dialectica

Vol 76 : 4
December 2022

Responsibility First: How to Resist Agnosticism about Moral Responsibility

László Bernáth & Tamás Paár

doi:10.48106/dial.v76.i4.01

László Bernáth & Tamás Paár. 2022. "Responsibility First: How to Resist Agnosticism about Moral Responsibility." *Dialectica* 76(4): 1–1. doi:10.48106/dial.v76.i4.01.

Responsibility First: How to Resist Agnosticism about Moral Responsibility

László Bernáth & Tamás Paár

We argue against the view that one should suspend belief in the existence of moral responsibility. We start with a simple argument based on the claim that the existence of obligations entails the existence of moral responsibility. If this is true, then agnosticism about moral responsibility is incoherent. However, this simple argument is insufficient. It can be repaired by focusing on agents who rationally believe in a particular conception of obligation (the "Responsibility First View" (RFV)). On that conception, non-moral obligations that are not appropriately related to moral obligations can be freely ignored and the property of being morally responsible is identical to the property of fulfilling all necessary conditions for bearing moral obligations. Those agents who rationally hold RFV can still rationally believe in moral responsibility even if they lack direct evidence for the existence of moral responsibility.

Even if we lack evidence for the existence of moral responsibility or if scientific research makes it unlikely that moral responsibility is real, one can still rationally maintain belief in it as long as one adopts a specific view of moral obligations (the "Responsibility First View"). Or so we will argue.

We first outline the case for agnosticism about responsibility (section 1), then we sketch a simple objection against it (section 2) and the reasons why the simple objection fails (section 3). Next, we outline the Responsibility First View (section 4) and we reformulate the case against agnosticism in a form that is not subject to the earlier difficulties (section 5). Finally, we consider free will, fairness and the circle of responsible agents.

1 Agnosticism about Moral Responsibility

There are three basic epistemological stances about moral responsibility. Some believe that normal human adults are often morally responsible, others deny that we know they are. The second approach, in turn, has two distinct versions. Members of the first group maintain that nobody is ever morally responsible (see Strawson, G. 1994; Pereboom 2001; Levy 2011), and they imply that this belief is justified for all of us. Members of the second group argue that we don't have enough evidence to tell. Our evidence is not decisive with regard to the existence of moral responsibility. Call philosophers who belong to this group "agnostics about moral responsibility." Typically, they suspend belief in the existence of moral responsibility, and they think that others should join them in doing so. A number of philosophers have put forward arguments to (roughly) that effect. They do not explicitly deny the reality of responsibility, but they argue that our most popular (compatibilist or incompatibilist) theories of moral responsibility make it unlikely that we could tell whether anyone is morally responsible (Byrd 2010, 2021; Sehon 2013, 2016; Kearns 2015).[1]

1 Note that although the papers cited here are centered around arguments that seem to support responsibility agnosticism, there are important differences between the accounts of the authors. Byrd (2010) embraces agnosticism about moral responsibility, claiming that the current debate should lead us to accept that we do not *now* know whether anyone is ever morally responsible. Nevertheless, in Byrd (2021) he seems to mitigate the strength of his argument, arguing in his conclusion that even if at present agnosticism about free will is reasonable, there seems to be hope that we can overcome our ignorance *in the future*. Kearns argues mostly for agnosticism about *free will*, but he also believes that "the thesis that we (don't) have free will [...] entails the moral claim that we are (not) morally responsible" (2015, 249), hence what he says amounts to a version of moral responsibility agnosticism too. What he insists is that we do not *know* whether there is free will or not, not that we are *unjustified* in believing it, nevertheless, by "not know" he means that our justification is so weak that it doesn't even meet a "low standard" (2015, 236). But having such a weak justification in a given question might very well warrant suspending belief. Sehon (2013, 2016) seems to be the furthest from the position of these agnostics, as he develops a certain variant of compatibilism in order to counter his own challenge against belief in moral responsibility. However, there is a good reason to consider and also to answer the agnostic *arguments* of these three authors together. The reason is that their agnostic arguments are logically independent from answers that they might come up with answering or at least mitigating moral responsibility agnosticism itself, as Sehon (2016) does. Naturally, one could be consistent in accepting their arguments that support agnosticism about moral responsibility while rejecting ways they may propose to evade this kind of agnosticism. (This is why it is no surprise that Sehon 2013 basically employs agnostic arguments, without offering a solution.) Therefore, one can scrutinize these arguments independently of the full-blown theories of the aforementioned authors. It is worth noting that their ways to avoid the agnostic conclusion seem

Those who try to resist agnosticism seek to show that we do have sufficient evidence—be it moral (van Inwagen 1983, 206–223; Coffman 2016), phenomenological (Guillon 2014), conceptual (Latham 2019), transcendental (Lockie 2018), or practice-based (Strawson, P. F. 1969)—to make our belief in the reality of moral responsibility justified. Agnostics, however, retort that such pieces of evidence are unreliable and open to objections. In this paper, we do not engage with that debate.[2] Rather, for the sake of argument, we take it for granted that there is no sufficient evidence for the reality of moral responsibility, and we try to show that one can rationally attribute moral responsibility to herself and others even in that case. Those who prefer to argue against responsibility agnosticism more directly, and believe that there is sufficient evidence in favor of moral responsibility, may still welcome our argument as an additional way to counter the agnostic.

