

**The Journal of Ethics**  
**Obligation, Responsibility, and History**  
--Manuscript Draft--

<b>Manuscript Number:</b>	JOET-D-17-00231
<b>Full Title:</b>	Obligation, Responsibility, and History
<b>Article Type:</b>	Manuscript
<b>Keywords:</b>	Ability, Blameworthiness, Desert. Historicism, Obligation incompatibilism, Role reversal cases
<b>Corresponding Author:</b>	I. Haji, Ph.D. University Of Calgary Calgary, Alberta, CANADA
<b>Corresponding Author Secondary Information:</b>	
<b>Corresponding Author's Institution:</b>	University Of Calgary
<b>Corresponding Author's Secondary Institution:</b>	
<b>First Author:</b>	I. Haji, Ph.D.
<b>First Author Secondary Information:</b>	
<b>Order of Authors:</b>	I. Haji, Ph.D.
<b>Order of Authors Secondary Information:</b>	
<b>Funding Information:</b>	
<b>Abstract:</b>	I argue that, each of the following, appropriately clarified to yield a noteworthy thesis, is true. (1) Moral obligation can affect moral responsibility (2) Obligation succumbs to changes in responsibility. (3) Obligation is immune from changes in responsibility.

## Author Details

Name: Ishtiyaque Haji

Email: [ihaji@ucalgary.ca](mailto:ihaji@ucalgary.ca)

**Institution:** Department of Philosophy

University of Calgary

2500 University Drive NW

Calgary, AB T2N 1N4

# Obligation, Responsibility, and History

## Abstract

I argue that, each of the following, appropriately clarified to yield a noteworthy thesis, is true. (1) Moral obligation can affect moral responsibility (2) Obligation succumbs to changes in responsibility. (3) Obligation is immune from changes in responsibility.

**Keywords** Ability. Blameworthiness. Desert. Historicism. Obligation incompatibilism. Role reversal cases.

In this paper, I begin with the uncontentious view that factors beyond our control can effect changes in moral obligation over time or imperil obligation. Its discussion lays the foundation to motivate the following theses, which may, preliminarily, be put in this way: (1) The *obligation impact thesis*: Moral obligation can affect moral responsibility (that is, moral praiseworthiness or blameworthiness). (2) The *responsibility impact thesis*: Obligation succumbs to changes in responsibility. (3) The *obligation immunity thesis*: Obligation is resistant to changes in responsibility. I argue that each of these theses, fittingly elucidated, is true.

## 1. Luck, Obligation, and Alternative Possibilities

It is uncontroversial that various factors, all beyond our control, or, if you want, luck, can influence moral obligation. Some of these factors “change” what is obligatory with the passage of time. Others altogether undermine obligation.

Regarding the former, imagine you promise Fred that you will return a book not later than Friday morning that you borrow from him on Monday. In virtue of promising, you morally ought to return the book by Friday. On Wednesday, you sell the valuable

book because you're short on cash. Prior to selling the book, you ought to return it by Friday, partly, because you *can*, before selling it, keep your promise. After selling it, though, it is not true that you ought to return it by Friday because you're unable to do so. Up to the time of selling the book, you ought to return it later, and as this (remote) obligation is *not* satisfied, a wrong *is* done. But this is consistent with holding that upon selling it, your returning the book is not obligatory for you. Moreover, once you sell it, you acquire new obligations, such as the obligation to apologize, make reparation, or attempt to recover the book. It should be borne in mind that obligations do expire. If you ought to return the book not later than 9:00 a.m., and it is now 9:32 a.m., even if you still have the book *this* obligation has now expired. At 9:32 a.m., you no longer have an obligation to return the book at 9:00, and you don't because as of 9:32 you cannot return the book at 9:00. If obligations can expire after the proper time of discharge (9:00 a.m. in this example), why suppose that they cannot expire before the slated time of discharge if the relevant inability has set in prior to this slated time? It is, then, perfectly in order to conclude that once you've sold the book, and thus the relevant inability sets in, your obligation to return it by Friday expires.<sup>1</sup> If one still has concerns about not having an obligation to return the book because the inability to return it is *self*-imposed, simply alter the case so that the inability is *other*-imposed: Rollo steals the book from you on Wednesday, reads it, and then burns it. It's still true that as of Monday, you ought to return the book on Friday; but as of the time the book is stolen, you don't have this obligation.

---

<sup>1</sup> See, e.g., Zimmerman (1996: 98-100); suppressed (2002: 47-52); Vranas (2007: 175-79).

What is beyond your control may also *undercut* obligation. To take a somewhat circuitous but instructive route to see this, begin with the proposition that obligation, permissibility, and impermissibility require alternative possibilities. For instance, nothing can be impermissible for you unless both you could have done, and you could have refrained from doing, it.

The primary argument for obligation's requiring alternative possibilities or two-way control is straightforward.

### ***Primary Argument***

(1O) If you ought to refrain from doing something, *A*, then it is impermissible for you to do *A*.

(2O) If it is impermissible for you to do something, *A*, then you can do *A*.

(3O) Therefore, if you ought to refrain from doing something, *A*, then you can do *A*.

(4O) If you ought to refrain from doing something, *A*, then you can refrain from doing *A*.

(5O) Therefore, if you ought to refrain from doing something, *A*, then both you can do, and you can refrain from doing, *A*.

(6O) If it is true that if you ought to refrain from doing something, *A*, then both you can do, and you can refrain from doing, *A*, then it is also true that if you ought to do, something, *A*, then both you can do, and you can refrain from doing, *A*.

(7O) Therefore, if you ought to do, something, *A*, then both you can do, and you can refrain from doing, *A*.

(1O) appeals to *Equivalence*: It is obligatory for you to do something, A, if and only if it is impermissible for you to refrain from doing A.

(2O) rests on *Kant's Law/Impermissible*: If it is impermissible for you to do something, A, then you can to A. (In other words, the principle that "*impermissible*" implies "*can*" sustains (2O).)

(4O) just is the second conjunct of *Kant's Law/Obligation*: If you ought to do something, A, then you can do A; and if you ought to refrain from doing A, then you can refrain from doing A.

Obligation, like responsibility, requires freedom. Nothing can be obligatory, permissible, or impermissible for you unless you have obligation-relevant freedom concerning this thing. The principle that "ought" implies "can" and "ought not" implies "can refrain from"—*Kant's Law/Obligation*—captures one element of this freedom. Furthermore, just as obligation requires freedom, so do permissibility and impermissibility. Hence, we should expect the permissibility and impermissibility counterparts of *Kant's Law/Obligation* to be true too. We'll see that this is indeed so.

(6O) is based on considerations of symmetry: Barring special reasons to believe otherwise, obligations to refrain from doing things do not differ in their control requirements from obligations to do things. So, just as there is a dual control requirement for obligations to refrain from doing things, there is such a requirement for obligations to do things.

*Equivalence* is a deontic axiom. But why accept the obligation, permissibility, and impermissibility variants of *Kant's Law*? A compelling reason is that these principles derive support from a powerful analysis of obligation. To elaborate briefly, an enduring

view in moral philosophy is that the concept of obligation is to be explained in terms of the concepts of possibility and goodness. The underlying idea is that something is obligatory if and only if it is the best of the possibilities. Call the value in terms of which possibilities or alternatives are to be ranked *deontic value*. Then we may provisionally assume that the alternative with the highest deontic value is obligatory. This principle may be refined in various ways. Here is an elegant version that Fred Feldman (1986) develops and defends: If (at the time of choice) there is a complete life history—a complete way in which your life can thereafter unfold—in which you perform an action, and no better life history in which you don't perform this action, then it is obligatory for you to perform this action. More rigorously:

*DBWC*: A person, *S*, morally ought, as of *t*, to see to the occurrence of a state of affairs, *p*, if and only if there is a world, *w*, accessible to *S* at *t* in which *S* brings about *p*, and it is not the case that *S* refrains from bringing about *p* in any accessible world as good as or better than *w*.<sup>2</sup>

