everything that you can see.'<sup>33</sup> But what we do not see is God, to whom we must credit *inter alia* our affective motives. It is no part of Augustine's thought that the claims of morality derive ultimately from pure reason, much less that those claims are self-legislated. He thus leaves himself an appeal that Kant does not have.

### Conclusion

Augustine's account of moral agency in *De libero arbitrio* is strikingly Kantian, especially in the role it accords to law. This was clearly recognized by O'Connell nearly fifty years ago, but the similarity was never pursued to its roots. We have seen that those roots reach deep into Augustine's account of moral agency. Removing from Augustine's account of moral agency his concept of law, what remain are some thorny puzzles, including one familiar from Stoic philosophy. Restoring the concept not only solves the puzzles, it renders Augustine's entire account of moral agency recognizably Kantian. This similarity is not coincidental: if one is committed to the view that moral agents are fully responsible for their moral success or failure, then some way must be found to accommodate bad luck, and for that purpose the concept of law is useful, indeed. There are, however, important differences between the two accounts of moral agency, some of which I have explored. It has not been my purpose to judge which account is better.<sup>34</sup>

## References

- AUGUSTINE, SAINT (1991a) *Confessions*, H. Chadwick (tr.) (Oxford: Oxford University Press).
- AUGUSTINE, SAINT (1991b) *The Trinity (De Trinitate)*, J. E. Rotelle (ed.) & E. Hill (tr.) (Hyde Park NY: New City Press).
- AUGUSTINE, SAINT (1992) *Sermons III/5 (148–183) on the New Testament*, J. E. Rotelle (ed.) & E. Hill (tr.) (New Rochelle NY: New City Press).
- AUGUSTINE, SAINT (1999) *Unfinished Work in Answer to Julian*, J. E. Rotelle (ed.) & R. J. Teske (tr.) (Hyde Park NY: New City Press).
- AUGUSTINE, SAINT (2010) *On the Free Choice of the Will, On Grace and Free Choice, and Other Writings*, P. King (tr.) (Cambridge: Cambridge University Press).
- BYERS, S. C. (2012) *Perception, Sensibility, and Moral Motivation in Augustine: A Stoic-Platonic Synthesis* (Cambridge: Cambridge University Press).
- BYRD, B. S. (2015) 'The elusive story of Kant's permissive laws', in L. Denis (ed.) *Kant's Lectures on Ethics: A Critical Guide* (Cambridge: Cambridge University Press), 156–169.
- CHROUST, A.-H. (1944) 'The philosophy of law of St. Augustine', *The Philosophical Review*, 53, 195–202.
- CICERO (1928) *On the Republic. On the Laws*, C. Keyes (tr.) (Cambridge MA: The Loeb Classical Library of Harvard University Press).
