

# Reforming Responsibility Practices without Skepticism

Marcelo Fischborn

Farroupilha Federal Institute of Education, Science, and Technology, Uruguiana, Brazil

The Version of Record of this manuscript has been published and is available in Philosophical Psychology <June 21, 2022> <https://www.tandfonline.com/doi/full/10.1080/09515089.2022.2090913>

## Abstract:

Derk Pereboom and Gregg Caruso argue that humans are never morally responsible for their actions and they take that thesis as a starting point for a project whose ultimate goal is the reform of responsibility practices, which include expressions of praise, blame, and the institution of legal punishment. This paper shares the skeptical concern that current responsibility practices can be suboptimal and in need of change, but argues that a non-skeptical pursuit of those changes is viable and more promising. The main lines of the argument are developed by assessing the prospects of implementing one of the changes favored within the skeptical project (namely, the reduction of punishment severity) in light of how human moral psychology works. An original vignette experiment (N = 180; participants from Facebook groups related to Brazilian universities) asked participants to recommend sentences for a fictitious criminal after considering alternatives to regular punishment that varied in their effectiveness to prevent reoffending. The results suggest that people can become less punitive even if they continue to believe in moral responsibility and free will. The paper further argues that reforming responsibility practices is more likely to occur without the endorsement of skepticism.

Keywords: moral responsibility; responsibility practices; blame; punishment; moral psychology

“I would like to see retributivistic systems of legal punishment move toward consequentialism, but I do not believe that claiming that science has shown that free will is an illusion is a good means to this end.” (Mele, 2013, p. 189)

“This is in no way to deny the possibility and desirability of redirection and modification of our human attitudes in the light of these studies. But we may reasonably think it unlikely that our progressively greater understanding of certain aspects of ourselves will lead to the total disappearance of those aspects.” (Strawson, 1962, p. 170)

## 1. Introduction

Throughout the history of philosophy, most approaches to moral responsibility and free will have been mainly theoretical, focusing on metaphysical issues related to causation,

determinism, powers, and the like. Some recent approaches also include more practical concerns. One example is the skeptical view developed by Derk Pereboom and Gregg Caruso (see, e.g., Caruso, 2016; Caruso & Pereboom, 2020; Pereboom, 2001, 2014). In addition to traditional theoretical issues, Pereboom and Caruso also discuss how responsibility practices—which include responses such as blame, praise, and punishment—could be reformed in light of their skeptical view. Even though skeptics have taken the lead in developing reformative proposals, however, the assumption that responsibility practices are sometimes defective and in need of change does not require skepticism. As the quotes from Alfred Mele and P. F. Strawson in the epigraph suggest, the reform of responsibility practices is also consistent with the affirmation of moral responsibility and free will.

This paper assesses the prospects of skeptical and non-skeptical reform. In contrast to more standard theoretical views, the assessment of reformative proposals needs to consider not just whether the views endorsed are true or false, but also the empirical viability of implementing proposed changes. Here I present results from an original experiment that support the viability of non-skeptical reform. I also argue, by taking into account additional studies that are in tension with the skeptical view, that non-skeptical reform is, at least so far, more promising. In the remainder of the paper, section 2 describes the skeptical project in more detail and identifies a move toward less severe punishment in the legal sphere as one of the goals of skeptical reform. Section 3 reports a short experimental study about folk judgments and attitudes about responsibility practices, including punishment recommendations, whose results indicate that people can recommend less severe punishments even if they preserve their (naturally strong) beliefs in moral responsibility and free will. Section 4 discusses the implications and limitations of the experimental results in light of some other studies and briefly compares the viability of skeptical and non-skeptical reform. The conclusion (section 5) is that human moral psychology is significantly hostile to skepticism, even though it leaves room for reforming responsibility practices in a way that is not premised on the rejection of moral responsibility.