The axiological issue of just what is deontically best need not detain us. Some may identify what is so best, for example, with what is intrinsically best, others with choice behind a Rawlsian veil of ignorance, and so forth. *DBWC* has it that if there is a world or there are worlds accessible to you in which you do something, and there is no better world or there are no better worlds accessible to you in which you refrain from doing this thing, then you ought to do this thing. More intuitively, and simplifying somewhat as I

---

<sup>2</sup> See, e.g., Feldman (1986): ch. 2. See, also, Zimmerman (1996): ch. 2; suppressed (2016): ch. 2.

customarily will, if *DBWC* is true, at some time, you morally ought to perform an act—an act is *morally obligatory* for you—if and only if you can do it, and it occurs in all the best worlds accessible to you at this time. As of a time, it is morally permissible for you to perform an act if and only if you can do it and it occurs in some of the best worlds accessible to you at this time. And at a time, it is impermissible for you to perform an act if and only if you can do it and it does not occur in any of the best worlds accessible to you at this time. *DBWC* validates both “ought” implies “can” and “permissibility” implies “can”; if *DBWC* is true, these principles are true too. So we have:

*Kant’s Law/Obligation:* If, at  $t$ ,  $S$  morally ought to do  $A$  at  $t^*$ , then, at  $t$ ,  $S$  can do  $A$  at  $t^*$ ; and if at  $t$ ,  $S$  has a moral obligation to refrain from doing  $A$  at  $t^*$ , then, at  $t$ ,  $S$  can refrain from doing  $A$  at  $t^*$ .

*Kant’s Law/Permissible:* If, at  $t$ , it is morally permissible for  $S$  to do  $A$  at  $t^*$ , then, at  $t$ ,  $S$  can do  $A$  at  $t^*$ .

Albeit somewhat indirectly *DBWC*, in conjunction with other considerations, also confirms that “impermissibility,” like “obligation,” implies “can.” Start with this highly plausible principle that is a *DBWC* theorem:

*Impermissible/Obligation Possibility:* If, at  $t$ , it is impermissible for  $S$  to do  $A$  at  $t$ , then, at  $t$ ,  $S$  can do something else, such as refraining from doing  $A$ , which it is obligatory for  $S$  to do at  $t$ .



Assume, now (for *reductio*), that you deny that “impermissible” implies “can.” However, you endorse the view that “ought” and “permissible” each implies “can” perhaps because you accept *DBWC*. In addition, include among these assumptions the additional one that there are no genuine conflicts of obligation (*DBWC* rules out such conflicts). If you have an indirect obligation to do something, you have this obligation in virtue of having an obligation to do something else. You have a direct obligation to do something if and only if this obligation is not indirect. Direct obligation (permissibility and impermissibility) is restricted to intentional actions; not so with indirect obligation.<sup>3</sup>

Consider this case. Unbeknownst to you Ergon, the infallible mind reader, can manipulate you into performing or not performing various actions without your being aware of the manipulation. You have a burning desire to run over Betty, your old foe, but Ergon, Betty’s clandestine protector, sees to it that you will never succeed: If you were to try, you would fail because at the appropriate time, ever watchful Ergon will bring it about that you have a change of heart. He will eradicate any motivation you have to kill Betty. In addition, he will implant in you an irresistible desire to refrain from killing Betty and also pertinent beliefs or other relevant mental elements to ensure that you will not trample Betty. The psychological engineering, aptly timed, suffices for your intentionally whipping around Betty. (You may have an irresistible desire to do something consistent with your intentionally doing that thing.) So, you cannot intentionally kill Betty but you can intentionally refrain from doing so.

---

<sup>3</sup> See, e.g., Zimmerman (2006): 602; suppressed (2012): 30.

Assume, as those who question whether “impermissibility” implies “can” do, that it is impermissible for you to run over innocent Betty. It is, however, false that there is something else you can do that it is obligatory for you to do: Your refraining from killing Betty—something you can do—is not obligatory for you since your refraining from killing Betty is unalterable for you, and anything that is unalterable for you is neither obligatory nor permissible for you. (*That the sun rises tomorrow is unalterable for you, but, surely, it is not obligatory for you to see to the occurrence of this state of affairs.*) Since principle *Impermissible/Obligation Possibility* is violated, conclude that “impermissible” implies “can.”

Modifying the case in this way sidesteps the appeal to unalterability: To safeguard Betty, Ergon reconfigures appropriate parts of your brain at the apt time in such a way that you no longer have any motivation to kill Betty (and you cannot acquire motivation to kill her during the right time interval). Furthermore, the manipulation leaves you with solely two engineered-in mutually exclusive alternatives: you can intentionally avoid killing Betty by veering to the left or to the right of her. Finally, Ergon engineers-in a healthy dose of chance. No matter which of these alternatives you perform, either is indeterministically caused in a way in which your performance is a matter of luck. Whether you veer to the left or right, your avoiding hitting Betty will be intentional but a matter of luck. Once again, you cannot intentionally kill Betty but you can intentionally refrain from killing her. Either of your veering to the left or to the right is not obligatory for you because either is too luck-infected to qualify as an obligation.

*Impermissible/Obligation Possibility* is yet again violated.

To add some detail, conceive of the manipulation in this way: An effort of will is the mental action of an agent's trying to make a particular decision or trying to bring it about that she makes a particular decision (Kane 1999: 231, 233-34). Ergon's manipulation results in your trying to make a decision to veer to the left, that is, in exerting an effort to veer left and, concurrently, trying to make a decision to veer to the right or your exerting an effort to veer right. Indeterminism is located between your effort and your subsequent decision. Given exactly the same past and the laws, if an effort of will to make a particular decision leaves it open that that decision will be made, it seems that whatever decision you make may be intentional but too luck-infected to count as free. Hence, the action or, more generally, the performance that results from this decision, will be too luck-infused to be obligatory.

In Ergon-2, since *Impermissible/Obligation Possibility* is yet again violated, conclude that "impermissible" implies "can." So we've now validated:

*Kant's Law/Impermissible*: If, at  $t$ , it is morally impermissible for  $S$  to do  $A$  at  $t^*$ , then, at  $t$ ,  $S$  can do  $A$  at  $t^*$ .

Next, it is easy to see that impermissibility, like obligation, requires two-way control:<sup>4</sup> If it is impermissible for you to do something, you ought not to do it. If you ought not to do something, you can refrain from doing it. Hence, if it is impermissible for you to do something, you can refrain from doing it (*Impermissibility/Can Refrain*).<sup>5</sup> But it

---

<sup>4</sup> I address permissibility in (suppressed 2016: 27-28).

<sup>5</sup> Alternatively, *Impermissibility/Can Refrain* can easily be derived from *Impermissible/Obligation Possibility*.

is also true that if it is impermissible for you to do something, you can do it (as *Kant's Law/Impermissible* is true). So, if something is impermissible for you, then both you can do, and you can refrain from doing, it.<sup>6</sup>

### *Toward Obligation Incompatibilism*

Digressing slightly, determinism is the thesis that there is, at any instant, exactly one physically possible future (van Inwagen 1983: 3). A possible world is deterministic if the proposition expressing the thesis of determinism is true at this world. It's tempting to argue that if (1O) obligation requires alternatives, and (2O) determinism precludes anyone from having alternatives, then *obligation incompatibilism*—the doctrine that determinism and obligation are incompatible—follows. But were it only this simple to derive obligation incompatibilism from these two premises! A confounding factor is whether the alternatives obligation requires or the “can do otherwise” of obligation, is weak in the sense that you can do otherwise even if determinism is true, or strong in that determinism precludes you from doing otherwise. Regarding the latter, *strong can* is be understood in this way:

*Strong Can:* In possible world,  $w$ , at  $t$ ,  $S$  can do other than  $A$  at  $t^*$  ( $t^*$  may be identical to or later than  $t$ ) =df. There is possible world,  $w^*$ , with the same laws of nature as  $w$ , and the same pre- $t$  past as  $w$  in which, at  $t$ ,  $S$  does something other

---

<sup>6</sup> The following clause may be added to *DBWC* to capture obligation's dual ability requirement.  $S$ , morally ought, at  $t$ , to see to the occurrence of a state of affairs,  $p$ , only if there is a world,  $w^*$ , accessible to  $S$  at  $t$  in which  $S$  refrains from bringing about  $p$ . See suppressed (2012: 61, 2016: 34). See, also, Zimmerman (1996: 26-27); and Hebert (2016: ch. 3).

than  $A$  at  $t^*$  ( $S$  may refrain from doing anything at all or may do something else,  $B$ , instead).