- CICERO (1991) *Cicero on Stoic Good and Evil*, M. R. Wright (tr.) (Warminster: Aris & Phillips Ltd.).
- COLISH, M. (1985) *The Stoic Tradition*, 2 vols (Leiden: E. J. Brill).
- DAVENPORT, J. (2002) 'Liberty of the higher-order will: Frankfurt and Augustine', *Faith and Philosophy: Journal of the Society of Christian Philosophers*, 19, 437–461.
- DIHLE, A. (1982) *The Theory of Will in Classical Antiquity* (Berkeley CA: University of California Press).
- DILLON, J. & GERSON, L. P. (trs) (2004) *Neoplatonic Philosophy: Introductory Readings* (Indianapolis: Hackett Publishing Company).
- EPICETUS (1979) *The Discourses as Reported by Arrian, The Manual, and Fragments*, I, W. A. Oldfather (tr.) (Cambridge MA: The Loeb Classical Library of Harvard University Press).
- EPICETUS (1985) *The Discourses as Reported by Arrian, The Manual, and Fragments*, II (Cambridge MA: The Loeb Classical Library of Harvard University Press, 1985).
- FRANKFURT, H. (1971) 'Free will and the concept of a person', *The Journal of Philosophy*, 68, 5–20.
- FREDE, M. (2002) 'John of Damascus on human action, the will, and human freedom', in K. Lerodiankonou (ed.) *Byzantine Philosophy and its Ancient Sources* (Oxford: Oxford University Press, 2002), 63–95.
- FREDE, M. (2011) *A Free Will: Origins of the Notion in Ancient Thought* (Berkeley: University of California Press).
- GILBERT, N. W. (1963) 'The concept of the will in early Latin philosophy', *Journal of the History of Philosophy*, 1, 17–35.
- GREEN, W. M. & DAUR, K.-D. (eds) (1970) *Sancti Aurelii Augustini: Contra academicos. De beata vita. De ordine. De magistro. De libero arbitrio (Corpus Christianorum Series Latina, Volume XXIX)* (Turnhout: Typographi Brepols).
- HOGAN, D. (2009) 'Noumenal affection', *The Philosophical Review*, 118, 501–532.
- KANT, I. (2002a) *Critique of Practical Reason*, W. Pluhar (tr.) (Indianapolis: Hackett Publishing).
- KANT, I. (2002b) *Groundwork for the Metaphysics of Morals*, A. Wood (tr.) (New Haven: Yale University Press).
- KIRWAN, C. (1989) *Augustine* (London: Routledge).
- MANN, W. E. (1999) 'Inner-life ethics', in G. B. Matthews (ed.) *The Augustinian Tradition* (Berkeley: University of California Press, 1999), 140–165.
- O'CONNELL, R. J. (1970) 'De libero arbitrio I: Stoicism Revisited', *Augustinian Studies*, 1, 49–68.
- O'DALY, G. (1987) *Augustine's Philosophy of Mind* (Berkeley: University of California Press).
- STANG, N. (2013) 'Freedom, knowledge and affection: reply to Hogan', *Kantian Review*, 18, 99–106.
- THOMPSON, S. E. (2012) 'What goodness is: order as imitation of unity in Augustine', *Review of Metaphysics*, 65, 525–553.
- WETZEL, J. (1992) *Augustine and the Limits of Virtue* (Cambridge: Cambridge University Press).

