

# Responsibility First

How to Resist Agnosticism About Moral  
Responsibility

LÁSZLÓ BERNÁTH & TAMÁS PAÁR

doi:[10.48106/dial.v76.i4.01](https://doi.org/10.48106/dial.v76.i4.01)

László Bernáth & Tamás Paár. 2022. “Responsibility First:  
How to Resist Agnosticism About Moral Responsibility.”  
*Dialectica* 76(4): 525–545. doi:[10.48106/dial.v76.i4.01](https://doi.org/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”). 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.

## 2.1 Agnosticism About Moral Responsibility

26 There are three basic epistemological stances about moral responsibility. Some  
 27 believe that normal human adults are often morally responsible, others deny  
 28 that we know they are. The second approach, in turn, has two distinct versions.  
 29 Members of the first group maintain that nobody is ever morally responsible  
 30 (see Strawson, G. 1994; Pereboom 2001; Levy 2011), and they imply that this  
 31 belief is justified for all of us. Members of the second group argue that we  
 32 don't have enough evidence to tell. Our evidence is not decisive with regard  
 33 to the existence of moral responsibility. Call philosophers who belong to this  
 34 group "agnostics about moral responsibility." Typically, they suspend belief in  
 35 the existence of moral responsibility, and they think that others should join  
 36 them in doing so. A number of philosophers have put forward arguments to  
 37 (roughly) that effect. They do not explicitly deny the reality of responsibility,  
 38 but they argue that our most popular (compatibilist or incompatibilist)  
 39 theories of moral responsibility make it unlikely that we could tell whether  
 40 anyone is morally responsible (Byrd 2010, 2021; Sehon 2013, 2016; Kearns  
 41 2015).<sup>1</sup>

---

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.<sup>2</sup> 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.

<sup>2</sup> 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.

63 S. We should suspend those of our beliefs that are not justified.<sup>3</sup>

64 And so, one can conclude that

65 ARC. We should suspend belief about the existence of moral re-  
66 sponsibility.

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

## 72 The Simple Objection

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

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

### 85 SIMPLE OBJECTION.

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

88 (SO2) If nobody has normatively binding obligations, then nobody has a nor-  
89 matively 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).

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

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

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

106 RATIONALITY PREMISE. If (SO4) is true, then nobody has a norma-  
 107 tively binding obligation to suspend belief in moral responsibility.

108 If the argument so far is sound, it follows that

109 (SOC) No one has a normatively binding obligation to suspend belief in moral  
 110 responsibility.

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

### 1193 Why the Simple Objection Fails

120 One can challenge the **SIMPLE OBJECTION** on a number of grounds. Here we  
 121 take into account two objections to (SO<sub>1</sub>) and one against the **RATIONALITY**  
 122 **PREMISE**.

123 The first problem about (SO<sub>1</sub>) is the following. Even if being morally re-  
 124 sponsible is a precondition of having moral obligations, some normatively  
 125 binding obligations might not be moral in nature, and so having them does  
 126 not entail being morally responsible. Indeed, normatively binding obligations  
 127 come in many varieties. One might have epistemic, aesthetic, prudential, legal,  
 128 as well as role obligations. Moreover, on many theories of epistemic obligation,  
 129 epistemic obligations are not moral at all (see, for instance, [Feldman 1988](#);  
 130 [Russell 2001](#)). If the obligation to suspend judgment about moral responsibil-  
 131 ity is a non-moral, epistemic obligation, then one could have it even if one is  
 132 not morally responsible. Hence (SO<sub>1</sub>) seems to be false.