Assuming the “can” of obligation is strong, and determinism precludes anyone from strongly doing otherwise, obligation incompatibilism may well ensue from (1O) and (2O). Here, I merely gesture toward obligation incompatibilism.

Plausibly, if you have an ability to do something,  $A$ , then there is a world *accessible* to you in which you do  $A$ ; you can make this world actual. To build on this account of ability, suitable constraints must be placed on the accessible worlds to capture different sorts of ability (and avoid counterexamples). The constraints will be dictated, partly, by the sort of normative assessment in question. With moral obligation, one primary consideration is that the “can” of obligation expresses specific as opposed to general ability, or in the spectrum from general to specific ability, an ability much closer to the specific end of the spectrum. For example, with no cello in the cab on the way to the concert, Yo-Yo Ma does not have the specific ability, while in the cab, to play the cello, although he has the general ability to do so. Indeed, he will display his marvelous skills in an hour’s time when he opens the performance. For another example and elaboration, purely in virtue of having the requisite capacity, motor skill, and know-how, you have a general ability to ride a bike, even if you’re asleep, intoxicated, and there’s no bike nearby for you to ride. Your ability is increasingly more specific to the extent that your general state correspondingly enables the exercise of your general bike-riding ability, such as when you are adequately awake and sober and your bodily integrity is intact. Likewise, your ability is increasingly more specific to the

degree that prevailing environmental conditions enable the exercise of your general bike-riding ability, such as when there is a well-functioning bike available for use and the road conditions aren't too inhibiting. If you have the specific ability to do something, you have the skills, know-how, psychological and physical capacities to do this thing, and the opportunity to do it; in the circumstances in which you find yourself, you can do this thing.

A second consideration is that your specific ability is reliable. If you don't know the combination of a safe and, amazingly enough, you dial the correct numbers on your first attempt, there is a world accessible to you in which you open the safe. You do have the specific ability to open the safe, but your ability is too tenuous—it's too luck-laden—to qualify as the sort of ability obligation requires.

A third consideration is that if you have a direct obligation to do something, your ability to do it is not just specific and reliable. You must also intentionally be able to do it.

Undoubtedly, there are other constraints to be taken into account when it is the ability requirements of obligation that are sought. Enough, though, has been adumbrated to appreciate the very broad outlines of a framework that aspires to show that the "can" of obligation is strong. Assuming you have a direct obligation, at  $t$ , to do  $A$  at  $t^*$ , conceive of the relevant ability in this way:

*Ability/Obligation:* You can<sub>obligation</sub>, at  $t$ , do  $A$  at  $t^*$  if and only if you have access, at  $t$ , to worlds with the same laws (as the actual world) in which (i) you have the skills, know-how, psychological and physical capacities to do this thing; (ii) you have the opportunity to do  $A$  (roughly, the right environmental conditions prevail);

(iii) you do *A* in a relatively large number of these worlds, perhaps even worlds where it is increasingly difficult for you to do *A*, and (iv) you do *A* intentionally.

A subset of these worlds will be best worlds, and if you morally ought to do *A*, then you do *A* in all of these best worlds. Keep in mind, again, that *Ability/Obligation* is skeletal. Clause (i) is meant to capture specific ability, clause (ii) opportunity, clause (iii), reliable ability, and clause (iv) intentional or robust ability. Needless to say, in a fully worked out version of this view, these clauses will undergo considerable refinement.<sup>7</sup>

We may now understand *Kant's Law* in this way:

*Kant's Law/Obligation\**: If, at *t*, *S* has a moral obligation to do *A* at *t\**, then, at *t*, *S* can<sub>obligation</sub> do *A* at *t\**; and if at *t*, *S* has a moral obligation to refrain from doing *A* at *t\**, then, at *t*, *S* can<sub>obligation</sub> refrain from doing *A* at *t\**.

Other deontic modals have similar ability requirements as well. For instance:

*Kant's Law/Permissible\**: If, at *t*, it is morally permissible for *S* to do *A* at *t\**, then, at *t*, *S* can<sub>obligation</sub> do *A* at *t\**.

*Kant's Law/Impermissible\**: If, at *t*, it is morally impermissible for *S* to do *A* at *t\**, then, at *t*, *S* can<sub>obligation</sub> do *A* at *t\**.

---

<sup>7</sup> See Herbert (2016) for an elegant development of this sort of account of ability.

Now, reconsider this derivation of “ought not” implies “can”:

1. If, at  $t$ ,  $S$  ought not to do  $A$  at  $t^*$ , then, at  $t$ , it is impermissible for  $S$  to do  $A$  at  $t^*$ . [From *Equivalence\**]
2. If, at  $t$ , it is impermissible for  $S$  to do  $A$  at  $t^*$ , then, at  $t$ ,  $S$  can<sub>obligation</sub> do  $A$  at  $t^*$ . [From *Kant's Law/Impermissible\**]
3. Therefore, if, at  $t$ ,  $S$  ought not to do  $A$  at  $t^*$ , then, at  $t$ ,  $S$  can<sub>obligation</sub> do  $A$  at  $t^*$ .
4. If, at  $t$ ,  $S$  ought not to do  $A$  at  $t^*$ , then, at  $t$ ,  $S$  can<sub>obligation</sub> refrain from doing  $A$  at  $t^*$ .  
[From *Kant's Law/Obligation\**]

Therefore, if, at  $t$ ,  $S$  ought not to do  $A$  at  $t^*$ , then, at  $t$ ,  $S$  can<sub>obligation</sub> do  $A$  at  $t^*$  and, at  $t$ ,  $S$  can<sub>obligation</sub> refrain from doing  $A$  at  $t^*$ . Since the control requirements of an obligation to refrain from doing something do not differ from those of an obligation to do something, we may infer:

5. If, at  $t$ ,  $S$  ought to do  $A$  at  $t^*$ , then, at  $t$ ,  $S$  can<sub>obligation</sub> do  $A$  at  $t^*$  and, at  $t$ ,  $S$  can<sub>obligation</sub> refrain from doing  $A$  at  $t^*$ .

In other words, the dual freedom requirements of obligation entail that if you have (at  $t$ ) an obligation to do something,  $A$  (at  $t^*$ ), then in the situation in which you find yourself, *then and there*, given your skills, psychological capacities, and so forth, and the environmental conditions, you can intentionally and reliably, at  $t$ , both do, and refrain from doing,  $A$  at  $t^*$ . This pretty much seems to amount to the claim that the “can” of



obligation is strong. A suitably refined successor of *Ability/Obligation* should, presumably, verify the view that obligation has robust dual specific ability requirements, and, thus, the case for obligation incompatibilism seems strong. The relevance of obligation incompatibilism to assessing whether the three theses with which we opened are true will emerge shortly.

### *Luck, Obligation, and Frankfurt Examples*

Reverting, now, to luck and obligation, a Frankfurt example perspicuously illustrates how factors beyond our control can undermine obligation (Frankfurt 1969). Partition such an example into two stages. In Stage 1 in *Theft*, despite Augustine's being blameworthy for stealing the pears that he did, he could have refrained from stealing them. In Stage 2, a "rerun" of Stage 1, something precludes Augustine from doing anything incompatible with stealing but *without* in any way interfering in Augustine's actually stealing as it turns out. A mind reader, Black, who can tell what Augustine is about to do, will do nothing if he detects some reliable sign Augustine displays that he, Augustine, is about to steal, but will force Augustine to steal if he discerns the reliable sign that Augustine is about to refrain from stealing. But Augustine proceeds exactly as before, so Black has no need to intercede.

Assume that it is impermissible for Augustine to steal the pears in Stage 1. In Stage 2, however, owing entirely to factors beyond his control—roughly, owing to Black's being ever vigilant—it is neither impermissible nor obligatory for Augustine to steal the pears because he cannot refrain from stealing them.