The argument for agnosticism can be formulated in the following way:

(**AR1**) Nobody is justified in believing that the metaphysical conditions of moral responsibility are ever satisfied.

(**AR2**) If you are not justified in believing that a necessary condition of *X*'s existence is satisfied, then you are not justified in believing that *X* exists.

Therefore,

(**AR3**) Belief in the existence of moral responsibility is unjustified.

On a common interpretation of justification, the following principle is true:

to be controversial. For example, Sehon (2016) offers a non-standard, non-causal and at present unpopular blend of compatibilism that has met with serious criticism (Mele 2019). If this kind of criticism is correct, the argument of this paper may still be sound, as it offers a different way to counter agnosticism about moral responsibility. Furthermore, even if the ways to respond to moral responsibility agnosticism that are suggested by those who themselves employ agnostic arguments might work, our argument does not lose its significance: it is an additional way to counter the kind of agnosticism in question that could strengthen belief in moral responsibility even more.

2 It may be worthwhile to point out that our argument is somewhat akin to the transcendental arguments for free will and responsibility such as Robert Lockie's recent arguments (2018). For example, like Lockie's argument, we argue for the rationality of believing in moral responsibility is based on some analysis of conditions for bearing obligations. However, transcendental arguments aim to show that all rational (human) agents should believe in free will and responsibility (and Lockie's transcendental arguments share the same ambitions) whereas the argument we present attempts to show only that some agents can rationally believe in moral responsibility.

doi: 10.48106/dial.v76.i4.01

S. We should suspend those of our beliefs that are not justified.[3]

And so one can conclude that

ARC. We should suspend belief about the existence of moral responsibility.

The agnostic ascribes an epistemic obligation to those who assess the evidence regarding the existence of moral responsibility. The core intuition of our paper is that there is a serious tension between suspending beliefs about moral responsibility and ascribing epistemic obligations to oneself and others—intuitively, someone who is not morally responsible cannot have obligations. One can argue for this in two ways. First, one could argue that no one can be obliged to suspend belief about moral responsibility. Alternatively, one can say that holding a specific conception of moral responsibility makes it irrational to believe in obligations to suspend belief about moral responsibility. In the next section, we explore the first idea in order to see if a simple and intuitive argument could support it.

2 The Simple Objection

It might seem *prima facie* plausible that moral responsibility is a precondition of having obligations. If it is, then ascribing obligations while suspending belief about moral responsibility is irrational.

Why is it plausible that moral responsibility is a precondition of having obligations? One could appeal to the idea that only morally responsible agents can have normatively binding obligations. Consider the following argument:

The Simple Objection.

(**SO1**) If nobody is morally responsible, then nobody has normatively binding obligations.

(**SO2**) If nobody has normatively binding obligations, then nobody has a normatively binding obligation to suspend any of her beliefs.

3 For example, one of the most prominent moral skeptics writes: "To call a belief 'justified' is to say that the believer ought to hold that belief as opposed to suspending belief, because the believer has adequate epistemic grounds for believing that it is true (at least in some minimal sense)" (Sinnott-Armstrong 2008, 48).

(**SO3**) If nobody has a normatively binding obligation to suspend any of her beliefs, then nobody has a normatively binding obligation to suspend belief in moral responsibility.

(**SO4**) If nobody is morally responsible, then nobody has a normatively binding obligation to suspend belief in moral responsibility.

Either there are morally responsible agents or not. By **SO4**, if there are no such agents, then nobody is obliged to suspend belief in moral responsibility. On the other hand, if there *are* morally responsible agents, some of whom are obliged to suspend some of their beliefs which could not be the case were they not morally responsible, then, one might argue, nobody can have a good reason to suspend belief in moral responsibility. For—assuming that moral responsibility is a precondition of having normatively binding obligations—an agent cannot ascribe to herself an obligation to suspend belief in moral responsibility unless she also takes herself to be morally responsible. That sounds incoherent, so no one can consistently believe, in the light of **SO4**, that there is an obligation to be agnostic about moral responsibility. In short:

> Rationality Premise. If **SO4** is true, then nobody has a normatively binding obligation to suspend belief in moral responsibility.

If the argument so far is sound, it follows that

> SOC. No one has a normatively binding obligation to suspend belief in moral responsibility.

The conclusion of the Simple Argument is a threat to agnosticism because it implies that nobody has a normatively binding obligation to suspend her belief in moral responsibility, even if there is no direct evidence for the reality of moral responsibility. On the other hand, if nobody is in fact morally responsible, then the lack of evidence for the existence of moral responsibility does not matter, since the lack of evidence fails to have normatively binding consequences. The agnostic's claim that we ought to suspend belief in moral responsibility is thus refuted.

3 Why the Simple Objection Fails

One can challenge the Simple Objection on a number of grounds. Here we take into account two objections to **SO1** and one against the Rationality Premise.

doi: 10.48106/dial.v76.i4.01

The first problem about **SO1** is the following. Even if being morally responsible is a precondition of having moral obligations, some normatively binding obligations might not be moral in nature, and so having them does not entail being morally responsible. Indeed, normatively binding obligations come in many varieties. One might have epistemic, aesthetic, prudential, legal, as well as role obligations. Moreover, on many theories of epistemic obligation, epistemic obligations are not moral at all (see, for instance, Feldman 1988; Russell 2001). If the obligation to suspend judgment about moral responsibility is a non-moral, epistemic obligation, then one could have it even if one is not morally responsible. Hence **SO1** seems to be false.