133 Another important objection to (SO<sub>1</sub>) is that moral responsibility may not  
 134 be a precondition of having moral obligations, or so the agnostic could argue.  
 135 She could rightly claim that if we conceive of moral obligations in a certain  
 136 way, then it is logically possible for agents who are not morally responsible  
 137 to be nonetheless morally obliged to do something. For example, one might  
 138 conceive of moral obligation in a consequentialist fashion and say that we  
 139 have a moral obligation to maximize pleasure and minimize suffering. And  
 140 it is possible that whether or not anyone is morally responsible, suspending  
 141 belief in moral responsibility would minimize the amount of suffering. Thus,  
 142 it could be the case that someone bears a (consequentialist) moral obligation  
 143 to suspend belief in moral responsibility regardless of whether she is a morally  
 144 responsible agent (see [Smilansky 1994](#); [Pereboom 2001](#); [Waller 2004](#); [Trakakis](#)  
 145 [2007](#)).

146 Further, the **RATIONALITY PREMISE** is open to the objection that there  
 147 is a gap between the truth of a proposition and rationally believing that  
 148 proposition. Even if (SO<sub>1</sub>), (SO<sub>2</sub>), and (SO<sub>3</sub>) are true, it does not necessarily  
 149 follow that everyone is rational in believing any of those premises. An agent's  
 150 epistemic position might be such that her evidence either contradicts one  
 151 of (SO<sub>1</sub>)–(SO<sub>3</sub>) or does not justify any of them. The agent's evidence may  
 152 even be such that it is rational for her to believe in the soundness of the  
 153 agnostic's argument. So even if (SO<sub>4</sub>) is true, there could be a normatively  
 154 binding obligation to suspend belief in moral responsibility. Therefore, the  
 155 **RATIONALITY PREMISE** appears to be false.

156 In order to avoid these difficulties, we need to modify the **SIMPLE OBJEC-**  
 157 **TION**. Instead of talking about normatively binding obligations, we will focus  
 158 on more specific ones. To evade the second challenge, we will base the argu-  
 159 ment on a particular conception of moral responsibility, one that, if rationally  
 160 upheld, renders it irrational to ascribe to oneself an obligation to be agnostic  
 161 about moral responsibility. Finally, to avoid the difficulties with the **RATIO-**  
 162 **NALITY PREMISE**, we will defend only those agents' beliefs who rationally  
 163 accepted such a conception. The next section describes the conception that  
 164 we will work with, the Responsibility First View, in detail.

#### 164 **The Responsibility First View**

166 Consider the following famous passage from Wittgenstein:

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

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



191 However, the second example suggests that violating an obligation *is* a norma-  
 192 tive error that is unacceptable to such a great extent that one should feel  
 193 remorse and should avoid repeating the error with all her strength. These vio-  
 194 lations are just not “all right”; they cannot be shrugged off. Wittgenstein and  
 195 many other philosophers claim that moral obligations fall into this category.  
 196 Violating them results in unacceptable normative errors. Further, Wittgen-  
 197 stein’s paper seems to imply that *only* the violation of moral obligations results  
 198 in such an error. We will call this idea the Moral Primacy Thesis (MPT).

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

202 ALL THINGS CONSIDERED OBLIGATION TO  $\varphi =_{df}$ . An obligation  
 203 which is not overridden by any other obligation (in the given case)  
 204 and which prescribes doing  $\varphi$  to agent *S* in a way that *S* should avoid  
 205 violating the obligation with all her strength; and if *S* fails to observe  
 206 the obligation to  $\varphi$ , then *S* should feel remorse.

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

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

225 STRONG OBLIGATION TO  $\varphi =_{df}$ . An obligation that constitutes an  
 226 all things considered obligation to  $\varphi$  (in the given case) unless it is  
 227 overridden by some other strong obligation(s) to do something else.

228 So in some cases, strong obligations to  $\varphi$  constitute an all things considered  
 229 obligation to  $\varphi$ , and in other cases, strong obligations to  $\varphi$  do not constitute  
 230 an all things considered obligation to  $\varphi$  (if they are overridden by other strong  
 231 obligations).<sup>4</sup>

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

239 WEAK OBLIGATION TO  $\varphi =_{df}$ . An obligation that does not constitute  
 240 an all things considered obligation to  $\varphi$  unless it appropriately relates  
 241 to a strong obligation in the given case.