## 2. Obligation's Influence on Responsibility

Thesis one—the *obligation impact thesis*—can be loosely formulated as the thesis that moral obligation may affect either moral praiseworthiness or blameworthiness. It is the thesis that what it is obligatory or impermissible for you to do has, or may have, a direct bearing on whether you are morally praiseworthy or blameworthy for what you do. If this thesis is true, then whatever affects obligation itself—whatever undermines obligation, for instance—will undermine responsibility too. Hence, if factors beyond our control, such as the world’s being deterministic, undermine obligation, then we should expect blameworthiness or praiseworthiness to be adversely affected as well. Exploring this in more detail, I confine discussion primarily to blameworthiness and impermissibility.

Many people have championed the principle that blameworthiness requires impermissibility:

*BRI*: Necessarily, you are blameworthy for doing something only if it is impermissible for you to do it.<sup>8</sup>

It’s easy to see how a change in obligation can influence blameworthiness if *BRI* is true. It is not impermissible for Augustine to steal the pears in Stage 2 in *Theft* (because he could not have refrained from stealing). Proponents of *BRI* are committed to the result that in this stage, unlike in Stage 1, Augustine is not blameworthy for stealing the pears. They may, consequently, reject Frankfurt examples. Although the matter is delicate, I’m inclined to go the other way: Since Augustine *is* blameworthy for stealing the pears despite its not being impermissible for him to steal them in Stage 2, *BRI* ought to be

---

<sup>8</sup> See, e.g., Mellema (1991: 87); Smith (1991: 271); Widerker (1991: 223); Copp (1997, 2003: 286-87); Fischer (2006: 218); Arpaly (2006: 91, n. 3); Campbell (2011: 33-34).

renounced. The possibility of being to blame for something that it is permissibly suboptimal for you to do—for example, for something that it is suberogatory for you to do—also casts serious doubt on *BRI*.<sup>9</sup>

If one holds that Augustine is blameworthy for stealing the pears in Stage 2, just why is he blameworthy if it is not impermissible for him to steal them? If what many take to be the obvious rationale that he did something impermissible is not available, an alternative is called for. In several works, I've attempted to defend the view that blameworthiness is associated with belief in what is impermissible. The core idea is encapsulated in this principle:

*BRBI-1*: Necessarily, you are morally blameworthy for doing something only if you believe that it is morally impermissible for you to do it.<sup>10</sup>

This principle, although not quite right, captures what contrasts it sharply from *BRI*. *BRBI-1*'s intuitive pull is straightforward. Suppose you freely do what you think is impermissible even if it is not in fact impermissible. Since you *willingly* did what from your perspective *is* impermissible for you to do, you are complicit; you are blameworthy for what you did. You were ready to, and did in fact, get your hands soiled. We may assume that as Augustine satisfies *BRBI-1* (or a more convincing variant) in Stage 2, he is blameworthy for stealing the pears even though it was not impermissible for him to steal them.

---

<sup>9</sup> One may also attempt to question *BRI* by counterexample.

<sup>10</sup> Zimmerman also favors a principle of this sort. See, e.g., Zimmerman (1988, 1997, 2014). Note that, as it stands, the possibility of suberogation calls *BRBI-1* into question.

Here, I won't refine *BRBI-1* because it is unnecessary to do so for immediate purposes: providing support for the *obligation impact thesis* that obligation can influence responsibility.<sup>11</sup> *BRBI-1*, just like *BRI*, may be invoked to ratify this thesis. Abe, a fictional agent, has beliefs that are all true. Necessarily, with any proposition, if Abe believes that proposition, that proposition is true. Replacing Augustine with Abe in *Theft*, when Abe steals the pears in Stage 1, he steals them on the basis of the belief that he is doing wrong in stealing them. We may assume that, here, he is blameworthy for stealing them. But when he steals the pears in Stage 2, it's false that he steals them while believing that it is impermissible for him to steal them (as he doesn't have this belief). Since Abe cannot refrain from stealing the pears in Stage 2, it is not impermissible for him to steal them since, again, impermissibility requires avoidability. Abe would then not be blameworthy for stealing the pears in Stage 2 if *BRBI-1* is true. We may conclude that even if *BRBI-1* and not *BRI* is the correct principle that conceptually links blameworthiness and impermissibility, a change in what it is obligatory (or impermissible) for you to do may affect blameworthiness.

Regarding the contentious matter of determinism's jeopardizing obligation, if determinism does indeed imperil obligation, if obligation incompatibilism is true, then responsibility suffers too. On the one hand, if *BRI* is true, then responsibility is undermined. On the other hand, if *BRBI-1* is true, then deontic irrationality pollutes appraisals of blameworthiness: assuming, again, that obligation incompatibilism is true, all these appraisals will be based on the view that blameworthy agents act in light of the *false* belief that they're doing what it is impermissible for them to do.

---

<sup>11</sup> Refinements may be found in (suppressed n.d.)

## 2. Responsibility's Influence on Obligation

What about responsibility's influence on obligation? The thesis to be defended—the *responsibility impact thesis*—may provisionally be stated in this way:

*R-Impact Thesis-1*: What it is morally obligatory for you to do can be affected by whether or not you are responsible for your actions.

Distinguish this thesis from the following not so interesting thesis.

*R-Impact Thesis-2*: What it is morally obligatory for you to do can be affected by whether or not *anyone* is responsible for their actions.<sup>12</sup>

It's not surprising that whether or not others are responsible for their actions can affect how someone morally ought to treat them. Whether or not Rollo is blameworthy for committing some crime may well bear on whether you or someone else is morally obligated to charge him for the crime.

Even so, *R-Impact Thesis-1* may be thought to be obviously true and this may detract from any interest in it. It may be morally obligatory for you to apologize to someone or to make amends depending on whether or not you are to blame for having done something to that person. If you have violated trust for no good moral reason, and you are blameworthy for having done so, you may incur an obligation to set things right. However, the following refinement is more arresting:

---

<sup>12</sup> I thank Erik Wielenberg for pressing me to differentiate these two theses.

*R-Impact Thesis-3*: Assuming that your alternatives are limited to *A*-ing or *B*-ing, and at time, *t*, you can perform either *A* or *B*, whether it is morally obligatory at *t* for you to do *A* or to do *B*, can be affected by whether or not you are (or would be) responsible for performing either of *these* actions in this way: If you are responsible or would be responsible for whichever action, *A* or *B*, you perform (or would perform), at *t* you ought to do *A*. But if you are not responsible (or would not) be responsible for whichever action you perform (or would perform), at *t* you ought to do *B*.

Perhaps many have thought that responsibility cannot influence obligation in the way previously described in which obligation can influence responsibility—changes in obligation altogether preclude blameworthiness (or praiseworthiness, whatever the case is)—or maybe in some other way because the following responsibility correlates of *BR/* and *BRBI-1* are false:

*IRB*: Necessarily, it is impermissible for you to do something only if you are blameworthy for doing it.

*IRBB*: Necessarily, it is impermissible for you to do something only if you believe you are or will be blameworthy for doing it.

Excuses undermine *IRB*. You may do inadvertent wrong and not be blameworthy for doing so. Against *IRBB*, you may do something wrong despite not having any belief whatsoever about whether you are or will be blameworthy for what you're doing.

Or maybe some have thought that responsibility cannot affect obligation because even if a world is devoid of responsibility it need not be shorn of obligation.<sup>13</sup> For example, some have argued that no one can be responsible for anything in a deterministic world because determinism precludes anyone's being the ultimate source of one's actions. But ultimate sourcehood (however precisely this is to be understood), they propose, is a precondition of responsibility (Kane (1996); Smilansky (2000); Pereboom (2001, 2014); Waller (2015)).

However, it *is* possible that changes in responsibility, specifically, the absence or elimination of responsibility, *can* influence obligation. I want to pay particular attention to cases of this sort. At *t*, you have a choice between *A*-ing at *t\** and *B*-ing at *t\**. In the first scenario or possible world, if at *t* you were to do either at *t\**, you would be responsible for doing either at *t\**. In the second scenario or possible world, whichever of the two alternatives, at *t*, you do, you would not be responsible for doing it at *t\** because no one in this world is ever morally responsible for anything. I will argue that in the first world, at *t*, you ought to do *A* at *t\**, and in the second, at *t*, you ought to do *B* at *t\**, and this difference is to be accounted for by the fact that responsibility can affect obligation in virtue of affecting desert.