## **2. Skepticism about moral responsibility and skeptical reform**

The kind of skepticism about moral responsibility to be discussed here was introduced by Pereboom (Pereboom, 2001, 2014) and later developed by Caruso (Caruso, 2013, 2016; Caruso & Morris, 2017) and by both authors in collaboration (Caruso & Pereboom, 2020; Pereboom & Caruso, 2018). Skepticism as endorsed by Pereboom and Caruso denies that people have moral responsibility in a specific sense, which they call ‘basic desert’ moral responsibility. Here is how Pereboom defines the notion:

in my view, for an agent to be morally responsible for an action is for this action to belong to the agent in such a way that she would deserve blame if the action were morally wrong, and she would deserve credit or perhaps praise if it were morally exemplary. The desert at issue here is basic in the sense that the agent, to be morally responsible, would deserve the blame or credit just by virtue of having

performed the action, and not, for example, by way of consequentialist considerations. (Pereboom, 2001, p. xx)<sup>1</sup>

The sense of moral responsibility Pereboom focuses on is characterized by what it arises from. That justification is entirely backward-looking: in order to be morally responsible in the basic desert sense, praise or blame should be deserved “just by virtue” of what the agent did.

Pereboom and Caruso offer two related reasons for focusing on moral responsibility in the basic desert sense. One of them has to do with what Pereboom takes to be at stake in the traditional debate about the (in)compatibility of free will and determinism. Pereboom considers the views of some philosophers who “identify themselves as compatibilists because they hold that some non-basic-desert notion of moral responsibility [...] is compatible with determinism”<sup>2</sup> and he objects that “if ‘compatibilism’ is defined so that such a position turns out to be compatibilist, virtually everyone in the debate stands to be a compatibilist.” (Pereboom, 2014, p. 2). A second reason seems to derive from a willingness to consider a sense of moral responsibility that is required for the justification of punishment according to certain retributivist theories. Caruso says that moral responsibility in the sense denied by his skeptical view is “the kind required for retributivism” (2016, p. 27). Caruso and Pereboom (2020, p. 356) again acknowledge the possibility of “non-basically deserved blaming and praising—for example, blaming that invokes desert grounded in consequentialist [...] or contractualist [...] consideration”, but leave it aside by saying that “punishment justified in this way would not be genuinely retributivist, since its ultimate justification would be consequentialist, and this is incompatible with retributivism as it has traditionally been understood.”.

Understanding the skeptics’ restricted focus on the notion of basic desert is critical because both Pereboom and Caruso acknowledge that there are other senses in which an agent can be morally responsible and that those senses are not affected by their views. So, for all the skeptics say, agents may still be morally responsible in senses that do not involve basic desert.<sup>3</sup> Yet, moral responsibility in the basic desert sense is claimed to be at the basis of some important aspects of responsibility practices: “A belief that an agent is morally responsible in this sense at least typically accompanies expressions of reactive attitudes such as moral resentment and indignation, and it is thus closely related to the notion of moral responsibility that P. F. Strawson brings to the fore” (Pereboom, 2014, p. 2). Therefore, despite its restricted scope, skepticism about basic desert has implications that need to be identified and assessed. In particular, Pereboom and Caruso have argued that their views entail that central aspects of

---

1 Pereboom (2014, p. 2) offers a slightly different definition: “The desert at issue here is basic in the sense that the agent would deserve to be blamed or praised *just* because she has performed the action, given an understanding of its moral status, and not, for example, *merely* by virtue of consequentialist or contractualist considerations.”.

2 The authors Pereboom refers to here are Frank Jackson, Daniel Dennett, and Manuel Vargas.

3 Pereboom (2014, p. 2) says: “there are other senses of moral responsibility that are not at issue in the free will debate, and are not challenged by skeptical arguments about free will, and the legitimacy of some of them is an important feature of my position”.

responsibility practices are unjustified and need to be either reformed or abandoned. In this paper, I focus on the changes they propose for the criminal justice system.