Another important objection to **SO1** is that moral responsibility may not be a precondition of having moral obligations, or so the agnostic could argue. She could rightly claim that if we conceive of moral obligations in a certain way, then it is logically possible for agents who are not morally responsible to be nonetheless morally obliged to do something. For example, one might conceive of moral obligation in a consequentialist fashion and say that we have a moral obligation to maximize pleasure and minimize suffering. And it is possible that whether or not anyone is morally responsible, suspending belief in moral responsibility would minimize the amount of suffering. Thus, it could be the case that someone bears a (consequentialist) moral obligation to suspend belief in moral responsibility regardless of whether she is a morally responsible agent (see Smilansky 1994; Pereboom 2001; Waller 2004; Trakakis 2007).

Further, the Rationality Premise is open to the objection that there is a gap between the truth of a proposition and rationally believing that proposition. Even if **SO1**, **SO2**, and **SO3** are true, it does not necessarily follow that everyone is rational in believing any of those premises. An agent's epistemic position might be such that her evidence either contradicts one of **SO1**–**SO3** or does not justify any of them. The agent's evidence may even be such that it is rational for her to believe in the soundness of the agnostic's argument. So even if **SO4** is true, there could be a normatively binding obligation to suspend belief in moral responsibility. Therefore, the Rationality Premise appears to be false.

In order to avoid these difficulties, we need to modify the Simple Objection. Instead of talking about normatively binding obligations, we will focus on more specific ones. To evade the second challenge, we will base the argument on a particular conception of moral responsibility, one that, if rationally upheld, renders it irrational to ascribe to oneself an obligation to be agnostic about moral responsibility. Finally, to avoid the difficulties with the Rationality

Premise, we will defend only those agents' beliefs who rationally accepted such a conception. The next section describes the conception that we will work with, the Responsibility First View, in detail.

4 The Responsibility First View

Consider the following famous passage from Wittgenstein:

> Supposing that I could play tennis and one of you saw me playing and said "Well, you play pretty badly" and suppose I answered "I know, I'm playing pretty badly but I don't want to play any better," all the other man could say would be "Ah, then that's all right." But suppose I had told one of you a preposterous lie and he came up to me and said, "You're behaving like a beast" and then I were to say "I know I behave badly, but then I don't want to behave any better," could he then say "Ah, then that's all right"? Certainly not; he would say "Well, you ought to want to behave better." (1965, 5)

When someone says "well, you play pretty badly", in most cases she is not merely offering a description but implies roughly the following: "you should do something about it if you don't want to look ridiculous." In Wittgenstein's story, the player in effect replies that he does not care about this implied "should." Using contemporary terms, one could say that the implied "should" expressed a prudential obligation to prevent an undesirable outcome, such as being ridiculed. Note that even if we suppose that the player would be unhappy if someone actually ridiculed him, he could nonetheless reply "that's all right," because being imprudent is not an unacceptable normative error. It seems to be implausible to think that one should avoid prudential errors with all her strength in every situation or that one should feel remorse if she made such an error. Sometimes, it is all right not to care about prudential obligations even if they actually bind the agent. In other words, it might be OK to neglect them even if violating them constitutes a basis for some negative treatment (such as ridicule).

However, the second example suggests that violating an obligation *is* a normative error that is unacceptable to such a great extent that one should feel remorse and should avoid repeating the error with all her strength. These violations are just not "all right"; they cannot be shrugged off. Wittgenstein and many other philosophers claim that moral obligations fall into this category.

doi: 10.48106/dial.v76.i4.01

Violating them results in unacceptable normative errors. Further, Wittgenstein's paper seems to imply that *only* the violation of moral obligations results in such an error. We will call this idea the Moral Primacy Thesis (MPT).

MPT is central to our case, so we would like to express it more precisely (incidentally explaining why the term "moral primacy" is apt). The following definition of "all things considered obligations" will be useful for that purpose:

> All things considered obligation to $\Phi =_{df}$. An obligation which is not overridden by any other obligation (in the given case) and which prescribes doing Φ to agent S in a way that S should avoid violating the obligation with all her strength; and if S fails to observe the obligation to Φ, then S should feel remorse.

We follow here Searle (1978) and many other philosophers who used the term "all things considered obligation." Nevertheless, we add that a genuine all things considered obligation to perform a specific action must have a normative weight that makes its violation normatively unacceptable. If an obligation does not have the significant normative weight, then—all things considered—it is permissible to ignore it.

Philosophers often talk about obligations that have a tendency to constitute all things considered obligations. They call these *pro tanto* or *prima facie* obligations (Ross 1930). These tend to constitute all things considered obligations if other, stronger obligations do not override them. (The paradigmatic examples are moral obligations.) However, for our purposes, it is better to not commit ourselves to any specific understanding of *pro tanto* or *prima facie* obligations because not only the difference between *pro tanto* and all things considered obligations is relevant for our argument but the difference between obligations that can constitute all things considered obligations in themselves and obligations that can do this only by the help of other obligations. So, instead of talking about *prima facie* and *pro tanto* obligations, we will use the term "strong obligation", defined as follows:

> Strong obligation to $\Phi =_{df}$. An obligation that constitutes an all things considered obligation to Φ (in the given case) unless it is overridden by some other strong obligation(s) to do something else.