Revert to this analysis of obligation, modified to include a clause that acknowledges the could-have-done-otherwise requirement of obligation:

---

<sup>13</sup> See, e.g., Pereboom (2001).

*DBWC*: A person, *S*, morally ought, at *t*, to see to the occurrence of a state of affairs, *p*, if and only if there is a world, *w*, accessible to *S* at *t* in which *S* brings about *p*, there is a world, *w\**, accessible to *S* at *t* in which *S* refrains from bringing about *p*, and it is not the case that *S* refrains from bringing about *p* in any accessible world as good as or better than *w*.

Now let's make a substantive assumption that is controversial but plausible: The deontically best is the intrinsically best. So, deontically best worlds are intrinsically best worlds. But what determines the intrinsic value of a world? For purposes of exposing how responsibility may affect obligation, I start with some crucial elements of Feldman's neo-hedonistic axiology. Roughly, Feldman recommends that the intrinsic value of a world is a function of both the amount of pleasure and pain the world contains and the distribution of pleasures and pains in accordance with whether the subjects of these pleasures and pains deserve to receive them (Feldman (1997, 2004)).<sup>14</sup> To elaborate briefly, the intrinsic value of a world is the sum of the intrinsic values of the basic intrinsic value states true at that world. Assume that the fundamental bearers of intrinsic value are episodes of pleasure and pain, all relevantly like the following states of affairs:

*B1*: At noon on Tuesday, October 16, 2001, Bob feels pleasure of intensity +8 in the fact that Bob's beer is frosty cold (Feldman 2004: 176).

---

<sup>14</sup> Compare this view with Ross' two-worlds objection (Ross 1930: 138). For simplicity, I assume that the pleasures and pains are sensory pleasures and pains. Nothing of substance would change if these the pleasures and pains were attitudinal.



*B2*: At noon on Tuesday, October 16, 2001, Ned feels pain of intensity -8 in the fact that Ned's beer is warm.

To determine the intrinsic value of a world, for each episode of pleasure (or pain) such as *B1*, adjust its value to reflect the extent to which the subject—the person who enjoys the pleasure or suffers the pain—deserves to be enjoying or suffering it (Feldman 2004: 193). Feldman proposes a number of principles for such adjustment. One of their primary imports is that the value of an episode of pleasure is adjusted either upward if the recipient deserves to enjoy the pleasure and enjoys it, or downward if the recipient deserves not to enjoy it and enjoys it.

Elaborating, a person has negative desert if she deserves pain, positive desert if she deserves pleasure, and neutral desert if she neither deserves pleasure nor pain. Feldman advances the following principles regarding the value adjustment of episodes of pleasure and pain.

P1: Positive desert enhances the intrinsic goodness of pleasure (Feldman 1997: 163).

If you receive the pleasure that you deserve, the value of that episode of pleasure is enhanced. More fully, if you receive the pleasure that you deserve, the value of this episode of pleasure *for the world* is enhanced. Feldman proposes that if a person fully deserves a certain good and receives that good, then the value of the *world* is

increased. The value for the world of his deserved receipt, because of the perfect match between receipt level and desert level, is greater than the value received by the person, taken just by itself. The other principles (P2 – P6) are also to be interpreted as saying something about the value of episodes of pleasure or pain for worlds.

P2: Negative desert base mitigates the intrinsic goodness of pleasure (Feldman 1997: 164).<sup>15</sup>

If you deserve pain and get pleasure instead, the value of the pleasure received is diminished.

P3: Neutral desert neither enhances nor mitigates the intrinsic goodness of pleasure (Feldman 1997: 166).

The intrinsic value of an episode of pleasure of this sort is directly proportional to the amount of pleasure it contains.

P4: Positive desert aggravates the intrinsic badness of pain (Feldman 1997: 166).

---

<sup>15</sup> Many agree it is welfare that people fundamentally deserve to receive, and that lack of fit between welfare received and welfare deserved lowers intrinsic value. See, e.g., Kagan (1997, 2003); and Skow (2012).

If you deserve pleasure but get pain instead, the badness of that pain is made yet worse.

P5: Negative desert base mitigates the intrinsic value of pain (Feldman 1997: 167).

If you deserve the pain that you receive, the value of that episode of pain is mitigated.

P6: Neutral desert neither enhances nor mitigates the evil of pain.

In sum, according to Feldman's desert-sensitive hedonistic axiology the intrinsic value of a world depends on the fit between the pleasures and pains people in this world deserve to receive and the pleasure and pains they receive, the better the fit the higher the intrinsic value of the world.<sup>16</sup>

Assuming that the desert is pleasure or pain, there are many but certainly not all cases in which the deserving person deserves in virtue of some action she performed. In such cases, the person is typically morally responsible for the action (although in other such cases, she may not be). Imagine that in *Theft* Augustine deserves some pain on the basis of stealing the pears. We would not think that he deserves the pain if he were not responsible—specifically, not blameworthy—for stealing the pears. If he stole solely owing to Black's manipulating him to steal, presumably he would not be blameworthy for stealing.

---

<sup>16</sup> See Carlson (1997); Persson (1997); and Arrhenius (2003) for critical discussion of these axiological principles.

In Feldman's axiology, the desert or the currency of desert, the things the deserver is said to deserve, are pleasures (the "positive" deserts) and pains (the negative counterparts).<sup>17</sup> Desert bases—the features in virtue of which the deserver deserves the desert, perhaps, at times, from a distributor of desert—can be many and varied. They include moral worthiness or, more generally, morally good behavior or morally bad behavior, excessive or deficient past receipt, legitimate claims or entitlements, having worked hard, having invested much effort or time in a project, being an entity of a certain sort (for example, being a rational agent), having inflicted harm on innocent others, and so forth.<sup>18</sup> Some of these bases presuppose responsibility in the following way. If agents were not morally responsible, perhaps never at all, or not in specific circumstances, then the putative desert base would not be a desert base. Suppose you were not responsible for writing the terrific book in philosophy, or investing a great deal of effort and time in nurturing the beautiful garden. You would not deserve praise for producing either the book or garden. If no agent is ever morally responsible for anything, then it's difficult to see how hard work could be a desert base. In contrast, suppose in virtue of being a rational agent or a person, you deserve respect. Here, the desert base seems unhitched from responsibility.

We may now see how responsibility can affect obligation as a result of its possible effect on desert. Suppose, holding all other relevant factors constant, as of a certain time, whether it is obligatory for Augustine to donate to a particular charity turns on whether *he* deserves pleasure in virtue of donating some money to that charity.

---

<sup>17</sup> Feldman (2016: 39-39) proposes that some deserts may not be a plus or a minus, a benefit or a burden.

<sup>18</sup> On desert bases, see, e.g., Rescher (1966: 73-83); Feinberg (1970); McLeod (1996); and Feldman (2016: 41-43).

Assume: (1) He would not deserve to receive any pleasure from donating if he were not responsible for donating. (2) His only options, at  $t$ , are to donate to charity C1 at  $t$  or to charity C2 at  $t$ . (3) If, at  $t$ , he were to donate to C1 at  $t$ , he would deserve 10 units of pleasure and receive exactly 10 units. The intrinsic value of the world in which he donates to C1 at  $t$  would then be, say, 115, given principle P1. If he were not responsible for donating, he would still receive 10 units of pleasure that he did not deserve to receive. In this case, the intrinsic value of the world in which he donates to C1 at  $t$  would be 110 (given principle P3). (4) If, at  $t$ , he were to donate to C2 at  $t$ , an unworthy charity, he would deserve 6 units of pain but receive 17 units of pleasure instead. The intrinsic value of the world in which he donates to C2 at  $t$  would then be 112 (given principle P2). If he were not responsible for donating, he would still receive 17 units of pleasure that he did not deserve to receive. In this case, the intrinsic value of the world in which he donates to C2 at  $t$  would be 117. Given these assumptions and the further assumption that he would be responsible for donating to either of these charities, we may suppose that, at  $t$ , he morally ought to donate to C1 at  $t$ .