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

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

---

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  $\varphi$ , although one cannot at the same time will that it becomes a universal law.

254 MORAL PRIMACY THESIS (MPT). All moral obligations are strong  
 255 obligations and every other kind of obligation is weak.

256 In other words, only moral obligations can constitute all things considered  
 257 obligations without being appropriately related to other kinds of obligations.  
 258 On the other hand, prudential, epistemic, role, legal, etc. obligations can only  
 259 constitute all things considered obligations if they are appropriately related  
 260 to moral obligations.

261 It is easy to see that one of the relevant consequences of **MPT** is the follow-  
 262 ing:

263 **NORMATIVE WEAKNESS OF NON-MORAL OBLIGATIONS.** Nobody  
 264 has to avoid violating with all her strength those non-moral obliga-  
 265 tions that do not relate appropriately to any of her moral obligations  
 266 and if someone fails to observe such an obligation, she should not  
 267 feel remorse.

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

275 **RESPONSIBILITY IDENTITY THESIS (RIT).** The property of being  
 276 morally responsible is identical to the property of fulfilling all con-  
 277 ditions for bearing moral obligations.

278 **RIT** makes moral responsibility into a precondition of having moral obliga-  
 279 tions, and it also makes the former prior to the latter in a certain respect. Since  
 280 moral obligations are, in turn, prior to any other type of obligations under  
 281 **MPT**, we will call the combination of **MPT** and **RIT** the Responsibility First  
 282 View (RFV).

## 285 5 The Primacy Argument

284 We can now turn to the revision of the **SIMPLE OBJECTION**. The proponent of  
 285 **RFV** can answer the agnostic's challenge as follows: "You claim that I should  
 286 suspend my belief in moral responsibility because I cannot prove that anyone  
 287 meets the conditions for being morally responsible. However, based on my  
 288 conception of morality and responsibility, if I am not morally responsible, then  
 289 I do not have any moral obligations. And if I have no moral obligations, I do not  
 290 have any obligations that I should fulfill with all my strength, any obligations  
 291 that should seriously concern me. In technical terms, I do not have all things  
 292 considered obligations. So if I am not morally responsible, then it is all right  
 293 for me to disregard your demand about suspension of belief. And in case you  
 294 claimed that I have an all things considered obligation to suspend my belief,  
 295 an obligation which I cannot disregard without committing a normative fault  
 296 I should regret, then I conclude on the basis of my conception of responsibility  
 297 that I am a morally responsible being after all. Either way, I can rationally  
 298 resist your challenge and keep believing in moral responsibility."

299 We would like to express this revised version of the **SIMPLE OBJECTION**  
 300 more formally:

### 301 PRIMACY ARGUMENT.

- 302 (PA1) No agent can rationally think that she has an all things considered  
 303 obligation to suspend her belief that she fulfills the necessary conditions  
 304 of having all things considered obligations.
- 305 (PA2) Someone who rationally upholds **RFV** cannot rationally think that she  
 306 has an all things considered obligation to suspend her belief that she is  
 307 morally responsible.
- 308 (PA3) If someone cannot rationally think that she has an all things considered  
 309 obligation to suspend her belief that she is morally responsible, then  
 310 she is rational to reject agnosticism about moral responsibility.
- 311 (PAC) Rejecting agnosticism about moral responsibility is rational for anyone  
 312 who rationally upholds **RFV**.

313 Until the agnostic does not challenge the moral framework that proponents of  
 314 **RFV** employ, she cannot undermine their belief in their own moral responsi-  
 315 bility. The agnostic cannot challenge belief in moral responsibility by merely  
 316 pointing out that evidence for the existence of moral responsibility is insuf-  
 317 ficient. What is more, if someone upholds **RFV** rationally, then it would be

318 straightforward irrational for her to accept the agnostic's conclusion—unless  
319 she finds out that her own moral framework is untenable.