So in some cases, Strong obligations to Φ constitute an all things considered obligation to Φ, and in other cases, strong obligations to Φ do not constitute

an all things considered obligation to Φ (if they are overridden by other strong obligations).[4]

In addition, there are obligations that fail to constitute all things considered obligations in spite of the fact that nothing overrides them. For instance, in many cases, prudential obligations do not constitute all things considered obligations even though the agent has no other kind of obligation. This is precisely the case in Wittgenstein's example: although the agent has a prudential obligation to play tennis better, he is free to ignore and violate it. We call these obligations *weak* obligations.

> Weak obligation to Φ $=_{df}$. An obligation that does not constitute an all things considered obligation to Φ unless it appropriately relates to a strong obligation in the given case.

We intentionally use the vague term "appropriately relates." It is a complicated question when and how strong obligations turn weak obligations into all things considered obligations. For the present purposes, what matters is that this is certainly possible—whatever the details. For instance, if the tennis player in Wittgenstein's example had previously promised his wife to do his best and avoid ridicule, and there was no strong obligation to override the obligation to keep his promise, then he would have an all things considered obligation to play better. In this case, his prudential obligation to play better would be an all things considered obligation, because it would be appropriately related to his moral obligation to fulfill his promise to avoid ridicule.

Using the terminology just introduced, we can now characterize MPT more precisely:

> Moral Primacy Thesis (MPT). All moral obligations are strong obligations and every other kind of obligation is weak.

In other words, only moral obligations can constitute all things considered obligations without being appropriately related to other kinds of obligations. On the other hand, prudential, epistemic, role, legal, etc. obligations can only

4 It is worthwhile to note that our notion of strong obligation resembles the Kantian notion of categorical imperative. The main difference is that our notion of strong obligation does not imply universalizability. That is, we do not deny the possibility that one may have an obligation that constitutes an all things considered obligation to Φ, although one cannot at the same time will that it becomes a universal law.

doi: 10.48106/dial.v76.i4.01

constitute all things considered obligations if they are appropriately related to moral obligations.

It is easy to see that one of the relevant consequences of MPT is the following:

> Normative Weakness of Non-Moral Obligations. Nobody has to avoid violating with all her strength those non-moral obligations that do not relate appropriately to any of her moral obligations and if someone fails to observe such an obligation, she should not feel remorse.

Attributing moral obligation to agents must have some conditions. For example, it would certainly be absurd to attribute moral obligations to beings that are incapable to act, because it makes no sense to say that they should avoid doing something with all their strength. Whatever the relevant conditions are, there is an obvious term, under MPT, for those beings who fulfill all of them: they are the morally responsible agents. So proponents of MPT are free to adopt the following thesis as a component of their moral framework:

> Responsibility Identity Thesis (RIT). The property of being morally responsible is identical to the property of fulfilling all conditions for bearing moral obligations.

RIT makes moral responsibility into a precondition of having moral obligations, and it also makes the former prior to the latter in a certain respect. Since moral obligations are, in turn, prior to any other type of obligations under MPT, we will call the combination of MPT and RIT the Responsibility First View.

5 The Primacy Argument

We can now turn to the revision of the Simple Objection. The proponent of RFV can answer the agnostic's challenge as follows: "You claim that I should suspend my belief in moral responsibility because I cannot prove that anyone meets the conditions for being morally responsible. However, based on my conception of morality and responsibility, if I am not morally responsible, then I do not have any moral obligations. And if I have no moral obligations, I do not have any obligations that I should fulfill with all my strength, any obligations

that should seriously concern me. In technical terms, I do not have all things considered obligations. So if I am not morally responsible, then it is all right for me to disregard your demand about suspension of belief. And in case you claimed that I have an all things considered obligation to suspend my belief, an obligation which I cannot disregard without committing a normative fault I should regret, then I conclude on the basis of my conception of responsibility that I am a morally responsible being after all. Either way, I can rationally resist your challenge and keep believing in moral responsibility."

We would like to express this revised version of the Simple Objection more formally:

The Primacy Argument.

(**PA1**) No agent can rationally think that she has an all things considered obligation to suspend her belief that she fulfills the necessary conditions of having all things considered obligations.

(**PA2**) Someone who rationally upholds RFV cannot rationally think that she has an all things considered obligation to suspend her belief that she is morally responsible.

(**PA3**) If someone cannot rationally think that she has an all things considered obligation to suspend her belief that she is morally responsible, then she is rational to reject agnosticism about moral responsibility.

(**PAC**) Rejecting agnosticism about moral responsibility is rational for anyone who rationally upholds RFV.

Until the agnostic does not challenge the moral framework that proponents of RFV employ, she cannot undermine their belief in their own moral responsibility. The agnostic cannot challenge belief in moral responsibility by merely pointing out that evidence for the existence of moral responsibility is insufficient. What is more, if someone upholds RFV rationally, then it would be straightforward irrational for her to accept the agnostic's conclusion—unless she finds out that her own moral framework is untenable.