Next, assume that no one is ever morally responsible for anything. Maybe determinism is true, and determinism undermines responsibility. Then Augustine is not responsible for any of his choices or actions, and because of this, the world in which, at  $t$ , he donates to C2 at  $t$  is intrinsically better than the one in which he donates to C1 at  $t$ . So this time around, at  $t$ , he ought to donate to C2 at  $t$ . We have a case, call it *Central*, in which what is or is not obligatory for Augustine essentially depends on whether or not agents are morally responsible for their actions (or omissions).

To summarize, in *Central*, Augustine can donate to one of two charities. On the supposition that Augustine is responsible for whichever action he performs, he ought to donate to the first charity. But on the supposition that Augustine is not responsible for whichever action he performs, he ought to donate to the second charity. That is so in virtue of fact that if Augustine is responsible for whichever action he performs, Augustine's desert and receipt levels more closely match one another if he donates to the first charity, whereas if Augustine is not responsible, then his desert and receipt levels more closely match one another if he donates to the second charity.

Even if one is hesitant about accepting Feldman's unconventional hedonistic axiology, it is reasonable to suppose that considerations of desert contribute to the intrinsic value of a world. There are very many deserts, very many things that the subject of desert—the deserver—may deserve. For example, people can deserve blame, compassion, sympathy, financial compensation, wages, commendation, prizes, respect, rights, contempt, penalties, love, hate, and so forth. As already noted, there are, too, very many desert bases, those features in virtue of which the deserver deserves the desert, sometimes from a distributor of deserts, including being morally virtuous, being morally vicious, being vulnerable to attack from invaders, being creative, being lonely, being hard working in business or academics, being the winner of a race, and so forth.

I can imagine some philosopher defending the view that the deontic value of a world is a function of the fit between the primary intrinsic goods and evils people in this world deserve to receive and the primary goods and evils they receive, the better the fit the higher the intrinsic value of the world. This time around, though, the primary intrinsic

goods would include pleasure and *other* goods, and the evils would include pain and *other* evils. It is this axiology that would then be wedded to the principle, roughly, that you ought to do an act if and only if it occurs in some world accessible to you and it does not fail to occur in any world accessible to you that is deontically better.

Next, simply modify Augustine's case so that pretty much everything stays about the same with the following exception: Suppose, holding all other relevant factors constant, as of a certain time, whether it is obligatory for Augustine to donate to a particular charity turns on whether he deserves various goods (in addition to pleasure) in virtue of donating some money to that charity. With the assumption that he would not deserve to receive any of these goods from donating if he were not responsible for donating, we would once again find that what it is morally obligatory for Augustine to do can be affected by whether or not Augustine is responsible for his actions.

Some have asked what life would be like if no one were ever morally responsible for anything. Derk Pereboom has argued that a conception of life without the freedom responsibility presupposes would not be devastating to morality or to our sense of meaning in life, and in certain respects it may even be beneficial (1995, 2001, 2014).<sup>19</sup> *Central* uncovers a cost of being deprived of responsibility, assuming, as Pereboom seems to, that there are worlds devoid of responsibility but not obligation, such as, perhaps, the actual world. We saw that the intrinsic value of the world in which Augustine donates to charity C1 at  $t$ , assuming he would be responsible for donating to C1, is 115. If he were not responsible for donating, the intrinsic value of the world in which he donates to C1 at  $t$  would be 110. Augustine could find himself in many other

---

<sup>19</sup> See, also, Waller (2015).

situations relevantly analogous to this one: He could make actual or bring about *better* worlds if he were responsible for his germane actions than if he were not. Being without the freedom responsibility requires deprives one of the opportunity to bring about better worlds in the fashion in which Augustine could bring about a better world if he were responsible.

In addition, not having the freedom responsibility requires undermines desert bases that presuppose desert (in the fashion previously explained). And this, of course, could affect what it is obligatory or impermissible for one to do provided it can be true that a world without the freedom moral responsibility requires need not preclude obligation.

### 3. Obligation, Responsibility, and History

Finally, I turn to the *obligation immunity thesis*, in coarse outline, the thesis that obligation is immune to changes in responsibility. What I have in mind can more fully (but still opaquely) be expressed in this way. Whereas responsibility is historical, obligation is not, and changes in responsibility appraisals, owing to changes in history, need have no impact on obligation appraisals. Demystification of these claims about obligation, responsibility, and history is the burden of this section.

Theorists about moral responsibility disagree with one another over whether how agents acquire their springs of action, such as desires, beliefs, and values, bears on the correct analysis of moral responsibility for actions. Historicists (or externalists) believe that history matters. They are of the mind that, for instance, various sorts of manipulation undermine responsibility because targeted actions of the manipulated agent causally derive from surreptitiously implanted actional springs. In contrast, anti-



historicists (or internalists or structuralists) believe that history is of little or no concern.

Frankfurt endorses anti-historicism as these passages reveal:

To the extent that a person identifies himself with the springs of his actions, he takes responsibility for those actions and acquires moral responsibility for them; moreover, the questions of how the actions and his identifications with their springs are caused are irrelevant to the questions of whether he performs the actions freely or is morally responsible for performing them. (Frankfurt 1988: 54)

If someone does something because he wants to do it, and if he has no reservations about that desire but is wholeheartedly behind it, then – so far as his moral responsibility for doing it is concerned – it really does not matter how he got that way. One further requirement must be added...: the person's desires and attitudes have to be relatively well integrated into his general psychic condition. Otherwise they are not genuinely his.... As long as their interrelations imply that they are unequivocally attributable to him... it makes no difference – so far as evaluating his moral responsibility is concerned – how he came to have them. (Frankfurt 2002: 27)

Mele, an historicist, questions these quoted claims with his radical reversal cases:

SWEET BETH. Beth is one of the kindest, gentlest people on Earth. She was not always that way, however. When she was a teenager, she came to view herself, with some justification, as self-centered, petty, and somewhat cruel. She worked hard to improve her character, and she succeeded.... She is, moreover, a person who is so [kind, generous and] gentle that even the thought of killing a pesky fly disturbs her. Her system of values plays a major role in generating her admirable behavior, of course. (Mele 2016: 72)

Whereas Beth is angelic, Chuck in THOROUGHLY BAD CHUCK is evil:

[Chuck]...enjoys killing people, and he is wholeheartedly behind his murderous desires, which are “well integrated into his general psychic condition.” When he was much younger, Chuck enjoyed torturing animals, but he was not wholeheartedly behind this. These activities sometimes caused him to feel guilty, he experienced bouts of squeamishness, and he occasionally considered abandoning animal torture. However, Chuck valued being the sort of person who does as he pleases and who unambivalently rejects conventional morality as a system designed for and by weaklings. He intentionally [and freely] set out to ensure that he would be wholeheartedly behind his torturing of animals and related activities, including his merciless bullying of vulnerable people [and he was morally responsible for so doing]. One strand of his strategy was to perform cruel actions with increased frequency in order to harden himself against feelings of guilt and squeamishness and eventually to extinguish the source of those feelings. His strategy worked. (Mele 2006: 171)<sup>20</sup>

Beth (from SWEET BETH) is reintroduced in ONE BAD DAY:

When [Beth] crawled into bed last night, she was a sweet person, as she always had been. But she awoke with a desire to stalk and kill a neighbor, George. Although she had always found George unpleasant, she is very surprised by this desire. What happened is that, while Beth slept, a team of psychologists that had discovered the system of values that make Chuck tick implanted those values in Beth after erasing hers. They did this while leaving her memory intact, which helps account for her surprise. Beth reflects on her new desire. Among other things, she judges, rightly, that it is utterly in line with her system of values. She also judges that she finally sees the light about morality—that it is a system designed for weaklings by weaklings. Upon reflection, Beth has no reservations about her desire to kill George and is wholeheartedly behind it. Furthermore, the desire is “well integrated into [her] general psychic condition” at the time. Seeing absolutely no reason not to stalk and kill George, provided that she

---

<sup>20</sup> See, also, Mele (2016: 73).

can get away with it, Beth devises a plan for killing him, and she executes it—and him—that afternoon. That she sees no reason not to do this is utterly predictable, given the content of the values that ultimately ground her reflection. (Mele 2006: 171–72)<sup>21</sup>

Mele judges that whereas Chuck may well be morally responsible for his killings, he “cannot help but see Beth as too much a victim of her manipulators to be morally responsible for killing George” (2006: 172).<sup>22</sup> I agree with this judgment.<sup>23</sup>

In ONE BAD DAY MODIFIED, “full-strength Chuck-like values were indeterministically produced in Beth, and those values were involved in the indeterministic causation of Beth’s intentionally killing George” (Mele 2016: 76). There is a very small chance, because of the indeterminism, that in this modified version of ONE BAD DAY, Beth refrains from killing George. Mele proposes that most who judge that Beth is not responsible for the killing in the original case will make the same judgment in the modified case.