320 The proponent of **RFV** gains a huge dialectical advantage by deploying the  
321 **PRIMACY ARGUMENT**. Due to the **PRIMACY ARGUMENT**, the debate shifts  
322 from the sufficiency of evidence to the tenability of a specific moral framework.  
323 Defending the tenability of **RFV** seems to be much easier than defending  
324 the sufficiency of evidence regarding the existence of moral responsibility.  
325 Especially so if the proponent of the **MORAL PRIMACY THESIS**, by investigating  
326 the nature of moral obligation and responsibility, comes to the conclusion  
327 that moral responsibility has heavy-weight metaphysical preconditions such  
328 as libertarian free will, since scientific evidence for libertarian freedom seems  
329 to be lacking. (We will say more about free will in the last section.)

330 Moreover, as far as we can tell, both **MPT** and **RIT** can be supported by  
331 considerable arguments. Even though some consequentialists deny that moral  
332 obligation implies moral responsibility, that principle seems to be fundamental  
333 and obvious for almost everyone—as even consequentialist critics note (**Waller**  
334 **2004, 427–428**). And someone who upholds **RIT** can explain why that principle  
335 is true: being a morally responsible agent is the same as being a potential  
336 bearer of moral obligations.

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

353 Of course, anyone, including the agnostic, can argue against **RFV**. Indeed,  
354 it seems that one can find not only prominent supporters of **RFV** (Kant seems  
355 to be the most obvious example) but able critics too. For instance, Bernard

356 Williams criticizes an ethical system under the label “morality” that contains,  
357 among other things, *RFV* (see Williams 2006, 174–196), because he believes  
358 that moral systems with such a strong notion of moral obligations threaten  
359 personal integrity. Even though the investigation of such counterarguments  
360 that are based on such wide-ranging considerations about the relation between  
361 a whole system of morality and other basic values is out of the scope of our  
362 paper, we can deal with another argument against *RFV* that is rather closely  
363 related to the problem of moral responsibility.

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

381 To be clear, *RFV* does not imply that epistemic obligations as such depend  
382 on (or are identical with) moral obligations. *RFV* does not exclude that they  
383 are totally unrelated to each other. What *RFV* implies is only that an epistemic  
384 obligation has to be appropriately related to some moral obligations in order  
385 to be true that agents have to avoid violating it with all their strength and  
386 if someone fails to observe an epistemic obligation which does not relate  
387 appropriately to any moral obligation, she should not feel remorse. In other  
388 words, in themselves, epistemic obligations do not have sufficient normative  
389 weight to constitute all things considered obligations. So, if one both accepts  
390 *RFV* and rejects moral responsibility based on her evidence-basis, she cannot  
391 rationally believe that there are moral and all things considered obligations,

---

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

392 but she still can rationally think that there are (weak) epistemic obligations.  
393 What she cannot rationally believe is that neglecting any epistemic obligation  
394 cannot be OK in the same way as neglecting the prudential obligation not to  
395 ridicule oneself. That is, even if one accepts **RFV** and, for instance, Pereboom's  
396 assessment of the evidence about free will and moral responsibility, she can  
397 rationally deny the existence of (a metaphysically rather extreme kind of)  
398 free will, moral responsibility, moral obligations, and all things considered  
399 obligations. The only thing that she cannot rationally maintain without re-  
400 jecting **RFV** is the idea that anyone should suspend belief in those things to  
401 avoid committing an unacceptable normative error that cannot be shrugged  
402 off. In other words, if there is a proponent of **RFV** who tries to heroically  
403 find the truth no matter the cost and finds that her evidence-basis strongly  
404 indicates the non-existence of moral responsibility, she can rationally believe  
405 that she has an epistemic (or even prudential) obligation to deny the existence  
406 of moral responsibility, but she cannot rationally think either that she has an  
407 all things considered obligation to reject moral responsibility or that anyone  
408 has an all things considered obligation to try to find the truth no matter the  
409 cost.