The proponent of RFV gains a huge dialectical advantage by deploying the Primacy Argument. Due to the Primacy Argument, the debate shifts from the sufficiency of evidence to the tenability of a specific moral framework. Defending the tenability of RFV seems to be much easier than defending the sufficiency of evidence regarding the existence of moral responsibility. Especially so if the proponent of the Moral Primacy Thesis, by investigating

doi: 10.48106/dial.v76.i4.01

the nature of moral obligation and responsibility, comes to the conclusion that moral responsibility has heavy-weight metaphysical preconditions such as libertarian free will, since scientific evidence for libertarian freedom seems to be lacking. (We will say more about free will in the last section.)

Moreover, as far as we can tell, both MPT and RIT can be supported by considerable arguments. Even though some consequentialists deny that moral obligation implies moral responsibility, that principle seems to be fundamental and obvious for almost everyone—as even consequentialist critics note (Waller 2004, 427–428). And someone who upholds RIT can explain why that principle is true: being a morally responsible agent is the same as being a potential bearer of moral obligations.

MPT also has notable advantages. Many people find it plausible that moral obligations can override all other obligations. MPT explains why this is the case: the set of moral obligations is identical to the set of strong obligations. Additionally, MPT provides a substantive definition of moral obligation: moral obligations are those obligations that can constitute all things considered obligation without the involvement of other types of obligation. Another notable advantage of MPT is that it helps understanding why some obligations can be neglected without normative costs in certain cases but not in others. Take, for example, the highway code, which prescribes various patterns of behaviour (call them "legal obligations"). Some of those prescriptions can be non-culpably neglected in a completely abandoned city. Still, in most cases, violating them is normatively unacceptable. One can use MPT to explain this phenomenon by pointing out that the highway code contains weak obligations. In most cases, they are appropriately related to moral obligations (for example, to the obligation to secure the safety of human beings). However, they are not appropriately related to moral obligations in an abandoned city.

Of course, anyone, including the agnostic, can argue against RFV. Indeed, it seems that one can find not only prominent supporters of RFV (Kant seems to be the most obvious example) but able critics too. For instance, Bernard Williams criticizes an ethical system under the label "morality" that contains, among other things, RFV (see Williams 2006, 174–196), because he believes that moral systems with such a strong notion of moral obligations threaten personal integrity. Even though the investigation of such counterarguments that are based on such wide-ranging considerations about the relation between a whole system of morality and other basic values is out of the scope of our paper, we can deal with another argument against RFV that is rather closely related to the problem of moral responsibility.

Namely, RFV seems to imply that epistemic obligations are identical with or, at least, not independent of moral obligations which means, in turn, that anyone who accepts RFV and would like to believe in epistemic norms is forced to believe in moral responsibility no matter which crazy theory about conditions of moral responsibility turns out to be true. For example, if Derk Pereboom's analysis on the conditions of moral responsibility is correct, moral responsibility needs not only agent-causation (which, according to Pereboom, may be a logically incoherent concept), but either systematic breaking of the laws of nature or inexplicable harmony between micro-physical statistical laws and the free decisions of the agents (see Pereboom 2001). For sure, believing that these conditions are met in reality would be a high price to pay for holding RFV. Insofar as the price is so high, it seems to be not only irrelevant, but weird that the proponent of RFV can and even should rationally defend believing in moral responsibility and its monstrous metaphysics by moving the battlefield from metaphysics to metaethics. After all, forming rational beliefs and fulfilling epistemic norms aim at the truth, and it is not too probable that this way of belief-formation leads us to true beliefs.[5]

To be clear, RFV does not imply that epistemic obligations as such depend on (or are identical with) moral obligations. RFV does not exclude that they are totally unrelated to each other. What RFV implies is only that an epistemic obligation has to be appropriately related to some moral obligations in order to be true that agents have to avoid violating it with all their strength and if someone fails to observe an epistemic obligation which does not relate appropriately to any moral obligation, she should not feel remorse. In other words, in themselves, epistemic obligations do not have sufficient normative weight to constitute all things considered obligations. So, if one both accepts RFV and rejects moral responsibility based on her evidence-basis, she cannot rationally believe that there are moral and all things considered obligations, but she still can rationally think that there are (weak) epistemic obligations. What she cannot rationally believe is that neglecting any epistemic obligation cannot be OK in the same way as neglecting the prudential obligation not to ridicule oneself. That is, even if one accepts RFV and, for instance, Pereboom's assessment of the evidence about free will and moral responsibility, she can rationally deny the existence of (a metaphysically rather extreme kind of) free will, moral responsibility, moral obligations, and all things considered obligations. The only thing that she cannot rationally maintain without re-

5 We would like to thank an anonymous reviewer for pointing out this possible objection.

doi: 10.48106/dial.v76.i4.01

jecting RFV is the idea that anyone should suspend belief in those things to avoid committing an unacceptable normative error that cannot be shrugged off. In other words, if there is a proponent of RFV who tries to heroically find the truth no matter the cost and finds that her evidence-basis strongly indicates the non-existence of moral responsibility, she can rationally believe that she has an epistemic (or even prudential) obligation to deny the existence of moral responsibility, but she cannot rationally think either that she has an all things considered obligation to reject moral responsibility or that anyone has an all things considered obligation to try to find the truth no matter the cost.