These reversal cases, however, do little, if anything to dislodge the judgment that it is morally impermissible for Beth to kill George in ONE BAD DAY MODIFIED. In all the best worlds accessible to her at the apt time, she refrains from killing George. Hence, such cases would not reinforce historicism concerning obligation, *O-historicism*, although I agree that they shore up historicism concerning responsibility, *R-historicism*.

Those interested in the freedom that obligation requires may be inclined to defend something like a Frankfurtian hierarchical account of this freedom (Frankfurt 1971), supplemented with the condition that the agent was able (in some defensible

---

<sup>21</sup> See, also, Mele (2016: 73-74).

<sup>22</sup> See, also, Mele (2016: 75-76).

<sup>23</sup> See, e.g., suppressed (1998: 115-22, 2009: 61-67, 2013); suppressed (2008: 15-41).

compatibilist sense of “able”) at the time to do otherwise then. It will not be a shortcoming of an account of this sort that it runs afoul of cases such as ONE BAD DAY MODIFIED: These cases don’t undermine obligation because obligation, unlike responsibility, is history-insensitive.

In addition, if obligation *were* history-sensitive and principle *BRI*—you are blameworthy for an action only if it is impermissible for you to perform it—were true, one could not be an anti-historicist about responsibility. There is, however, no prima facie difficulty in holding that whereas responsibility is history-sensitive, obligation is not.

In original design cases, a designer determines exactly how an agent’s life will unfold before the agent comes into existence. Here is Mele’s intriguing case, featuring such design, which may incline some toward responsibility incompatibilism (the doctrine that responsibility and determinism are incompatible):

Diana creates a zygote *Z* in Mary. She combines *Z*’s atoms as she does because she wants a certain event *E* to occur thirty years later. From her knowledge of the state of the universe just prior to her creating *Z* and the laws of nature of her deterministic universe, she deduces that a zygote with precisely *Z*’s constitution located in Mary will develop into an ideally self-controlled agent who, in thirty years, will judge, on the basis of rational deliberation, that it is best to *A* and will *A* on the basis of that judgment, thereby bringing about *E*. If this agent, Ernie, has any unsheddable values at the time, they play no role in his motivating his *A*-ing. Thirty years later, Ernie is a mentally healthy, ideally self-controlled person who regularly exercises his powers of self-control and has no relevant compelled or coercively produced attitudes. Furthermore, his beliefs are conducive to informed deliberation about all matters that concern him, and he is a reliable deliberator. (Mele 2006: 188)

The zygote story motivates this argument against responsibility compatibilism (the denial of responsibility incompatibilism) (Mele 2006: 189):

1. Ernie is not a free agent and is not morally responsible for anything in his deterministic universe because of the way his zygote was produced (by original design).
2. Concerning free action and moral responsibility of the beings into whom zygotes develop, there is no significant difference between the way Ernie's zygote comes to exist and the way any normal human zygote comes to exist in a deterministic universe.
3. If (1) and (2), then determinism precludes free action and moral responsibility.
4. Therefore, determinism precludes free action and moral responsibility.

Even if one accepts the first premise, which responsibility compatibilists should be inclined to question, one would be hard pressed to accede to its revision that addresses obligation: No event is morally obligatory for Ernie in his deterministic universe *because* of the way his zygote was produced. Indeed, such original design cases fail to provide support for the judgment that nothing can be obligatory for designed agents like Ernie. An appropriately modified version of the zygote argument would, thus, be an ineffective argument against obligation compatibilism (the denial of obligation incompatibilism) even though some may take the original argument to exert pressure against responsibility compatibilism.

Defending responsibility compatibilism is *the* prominent strategy to evade the deterministic horn of the famed responsibility dilemma: either determinism or its falsity is incompatible with responsibility. But responsibility-skeptics or responsibility-libertarians

may invoke arguments featuring reversal cases or original design arguments (such as the zygote argument) to undermine responsibility compatibilism.<sup>24</sup> These two sorts of argument make it harder to defend responsibility compatibilism; at least the arguments must be tackled head-on if the deterministic horn is to be evaded. But these kinds of argument will not stand in the way of developing and defending some version of obligation compatibilism. Not having to worry about these sorts of argument makes it relatively easier to defend obligation compatibilism. The requirement of alternative possibilities for obligation, however, makes it relatively harder to defend obligation compatibilism because obligation compatibilists incur the burden of developing a defensible compatibilist account of “can.”

In sum, the radical reversal and original design cases provide strong intuitive support for the view that obligation, unlike responsibility, is not sensitive to history; O-historicism is untenable. But why precisely is obligation anti-historical?

First, notably, not one of the major competing normative ethical theories that seek to specify necessary and sufficient conditions for obligation lend any credibility to obligation’s being historical. For example, on simple hedonistic utilitarian views, an act is right if and only none of its alternatives has a higher utility than it has (the utility of an action is the result of subtracting the total amount of pain it would produce if performed from the total amount of pleasure it would produce if performed). In *ONE BAD DAY MODIFIED*, nothing tells against the presumption that an alternative to Beth’s killing George maximizes utility. On Ross’ view, an act is right if and only if none of its alternatives is a more stringent prima facie duty than it is. Again, we may assume that

---

<sup>24</sup> Responsibility libertarianism is the conjunction of responsibility incompatibilism and the view that, at times, persons perform free actions for which they are morally responsible.

an alternative to Beth's killing George is the most stringent of her prima facie duties in her circumstances. One version of the categorical imperative says that an act is right if and only if its agent can consistently will that its maxim be a universal law of nature. It would seem that Beth could not consistently will that the appropriate maxim of her killing George in ONE BAD DAY MODIFIED be a universal law of nature. A Rawlsian view has it that an act is right if and only if it is performed in its society, in some reasonable sense of "its society," there is some moral code that members of this society would agree to abide by in the original position, and the code permits this act. Again, nothing about this theory supports obligation's being historical. Presumably, if Beth would abide by the prescriptions of the preferred code of her society, she would not kill George.<sup>25</sup>

In addition, *DBWC*—an *analysis* of obligation—does not favor obligation's being historical either. In reversal cases, such as Beth's, one of Beth's engineered-in options may have the property of *being such that it occurs in all the best worlds accessible to her* at the pertinent time (and it does not occur in some worlds accessible to her at this time). If there were such an option, it would be obligatory for her. Furthermore, if obligation were historical, we would be saddled with the untenable result that the worlds accessible to an agent at some time that turn out to be the *best* worlds accessible to her at that time is partly a function of, for instance, the sort of manipulation found in reversal cases: If Beth does not fall victim to the annoying psychologists, she does not kill George in all the best worlds accessible to her at the germane time; but if she is

---

<sup>25</sup> Each of these theories can be better reformulated in the form: at  $t$ , it is permissible for  $S$  to do  $A$  at  $t^*$  (roughly) if and only if  $S$  does  $A$  at  $t^*$  in at least some of the best worlds accessible to  $S$  at  $t$ . I won't, however, worry about the reformulations here.

manipulated, and assuming that apt manipulation affects what it is obligatory for her to do, she kills George in all these worlds.