## Notes

1. *De libero arbitrio* (henceforth *lib. arb.*) 1.4.10.28. Translation from Augustine (2010), 216. The original text, translated into English above, runs as follows: *Conuenit enim inter nos omnia malefacta non ob aliud mala*

*esse nisi quod libidine, id est improbanda cupiditate, fiunt.* I draw all Latin text from Green & Daur (1970). Terminological note: I reproduce the polyvalence of Augustine's '*voluntas*' (also, '*volo*') by treating 'desire' and 'will' and their respective cognates as intersubstitutable *salva veritate*, a treatment by and large consistent with contemporary English usage. Abbreviations of Augustine's works follow *Augustinus-Lexikon*.

2. *lib. arb.* 1.4.10.31.
3. Augustine introduces temporality with respect to laws at 1.6.14.48 and later extends it to objects, for example, at *lib. arb.* 1.15.31.105. In Plotinus the nearest parallel to an object's being eternal in Augustine's sense would be its being 'up to us': 'If, then, virtue is, in a way, another intellect and in a way a habit that makes the soul "intellectualized," what is "up to us" will again not come to be in action but in intellect detached from actions' (*Enneads*, VI.8.5.34–35, in Dillon & Gerson (2004), 165). We will see that Augustine's Stoic antecedents are even nearer the mark.
4. Epictetus, *Discourses* III.II.2, translation from Epictetus (1985), 21.
5. Cicero, *Stoic Paradoxes* [*Paradoxes*] II.17, translation from Cicero (1991), 85.
6. *Ibid.*, III.25, p. 89.
7. *Ibid.*, III, p. 86.
8. Augustine adds to the paradox a conception of the will as independent of the faculty of reason and responsible for evil because free to choose it or not. The introduction of this conception was standardly attributed to Augustine – see e.g. Gilbert (1963), 31–32, and, more imposingly, Dihle (1982). Recently, Augustine's priority has been disputed. To cite the most prominent example, Frede traces that conception of free will rather to Epictetus – see Frede (2011), ch. 5, particularly pp. 76–77, and ch. 9; also Frede (2002). For more on Augustine's Stoic antecedents generally, see Colish (1985), II; also Byers (2012).
9. Cf. Dihle (1982), 70ff, where limits to our knowledge of the future are discussed in relation to Hellenistic conceptions of rationalized choice and Biblical voluntarism.
10. *lib. arb.* 1.1.3.9.
11. *Ibid.* 1.5.11.33. The idea is familiar from Cicero: 'It may thus be clear that in the very definition of the term 'law' there inheres the idea and principle of choosing what is just and true', *De Legibus*, Book II, translation from Cicero (1928), 385.
12. *Ibid.*, 1.6.15.48.
13. Cf. *conf.* II.iv.9: 'Theft receives certain punishment by your law, Lord, and by the law written in the hearts of men, which not even iniquity itself destroys', also *conf.* III.viii; translation from Augustine (1991a), 28 and 46, respectively. Unconditionedness on compliance is, in fact, a *condicio sine qua non*. Maybe the clearest acknowledgement of this in *lib. arb.* comes at 2.12.34.133–135, where Augustine treats of 'rules of truth': 'The truth, remaining in itself, neither increases when we see more of it nor decreases when we see less', etc. For general discussion of the normative character of Augustine's hierarchies – including beasts and masters, as discussed in the first book of *lib. arb.* – see Mann (1999), 144. For discussion of antecedents of Augustine's view, see Wetzell (1992), 55–62. See also O'Daly (1987), 183.
14. *Ibid.* 1.6.15.51.
15. *lib. arb.* 2.1.2.3: 'conferring benefits on strangers is a sign of someone's goodness'.
16. *lib. arb.* 1.5.12.34.
17. O'Connell (1970), 64.
18. Willing is not something that a will can fail to do against itself, for the will against which it would fail is already the accomplishment of that at which it would have to fail in order to fail against itself. See *lib. arb.* 1.12.26.86–1.13.29.97; also *ciu.* 5.10. Divided will and contention between multiple wills, both within and between individuals, are different matters altogether, and not under discussion. At the urging of one reader, I note here the striking echo of Augustine's conception of free will in Frankfurt (1971). For more recent discussion, see Davenport (2002).
19. *lib. arb.* 1.6.15.51.
20. Kant (2002b), 55.
21. Kant characterizes the categorical imperative as a 'formal' principle; for a representative instance, see *Critique of Practical Reason, Analytic of Pure Practical Reason*, Theorem IV. When it comes time in *Critique of Practical Reason* to specify 'what form in a maxim is fitting for universal legislation', Kant offers only that it 'can be distinguished without instruction by the commonest understanding' before proceeding to an illustration (translation from Kant (2002a), 40). Kant's examples of failed maxims are principles of action that extinguish themselves when universalized, the most famous being that false promises are permissible when in distress (*Groundwork*, First Section). Kant has much to say about this kind of self-extinction;