Nonetheless, none of this undermines our point that the proponent of RFV, if she wants to defend the belief in moral responsibility, can move the battlefield from metaphysics to metaethics, and the latter seems to be much more advantageous for her, especially if she also holds that the sufficient conditions of moral responsibility are metaphysically rather demanding. The more demanding these conditions are, the less plausible is the claim that the existence of moral responsibility is obvious and/or probable in the light of the given evidence, so moving the battlefield from metaphysics to metaethics provides more strategic advantage.

Note, parenthetically, that one can construct a modified version of the Primacy Argument even if both MPT and RIT are untenable. One need not appeal to morality (or moral responsibility) at all. Anyone who rationally believes in all things considered obligations has the epistemic right to sustain belief in a specific kind of responsibility. As the first premise of the Primacy Argument says, no agent can rationally think she has an all things considered obligation to suspend the belief that she fulfills the necessary conditions of having all things considered obligations. In other words, someone who rationally attributes all things considered obligations to herself must also accept that she fulfills the necessary conditions of having all things considered obligations. Since having strong obligations is one of those necessary conditions, the agent in question must also accept that she fulfills the necessary conditions of having strong obligations. It is reasonable to say that being responsible "in a strong sense" requires fulfilling all necessary conditions for bearing strong obligations, so anyone who rationally attributes all things considered obligations to herself can rationally attribute "strong responsibility" to herself as well. It seems that this argument for "strong responsibility" can be threatened only by arguments against the existence of all things considered obligations.

To sum up the Primacy Argument, anyone who rationally accepts RFV can rationally maintain her belief in moral responsibility even if she does not have sufficient direct evidence that anyone fulfills the metaphysical conditions of being morally responsible. Until the agnostic refutes MPT or RIT, one can rationally resist the agnostic challenge.

6 Free Will, Fairness, and Others

Various questions could be raised about our argument. We will look at three. First, one might ask how the dialectic is related to free will. We claimed that someone who rationally believes in RFV does not have to suspend her belief in moral responsibility even if she lacks direct evidence for it. Could RFV be used to defend belief in free will as well?

The answer to this question depends on one's conception of free will. There are two basic approaches in the literature. According to the first, having free will means fulfilling a subset of conditions that guarantee necessary control over one's morally relevant actions (Clarke 1992). The present argument obviously extends to the defense of free will conceived this way. If someone rationally accepts RFV and also rationally thinks that she fulfills all necessary conditions for being morally responsible, then she cannot rationally believe that she fails to fulfill a subset of those conditions, namely those that are necessary for control. So our argument supports belief in free will for those who rationally believe RFV and identify having free will with fulfilling a subset of necessary control conditions for being morally responsible.

However, there is another prevalent conception, according to which free will is the ability to do otherwise (van Inwagen 1983). Our argument can be extended to this case as well, but only if one rationally upholds that the ability to do otherwise is a necessary precondition of being morally responsible. Given strong evidence that moral responsibility depends on free will of the second sort, then rational belief in RFV (together with the evidence in question) can ground rational belief in the existence of free will. And if the proponent of RFV has sufficient evidence that moral responsibility has further metaphysical conditions, she can also rationally believe that she fulfills all those further conditions, regardless of how demanding they are metaphysically.

These possible extensions of the Primacy Argument are especially significant if one takes into account that many philosophers and scientists insist that there is no sufficient scientific or other evidence for macro-level psychological indeterminism (which is a precondition of libertarian free will)

doi: 10.48106/dial.v76.i4.01

or the presence of compatibilist-friendly causal determinism in the brain. In light of the possible extensions of the Primacy Argument, proponents of RFV can rationally believe in responsibility-relevant free will (of either the incompatibilist or compatibilist sort) even in the absence of sufficient direct scientific or other evidence.

This last point regarding the absence of evidence leads us on to a potential objection implied by Scott Sehon. He emphasizes that we treat responsible and irresponsible agents very differently. If, for example, someone pushes another person into the traffic, we treat her act very differently depending on whether she was or was not responsible. If she was, then her act "certainly looks incredibly reprehensible and maybe even the stuff of an attempted murder charge" (Sehon 2013, 369). But if we know that the pusher is not responsible, we would not call her action "reprehensible" and would not make her face serious charges. Sehon adds that "[it] would be manifestly unfair to regard the agent as responsible if our degree of certainty on the matter is quite low" (2013, 36). One could extend this point and argue that if we lack strong direct evidence for moral responsibility, then, out of fairness, we should suspend belief about whether anyone is ever responsible in a way that would render retribution justified.