A second reason recommending obligation's being anti-historical has to do with appraisals of obligation being act- rather than agent-focused. With responsibility or agent-focused appraisals, the agent is of paramount significance; her springs of action must be "unequivocally attributable" (Frankfurt 2002: 27) to her as she is the primary object of evaluation. This is not so with *act*-focused assessments of obligation. Imagine it is imperative that a button in a hotel suite in New York be pressed to save millions of lives in Asia; it is obligatory at  $t$  for someone that she presses the button at  $t$ , no matter who this person is and no matter what her psychology—it could be Michael, or Bob, or Al, or Dana however psychologically different they may be. Roughly, the idea is that evaluations of obligation pay little or no heed to the sort of person you are, whereas those of responsibility do.

Approaching this point slightly differently, assessments of obligation are, in a certain respect, impersonal. In many mundane cases, the psychological profile of the agent matters little with judgments of obligation despite obligation's being agent- and time- relativized. Sweet Beth is by the hotel suite with the button that she can easily press in time. In all the best worlds accessible to her at the appropriate time, she presses the button at this time. Several others are like Beth in this respect too. If any one of them were to replace her, it would, at the apt time, be obligatory for this person to press the button at this time. But then it would be puzzling why history matters to the truth of judgments of obligation despite its bearing on the truth of judgments of responsibility. Suppose, just prior to pressing the button, the manipulators work on Beth



and turn her into a psychological twin of Chuck (they work very fast!) She awakes with no memories of being tinkered with. It is still obligatory for her to press the button even though she will now not press it. The switch in psychological profile makes no difference to what it is obligatory for her to do just as a switch with someone else would make no difference to what it would be obligatory for this substitute to do: she should press the button.

To conclude, of the three theses discussed, the *impact obligation thesis* is least disputable. If obligation incompatibilism is true (obviously, something disputable), then provided obligation presupposes our having strong alternatives, responsibility will be impacted if either *BRI* or its belief competitor, *BBRI-1*, is true. The *responsibility impact thesis* (*R-Impact Thesis-3*), among other things, casts doubt on the view that the elimination of responsibility need have no interesting bearing on obligation. Finally, the *obligation immunity thesis* brings to light one significant difference between obligation and responsibility: whereas the latter is historical, the former is not.<sup>26</sup>

**Acknowledgments** I'm most grateful to Ryan Hebert for his astute comments on an earlier draft. This paper was written during my tenure of a 2017-2021 Social Sciences and Humanities (SSHRC) grant. Many thanks to this granting agency for its support.

## References

Arpaly, Nomy. 2006. *Merit, meaning, and human bondage: An essay on free Will*. Princeton, NJ: Princeton University Press.

---

<sup>26</sup> Another difference is that whereas the latter may well presuppose that agents be the ultimate sources or originators of their actions, the former does not. See suppressed (2017).

- Arrhenius, Gustaf. 2003. Feldman's desert-adjusted utilitarianism and population ethics. *Utilitas* 15: 225-36.
- Feldman, Fred. 1986. *Doing the best we can*. Dordrecht: D. Reidel Publishing Company.
- Campbell, J. K. 2011. *Free will*. Oxford: Polity Press.
- Carlson, Erik. 1997. Consequentialism, distribution and desert. *Utilitas* 9: 207–216.
- Copp, David. 1997. Defending the principle of alternate possibilities: blameworthiness and moral responsibility. *Nous* 31: 441–56.
- Copp, David. 2003. 'Ought' implies 'can', blameworthiness, and the principle of alternative possibilities. In *Moral responsibility and alternative possibilities: Essays on the importance of alternative possibilities*, ed. David Widerker and Michael McKenna, 265–99. Aldershot UK: Ashgate Press.
- Feinberg, Joel. 1970. Doing and deserving. In *Doing and deserving: Essays in the theory of responsibility*, 55-94. Princeton, NJ: Princeton University Press.
- Feldman, Fred. 1986. *Doing the best we can*. Dordrecht: D. Reidel Publishing Company
- Feldman, Fred. 1997. *Utilitarianism, hedonism, and desert*. Cambridge: Cambridge University Press.
- Feldman, Fred. 2004. *Pleasure and the good life. Concerning the nature, varieties, and plausibility of hedonism*. Oxford: Clarendon Press.
- Feldman, Fred. 2016. *Distributive justice: Getting what we deserve from our country*. New York: Oxford University Press.

- Fischer, John M. 2006. *My way: Essays on moral responsibility*. New York: Oxford University Press.
- Frankfurt, Harry. 1969. Alternate possibilities and moral responsibility. *The Journal of Philosophy* 66: 829–39.
- Frankfurt, Harry. 1971. Freedom of the will and the concept of a person. *The Journal of Philosophy* 68: 5–20.
- Frankfurt, Harry. 1988. *The importance of what we care about*. Cambridge: Cambridge University Press.
- Frankfurt, Harry. 2002. Reply to John Martin Fischer. In *Contours of agency. Essays on themes from Harry Frankfurt* ed. S. Buss and L. Overton, 27-31. Cambridge, Mass.: The MIT Press.
- Hebert, Ryan. 2016. *Teleologism full stop: A general theory of ability, agency, know how, obligation, and justification*. PhD diss.: University of Calgary.
- Kagan, Shelly. 1997. Equality and desert. In *What Do We Deserve*, ed. Louis Pojman and Owen McLeod, 298–314. New York: Oxford University Press.
- Kagan, Shelly. 2003. Comparative desert. In *Desert and Justice*, ed. Serena Olsaretti, 93–122. New York: Oxford University Press.
- Kane, Robert. 1996. *The significance of free will*. New York: Oxford University Press.
- Kane, Robert. 1999. Responsibility, luck, and chance: Reflections on free will and indeterminism. *The Journal of Philosophy* 96: 217–40.
- McLeod, Owen. 1996. Desert and wages. *Utilitas* 8: 205-21.
- Mele, Alfred. 2006. *Free will and luck*. New York: Oxford University Press.

- Mele, Alfred. 2016. Moral responsibility: Radical reversals and original design. *The Journal of Ethics* 20: 69-82.
- Mellema, Gregory. 1991. *Beyond the call of duty: Supererogation, obligation, and offence*. Albany: State University of New York Press.
- Pereboom, Derk. 1995. Determinism al dente. *Nous* 29: 21–45.
- Pereboom, Derk. 2001. *Living without free will*. Cambridge: Cambridge University Press.
- Pereboom, Derk. 2014. *Free will, agency, and meaning in life*. New York: Oxford University Press.
- Persson, Ingmar. 1997. Ambiguities in Feldman's desert-adjusted values. *Utilitas* 9: 319-27.
- Rescher, Nicholas. 1966. *Distributive justice: A constructive critique of the utilitarian theory of distribution*. New York: Bobbs-Merrill Co. Inc.
- Ross, William David. 1930. *The right and the good*. New York: Oxford University Press.
- Skow, Bradford. 2012. How to adjust utility for desert. *The Australasian Journal of Philosophy* 90: 235-57.
- Smilansky, Saul. 2000. *Free will and illusion*. Oxford: Clarendon Press.
- Smith, Holly. 1991. Varieties of moral worth and moral credit. *Ethics* 101: 279–303.
- Van Inwagen, Peter. 1983. *An essay on free will*. Oxford: Clarendon Press.
- Vranas, Peter B. M. 2007. I ought, therefore I can. *Philosophical Studies* 136: 167-216.
- Waller, Bruce. 2015. *The stubborn system of moral responsibility*. Cambridge, MA: MIT Press.
- Widerker, David. 1991. Frankfurt on 'ought implies can' and alternative possibilities. *Analysis* 51: 222–24.

Zimmerman, Michael J. 1988. *An essay on moral responsibility*. Totowa, NJ: Rowman & Littlefield.

Zimmerman, Michael J. 1996. *The concept of moral obligation*. Cambridge: Cambridge University Press.

Zimmerman, Michael J. 1997. A plea for accuses. *American Philosophical Quarterly* 34: 229–43.

Zimmerman, Michael J. 2006. Moral luck: A partial map. *Canadian Journal of Philosophy* 36: 585–608.

Zimmerman, Michael J. 2014. *Ignorance and moral obligation*. Oxford: Oxford University Press.