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

419 Note, parenthetically, that one can construct a modified version of the  
420 **PRIMACY ARGUMENT** even if both **MPT** and **RIT** are untenable. One need  
421 not appeal to morality (or moral responsibility) at all. Anyone who rationally  
422 believes in all things considered obligations has the epistemic right to sustain  
423 belief in a specific kind of responsibility. As the first premise of the **PRIMACY**  
424 **ARGUMENT** says, no agent can rationally think she has an all things consid-  
425 ered obligation to suspend the belief that she fulfills the necessary conditions  
426 of having all things considered obligations. In other words, someone who ra-  
427 tionally attributes all things considered obligations to herself must also accept  
428 that she fulfills the necessary conditions of having all things considered obliga-  
429 tions. Since having strong obligations is one of those necessary conditions, the

430 agent in question must also accept that she fulfills the necessary conditions  
431 of having strong obligations. It is reasonable to say that being responsible “in  
432 a strong sense” requires fulfilling all necessary conditions for bearing strong  
433 obligations, so anyone who rationally attributes all things considered obliga-  
434 tions to herself can rationally attribute “strong responsibility” to herself as  
435 well. It seems that this argument for “strong responsibility” can be threatened  
436 only by arguments against the existence of all things considered obligations.

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

## 446 **6 Free Will, Fairness, and Others**

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

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

459 However, there is another prevalent conception, according to which free  
460 will is the ability to do otherwise (van Inwagen 1983). Our argument can be  
461 extended to this case as well, but only if one rationally upholds that the ability  
462 to do otherwise is a necessary precondition of being morally responsible. Given  
463 strong evidence that moral responsibility depends on free will of the second  
464 sort, then rational belief in **RFV** (together with the evidence in question) can  
465 ground rational belief in the existence of free will. And if the proponent of



466 RFV has sufficient evidence that moral responsibility has further metaphysical  
467 conditions, she can also rationally believe that she fulfills all those further  
468 conditions, regardless of how demanding they are metaphysically.

469 These possible extensions of the **PRIMACY ARGUMENT** are especially sig-  
470 nificant if one takes into account that many philosophers and scientists insist  
471 that there is no sufficient scientific or other evidence for macro-level psycho-  
472 logical indeterminism (which is a precondition of libertarian free will) or  
473 the presence of compatibilist-friendly causal determinism in the brain. In  
474 light of the possible extensions of the **PRIMACY ARGUMENT**, proponents of  
475 RFV can rationally believe in responsibility-relevant free will (of either the  
476 incompatibilist or compatibilist sort) even in the absence of sufficient direct  
477 scientific or other evidence.

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

492 Proponents of the **PRIMACY ARGUMENT** evidently disagree, as their sup-  
493 posedly rational belief in RFV makes them rational in holding that they can  
494 be morally responsible for their actions. It is important that Sehon brings up  
495 this issue in terms of fairness. The obligation to be fair with others is naturally  
496 understood to be a moral obligation, and hence a strong one—according to  
497 MPT. Those who uphold RFV will see the situation as follows. The obligation  
498 to be fair can only be attributed to morally responsible persons. If nobody is  
499 morally responsible, then the strong obligation to be fair cannot be attributed  
500 to anybody. And if that is the case, then nobody has to care about being fair  
501 to anybody. So if the proponent of RFV takes Sehon’s exhortation to be fair  
502 seriously, and if she thinks she has to care about it, then, in the light of MPT,  
503 she incidentally attributes a strong obligation to herself. As a result, she im-

504 plies that she fulfills all the conditions of having strong obligations, including  
505 having moral responsibility. That is, for proponents of RFV, Sehon's point can  
506 only have force if they take themselves to be morally responsible. They would  
507 need to assume, first, that they are morally responsible, and, second, they  
508 would have to suspend judgment about the existence of moral responsibility  
509 because of that very assumption—which seems incoherent. Thus, the argu-  
510 ment that insufficient direct evidence for moral responsibility should make us  
511 suspend our belief in moral responsibility because it might lead to the unfair  
512 treatment of others makes no sense to those who hold RFV to be true.