Proponents of the Primacy Argument evidently disagree, as their supposedly rational belief in RFV makes them rational in holding that they can be morally responsible for their actions. It is important that Sehon brings up this issue in terms of fairness. The obligation to be fair with others is naturally understood to be a moral obligation, and hence a strong one—according to MPT. Those who uphold RFV will see the situation as follows. The obligation to be fair can only be attributed to morally responsible persons. If nobody is morally responsible, then the strong obligation to be fair cannot be attributed to anybody. And if that is the case, then nobody has to care about being fair to anybody. So if the proponent of RFV takes Sehon's exhortation to be fair seriously, and if she thinks she has to care about it, then, in the light of MPT, she incidentally attributes a strong obligation to herself. As a result, she implies that she fulfills all the conditions of having strong obligations, including having moral responsibility. That is, for proponents of RFV, Sehon's point can only have force if they take themselves to be morally responsible. They would need to assume, first, that they are morally responsible, and, second, they would have to suspend judgment about the existence of moral responsibility because of that very assumption—which seems incoherent. Thus, the argument that insufficient direct evidence for moral responsibility should make us

suspend our belief in moral responsibility because it might lead to the unfair treatment of others makes no sense to those who hold RFV to be true.

The question of being fair to others and taking them to be morally responsible brings us to the crucial issue of the circle of agents whom one might attribute moral responsibility to, on the basis of the Primacy Argument. This is a crucial issue, as it might very well be the case that the individuals who accept RFV can attribute moral responsibility only to themselves and not to anyone else. This is because **PA1** takes only the agent's own perspective into account. The agent is considering her own obligations and moral responsibility, and the reason why she doesn't have to become an agnostic is that, were she to take agnosticism as a strong obligation, she would thereby attribute moral responsibility to *herself*. (As we have indicated, she might even go on to attribute free will as well.) But the incoherence would arise only in her own case, so the Primacy Argument's conclusion applies only to her: she is free to go on believing that she, for one, is morally responsible. And clearly, she can believe in the existence of moral responsibility on the basis on *that*, since moral responsibility exists even if only one agent has it.

Extending this rather small circle of responsible agents might look unreasonable or unfair indeed. However, there could be ways to do it. Remember that the agent in question reasonably believes in her own responsibility. If she considers agents who seem to be like her in every relevant respect, she may take them to be morally responsible as well. Nevertheless, the reasonableness of this move depends on two crucial factors. First, the agent must have a rationally held theory of what the relevant respects are. Second, were she to deem morally responsible any agent other than herself, her judgment that that person is similar to herself in every relevant respect must also be rational. This means that reasonably extending the circle of morally responsible agents to others is logically possible, but could be difficult in practice. Fortunately, there might be an easier way. It seems reasonable to think that all fully developed human beings have the same metaphysical structure. Insofar as this assumption is reasonable, a proponent of the Primacy Argument can extend the circle of morally responsible agents to all fully developed human adults who fulfill the non-metaphysical and empirically verifiable conditions of moral responsibility, whether or not she can identify the precise metaphysical conditions for being morally responsible.

Note that extending the circle of responsibility poses a challenge not only with regard to other agents, but also with regard to the agent who can rationally believe in moral responsibility based on RFV and PA. This is because The

doi: 10.48106/dial.v76.i4.01

Primacy Argument does not imply that the agent is morally responsible all the time. It only permits the agent to believe that she is morally responsible in her present state. Nevertheless, what we have said previously about the possibilities of extending the circle of morally responsible agents can also be used to extend this temporal limitation. This means that if an earlier or later state of the agent seems to be similar in every relevant respect to her present state, then she may take it that she was or is going to be morally responsible at those times. However, it might not be clear in every case that these conditions are fulfilled. Therefore, our argument is compatible with accepting that even though we are reasonable in thinking that we are morally responsible some of the time, we could be also reasonable in thinking that we are not responsible at other times, or thinking that we should be agnostics about the question whether we are morally responsible in certain situations.*

László Bernáth
Research Centre for the Humanities
Budapest
bernath.laszlo@abtk.hu

Tamás Paár
Central European University
Budapest
paar.tamas@gmail.com

* We would like to express our indebtedness to the anonymous referees of this paper, their insightful comments helped us developing our position a great deal. We also thank Dániel Kodaj for his suggestions and the stimulating exchanges we had on many occasions. László Bernáth was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences (grant no. BO/00432/18/2), the OTKA (Hungarian Scientific Research Fund by the National Research Development and Innovation Office) Postdoctoral Excellence Programme (grant no. PD131998), and another OTKA research grant (grant no. K132911). Tamás Paár was supported by the ÚNKP-19-3-III-PPKE-24 New National Excellence Program of the Ministry for Innovation and Technology. His research was supported by the University of Oxford project "New Horizons for Science and Religion in Central and Eastern Europe" funded by the John Templeton Foundation. (The opinion expressed in the publication are those of the author(s) and do not necessarily reflect the view of the John Templeton Foundation.) We owe special thanks to the organizers and the committee of the Oxford University Essay Competition for Central and Eastern European Perspectives on Science, Theology and Philosophy in which this paper has been awarded first prize.

References

Byrd, Jeremy. 2010. "Agnosticism about Moral Responsibility." *Canadian Journal of Philosophy* 40(3): 411–432, doi:10.1080/00455091.2010.10716729.

—. 2021. "What Should we Believe About Free Will?" *Erkenntnis* 86(3): 505–522, doi:10.1007/s10670-019-00116-3.

Clarke, Randolph. 1992. "Free Will and the Conditions of Moral Responsibility." *Philosophical Studies* 66(1): 53–72, doi:10.1007/bf00668395.