- following standard practice, I characterize it broadly as contradiction or inconsistency. I should note here a further point, urged on me in a personal communication by Sally Markowitz. Kant's appeal to respect in a different formulation of the categorical imperative strikes a pointedly *informal* tone. However, I will continue to read Kant's account of moral agency as essentially formal for the simple reason that this hitches his account most directly to the machinery in *Critique of Pure Reason*. Also relevant is that we have no successful derivation of the respect-based formulation of the categorical imperative from the formal one. If we had a valid argument from consistent universalizability to respect, room would have to be made for the notion of respect wherever consistent universalizability took centre stage. But we have no such such argument.
22. Chroust (1944), 199.
  23. An example of its implicit appearance: 'If it neglects to do this it is justly blamed and rightly said to be not a good soul; for it diverges from one that does so act, and as this one is praiseworthy, so it follows that the one who does not [so] act is blameworthy', *De Trinitate* [*trin.*], VIII.2.4, translation from Augustine (1991b), 244 (I follow Hill's chapter and passage numerations). Interestingly, Epictetus states the principle outright: 'Wherein lies the good? - In moral purpose. - Wherein lies evil? - In moral purpose. - Wherein lies that which is neither good nor evil? - In the things that lie outside the domain of moral purpose' (Epictetus, *Discourses* II.XVI.1, in Epictetus (1979), 321). The Stoic commitment to semantic bivalence is well known.
  24. Proof from the contrary of P together with two truisms: if something is blameworthy then it is forbidden [truism]; if something is forbidden, it is avoidable [contrary of P]; and if something is avoidable, one who pursues it is responsible for pursuing it [truism].
  25. Mann (1999), 144. I here assimilate willings to what Mann calls 'mental states'. Kirwan (1989), 76, dissents, citing works that post-date *lib. arb.*
  26. Augustine's thought here recalls the two-fold Stoic distinction between the good and the preferred [*proegmena*] and the bad and the dispreferred [*apoproegmena*] - see *De ciuitate dei* IX.4 for an explicit statement, of course post-dating *lib. arb.* That said, as emphasized to me in personal communication by Jochen Schultheiß, Augustine's *uti-frui* distinction puts temporal things on one side and God on the other, while for the Stoics the *proegmena-apoproegmena* distinction is between two kinds of temporal things. So the parallel is limited.
  27. See e.g. Kant (2202b), 'Doctrine of Virtue', XVII. For a good recent discussion, see Byrd (2015).
  28. These identifications recur throughout Augustine's work. For a particularly apt formulation roughly contemporaneous with *lib. arb.*, see *De moribus ecclesiae catholicae et de moribus Manichaeorum*, II.6.8. The relation between, in particular, existence and order is explicated in Thompson (2012).
  29. This is perhaps the most startling claim Kant advances, as he himself realizes: 'The will is thought as independent of empirical conditions and hence, qua pure will, as determined by the mere form of law, and this determining basis is regarded as the supreme condition of all maxims. The thing is strange enough and has no equal in all the rest of practical cognition. For the a priori thought of a possible universal legislation, a thought which is therefore merely problematic, is commanded unconditionally as a law, without borrowing anything from experience or from any external will' (*Critique*, p. 45). The immunity of ethical principles to counterexample rules out empirical criteria for distinguishing genuine ethical principles from counterfeits, but consistency remains available as a criterion, and Kant embraces it.
  30. E.g. *Contra Iulianum opus imperfectum*, III, 106, where he chides Julian: 'You mention so many ways in which God helps us, that is, "by commanding, blessing, sanctifying, restraining, challenging, and enlightening," but you do not mention: by giving his love, though the apostle John says, *Love comes from God*.' Translation from Saint Augustine (1999), 334.
  31. Kant, *Groundwork*, p. 77.
  32. *Ibid.*, p. 78. Kant comes against the problem again in *Critique*, Book I, Chapter III, 'On the Incentives of Pure Practical Reason'. However, meeting the same limitation to our knowledge, he proposes to 'indicate a priori not the basis on which the moral law intrinsically yields an incentive, but what, insofar as it is an incentive, it brings about (or to put it better, must bring about) in the mind' (p. 95). The nature of noumenal affection is a topic of continuing interest among Kant scholars - for recent discussion, see Hogan (2009), and Stang (2013).
  33. *Sermon 165*, 3, from Saint Augustine (1992), 203.
  34. I am grateful to Jacob Swenson, Robert Chenault, Sally Markowitz, Randall Havas, Jochen Schultheiß, and four anonymous reviewers for comments and criticism, and to Betty Talbert for editorial suggestions. I wish to thank especially Willamette University's Center for Ancient Studies and Archaeology and its director Ortwin Knorr for their generous financial support of my research.