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

528 Extending this rather small circle of responsible agents might look unrea-  
529 sonable or unfair indeed. However, there could be ways to do it. Remember  
530 that the agent in question reasonably believes in her own responsibility. If she  
531 considers agents who seem to be like her in every relevant respect, she may  
532 take them to be morally responsible as well. Nevertheless, the reasonableness  
533 of this move depends on two crucial factors. First, the agent must have a  
534 rationally held theory of what the relevant respects are. Second, were she to  
535 deem morally responsible any agent other than herself, her judgment that that  
536 person is similar to herself in every relevant respect must also be rational. This  
537 means that reasonably extending the circle of morally responsible agents to  
538 others is logically possible, but could be difficult in practice. Fortunately, there  
539 might be an easier way. It seems reasonable to think that all fully developed  
540 human beings have the same metaphysical structure. Insofar as this assump-  
541 tion 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 *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

---

\* 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.

Tamás Paár

 0000-0003-0585-6388

Central European University, Budapest

paar.tamas@gmail.com

## References

- BYRD, Jeremy. 2010. "Agnosticism about Moral Responsibility." *Canadian Journal of Philosophy* 40(3): 411–432, doi:[10.1080/00455091.2010.10716729](https://doi.org/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](https://doi.org/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](https://doi.org/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 C. LYONS, pp. 78–98. Malden, Massachusetts: Wiley-Blackwell, doi:[10.1111/phis.12066](https://doi.org/10.1111/phis.12066).
- FELDMAN, Richard H. 1988. "Epistemic Obligations." in *Philosophical Perspectives 2: Epistemology*, edited by James E. TOMBERLIN, pp. 235–256. Atascadero, California: Ridgeview Publishing Co., doi:[10.2307/2214076](https://doi.org/10.2307/2214076).
- GUILLOIN, Jean-Baptiste. 2014. "Van Inwagen on Introspected Freedom." *Philosophical Studies* 168(3): 645–663, doi:[10.1007/s11098-013-0159-x](https://doi.org/10.1007/s11098-013-0159-x).
- VAN INWAGEN, Peter. 1983. *An Essay on Free Will*. Oxford: Oxford University Press.
- KEARNS, Stephen. 2015. "Free Will Agnosticism." *Noûs* 49(2): 235–252, doi:[10.1111/nous.12032](https://doi.org/10.1111/nous.12032).
- LATHAM, Andrew James. 2019. "The Conceptual Impossibility of Free Will Error Theory." *European Journal of Analytic Philosophy* 15(2): 99–120, doi:[10.31820/eja.p.15.2.5](https://doi.org/10.31820/eja.p.15.2.5).
- LEVY, Neil. 2011. *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*. Oxford: Oxford University Press, doi:[10.1093/acprof:oso/9780199601387.001.0001](https://doi.org/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, doi:[10.5040/9781350044180](https://doi.org/10.5040/9781350044180).
- 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](https://doi.org/10.4324/9780429506048-2).
- PEREBOOM, Derk. 2001. *Living Without Free Will*. Cambridge: Cambridge University Press, doi:[10.1017/cbo9780511498824](https://doi.org/10.1017/cbo9780511498824).
- ROSS, William David. 1930. *The Right and the Good*. Oxford: Oxford University Press.