Coffman, E. J. 2016. "Incompatibilist Commitment and Moral Self-Knowledge: The Epistemology of Libertarianism." in *Philosophical Issues 26: Knowledge and Mind*, edited by Christoph Kelp and Jack Lyons, pp. 78–98. Oxford: Wiley-Blackwell, doi:10.1111/phis.12066.

Feldman, Richard H. 1988. "Epistemic Obligations." in *Philosophical Perspectives 2: Epistemology*, edited by James E. Tomberlin, pp. 235–256. Oxford: Basil Blackwell Publishers, doi:10.2307/2214076.

Guillon, Jean-Baptiste. 2014. "Van Inwagen on Introspected Freedom." *Philosophical Studies* 168(3): 645–663, doi:10.1007/s11098-013-0159-x.

Kearns, Stephen. 2015. "Free Will Agnosticism." *Noûs* 49(2): 235–252, doi:10.1111/nous.12032.

Latham, Andrew J. 2019. "The Conceptual Impossibility of Free Will Error Theory." *European Journal of Analytic Philosophy* 15(2): 99–120, doi:10.31820/ejap.15.2.5.

Levy, Neil. 2011. *Hard Luck. How Luck Undermines Free Will & Moral Responsibility*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199601387.001.0001.

Lockie, Robert. 2018. *Free Will and Epistemology. A Defence of the Transcendental Argument for Freedom*. London: Bloomsbury Academic.

Mele, Alfred R. 2019. "Causalism: On Action Explanation and Causal Deviance." in *Explanation in Action Theory and Historiography. Causal and Teleological Approaches*, edited by Gunnar Schumann, pp. 45–58. London: Routledge, doi:10.4324/9780429506048-2.

Pereboom, Derk. 2001. *Living Without Free Will*. Cambridge: Cambridge University Press, doi:10.1017/CBO9780511498824.

Ross, William David. 1930. *The Right and the Good*. Oxford: Oxford University Press.

Russell, Bruce. 2001. "Epistemic and Moral Duty." in *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, edited by Matthias Steup, pp. 34–47. Oxford: Oxford University Press, doi:10.1093/0195128923.003.0003.

Searle, John R. 1978. "Prima Facie Obligations." in *Practical Reasoning*, edited by Joseph Raz, pp. 81–90. Oxford: Oxford University Press. Reprinted in van Straaten (1980, 238–259), doi:10.1093/acprof:oso/9780199693818.003.0007.

Sehon, Scott R. 2013. "Epistemic Issues in the Free Will Debates: Can we Know when We Are Free?" *Philosophical Studies* 166(2): 363–380, doi:10.1007/s11098-012-0044-z.

—. 2016. *Free Will and Action Explanation. A Non-Causal, Compatilist Account*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780198758495.001.0001.

Sinnott-Armstrong, Walter. 2008. "Framing Moral Intuitions." in *Moral Psychology, Volume 2. The Cognitive Science of Morality: Intuition and Diversity*, edited by Walter Sinnott-Armstrong, pp. 47–76. Cambridge, Massachusetts: The MIT Press.

Smilansky, Saul. 1994. "Ethical Advantages of Hard Determinism ." *Philosophy and Phenomenological Research* 54(1): 355–363, doi:10.2307/2108494.

Strawson, Galen. 1994. "The Impossibility of Ultimate Moral Responsibility." *Philosophical Studies* 75(1–2): 5–24. Reprinted in Strawson, G. (2008, 319–336), doi:10.1007/bf00989879.

—. 2008. *Real Materialism, and Other Essays*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199267422.001.0001.

Strawson, Peter Frederick. 1969. *Meaning and Truth*. Oxford: Oxford University Press. Reprinted in Strawson, P. F. (1971, 170–189) and in Strawson, P. F. (2004, 131–146).

—. 1971. *Logico-Linguistic Papers*. London: Methuen & Co. Reprinted as Strawson, P. F. (2004).

—, ed. 2004. *Logico-Linguistic Papers*. 2nd ed. Aldershot, Hampshire: Ashgate Publishing Limited, doi:10.4324/9781315250250.

Trakakis, Nick N. 2007. "Whither Morality in a Hard Determinist World? ." *Sorites* 19(1): 14–40, http://lorenzopena.es/Sorites/Issue_19/sorites19.pdf.

van Inwagen, Peter. 1983. *An Essay on Free Will*. Oxford: Oxford University Press.

van Straaten, Zak, ed. 1980. *Philosophical Subjects: Essays Presented to P.F. Strawson*. Oxford: Oxford University Press.

Waller, Bruce N. 2004. "Virtue Unrewarded: Morality without Moral Responsibility." *Philosophia: Philosophical Quarterly of Israel* 31(3–4): 427–447, doi:10.1007/bf02385194.

Williams, Bernard Arthur Owen. 1985. *Ethics and the Limits of Philosophy*. Hammersmith: Fontana Press.

—. 2006. *Ethics and the Limits of Philosophy*. London: Routledge. With a commentary on the text by A.W. Moore; original publication: Williams (1985), doi:10.4324/9780203969847.

Wittgenstein, Ludwig. 1965. "A Lecture on Ethics." *The Philosophical Review* 74(1): 3–12, doi:10.2307/2183526.