- 603 RUSSELL, Bruce. 2001. "Epistemic and Moral Duty." in *Knowledge, Truth, and Duty:*  
604 *Essays on Epistemic Justification, Responsibility, and Virtue*, edited by Matthias  
605 STEUP, pp. 34–48. Oxford: Oxford University Press, doi:10.1093/0195128923.003.0  
606 003.
- 607 SEARLE, John R. 1978. "Prima Facie Obligations." in *Practical Reasoning*, edited by  
608 Joseph RAZ, pp. 81–90. Oxford: Oxford University Press. Reprinted in van Straaten  
609 (1980, 238–259).
- 610 SEHON, Scott R. 2013. "Epistemic Issues in the Free Will Debates: Can we Know when  
611 We Are Free?" *Philosophical Studies* 166(2): 363–380, doi:10.1007/s11098-012-  
612 0044-z.
- 613 —. 2016. *Free Will and Action Explanation: A Non-Causal, Compatibilist Account*. Oxford:  
614 Oxford University Press, doi:10.1093/acprof:oso/9780198758495.001.0001.
- 615 SINNOTT-ARMSTRONG, Walter. 2008. "Framing Moral Intuitions." in *Moral Psychology,*  
616 *Volume 2. The Cognitive Science of Morality: Intuition and Diversity*, edited by  
617 Walter SINNOTT-ARMSTRONG, pp. 47–76. Cambridge, Massachusetts: The MIT  
618 Press, doi:10.7551/mitpress/7573.003.0004.
- 619 SMILANSKY, Saul. 1994. "The Ethical Advantages of Hard Determinism." *Philosophy*  
620 *and Phenomenological Research* 54(2): 355–363, doi:10.2307/2108494.
- 621 VAN STRAATEN, Zak, ed. 1980. *Philosophical Subjects: Essays Presented to P.F. Strawson*.  
622 Oxford: Oxford University Press.
- 623 STRAWSON, Galen. 1994. "The Impossibility of Moral Responsibility." *Philosophical*  
624 *Studies* 75(1/2): 5–24. Reprinted as "The Impossibility of Ultimate Moral Respon-  
625 sibility" in Strawson, G. (2008, 319–336), doi:10.1007/bf00989879.
- 626 —. 2008. *Real Materialism, and Other Essays*. Oxford: Oxford University Press, doi:10  
627 .1093/acprof:oso/9780199267422.001.0001.
- 628 STRAWSON, Peter Frederick. 1969. *Meaning and Truth*. Oxford: Oxford University Press.  
629 An Inaugural Lecture Delivered Before the University of Oxford on 5 November  
630 1969. Reprinted in Strawson, P.F. (1971, 170–189) and in Strawson, P.F. (2004,  
631 131–146).
- 632 —. 1971. *Logico-Linguistic Papers*. London: Methuen & Co. Reprinted as Strawson, P.F.  
633 (2004), doi:10.4324/9781351153607.
- 634 —. 2004. *Logico-Linguistic Papers*. 2nd ed. London: Routledge, doi:10.4324/97813152  
635 50250.
- 636 TRAKAKIS, Nick N. 2007. "Whither Morality in a Hard Determinist World?" *Sorites*  
637 19(1): 14–40, [http://lorenzopena.es/Sorites/Issue\\_19/sorites19.pdf](http://lorenzopena.es/Sorites/Issue_19/sorites19.pdf).
- 638 WALLER, Bruce N. 2004. "Virtue Unrewarded: Morality Without Moral Responsibility."  
639 *Philosophia: Philosophical Quarterly of Israel* 31(3/4): 427–447, doi:10.1007/bf0238  
640 5194.
- 641 WILLIAMS, Bernard Arthur Owen. 1985. *Ethics and the Limits of Philosophy*. London:  
642 Fontana Press.

- 643 —. 2006. *Ethics and the Limits of Philosophy*. London: Routledge. With a commentary  
644 on the text by A.W. Moore; original publication: Williams (1985), doi:[10.4324/97](https://doi.org/10.4324/9780203969847)  
645 [80203969847](https://doi.org/10.4324/9780203969847).
- 646 WITTGENSTEIN, Ludwig. 1965. "A Lecture on Ethics." *The Philosophical Review* 74(1):  
647 3–12, doi:[10.2307/2183526](https://doi.org/10.2307/2183526).

PROOF