Pereboom and Caruso have been elaborating the “public health-quarantine model”, which is “a non-retributive and indeed non-punitive alternative for addressing criminal behavior [...] which draws on the public health framework and prioritizes prevention and social justice” (Caruso & Pereboom, 2020, p. 363). The model has two parts. Concerning one part, the quarantine analogy, Pereboom says that

a theory of crime prevention that would be acceptable whether or not the skeptic is right can be developed by analogy with our rationale for quarantining carriers of dangerous diseases. The core idea is that the right to harm in self-defense and defense of others justifies incapacitating the criminally dangerous with the minimum harm required for adequate protection. The resulting account would not justify the sort of criminal punishment whose legitimacy is most dubious, such as death or confinement in the most common kinds of prisons in our society. More than this, it demands a certain level of care and attention to the well-being of criminals which would change much of current policy. (Pereboom, 2014, pp. 173–174)

The quarantine analogy says that society has a right to defend itself from those who pose serious risks to others by detaining them, just as it has a right to quarantine those who possess a serious contagious disease—and this may be so even if neither the criminal nor the diseased deserve to be set apart from others. The quarantine analogy also entails some significant departures from current punitive practices because quarantine requires a concern for the well-being of the diseased, while punishment is traditionally understood as the intentional imposition of something burdensome. The second part of the model, the public health framework, stresses the need to address the causes of criminality, many of which are claimed to be the same as the causes of health issues. Caruso (2016, p. 34) mentions “social injustice, poverty, systematic disadvantage, mental health issues, and addiction” as examples of health and social issues that also significantly increase the chances that someone will pose a threat to the safety of others (see also Caruso, 2017).

The quote from Pereboom in the previous paragraph seems to contain a tension between two points: it says both that the theory of crime prevention would be acceptable “whether or not” skepticism is right and that its implementation “would change much of current policy”. The tension can be eliminated by interpreting each of the claims in a more specific way. A quarantine policy based on the right to self-defense is independent from skepticism insofar as it is *not* meant to replace current criminal justice systems. Preventive detention, whether or not it can actually be made into an acceptable public policy, is something different from current punishment policies and, indeed, both types of policies could in principle be implemented simultaneously.<sup>4</sup> Insofar as having actually committed a crime is not a necessary condition for the attribution of criminal risk, some individuals who

---

4 Similarly, the public health approach described by Caruso can also be implemented alongside current practices to a large extent, particularly in those aspects that seek to promote health and social conditions before someone has posed any risk to society.

were never addressed by the criminal justice system might end up subjected to preventive detention. What would impact current punishment practices is the skeptical goal of *replacing* current punishment systems by a system of preventive detention. Specifically, there would be a change in the way those who actually commit crimes are treated because they would receive much less severe responses from the state. Therefore, one specific change in the criminal justice system that skeptics advocate is the reduction of punishment severity.

In sum, the kind of skepticism Pereboom and Caruso advocate denies the existence of free will and moral responsibility in a specific sense that they take to be required if responses such as legal punishment are ever to be truly deserved by their targets. But that skepticism does not directly translate into practical recommendations about how things such as criminal behavior should be dealt with. Pereboom and Caruso argue that they can consistently endorse some (non-punitive) preventive measures, which on their own, I argued, do not conceptually conflict with preserving current punitive practices. However, skeptics may also argue that those preventive measures should *replace* current punitive practices, which is a more substantive reformative proposal concerning responsibility practices. The implementation of that more substantive proposal involves treating criminal offenders less severely than they are treated within current criminal justice systems. My discussion of the reform of responsibility practices in this paper focuses on the more substantive proposal.

In order for a theory recommending the reform of responsibility practices to be successful, the changes it proposes need to be not just acceptable (something skeptics and non-skeptics are bound to disagree about) but also *implementable*. The implementability of a proposed change is an empirical matter, as Pereboom (2014, p. 130) himself acknowledges when he says that “[p]hilosophical method is not suited for determining whether living with [a skeptic] conception would be viable for us”. Hence, the question arises as to whether, and how, a reduction of punishment severity can be implemented.

I think the available evidence tells mostly *against* the viability of skeptical reform. On a general level, punitive systems have been present in human societies throughout history (Morris & Rothman, 1995; Stearns, 1936) and there are reasons to think a sense of desert is an evolved aspect of human moral psychology (Mameli, 2013; Tomasello, 2016, pp. 60–62). More specifically, researchers conducting empirical studies have reported, for example, that belief in free will and moral responsibility is strong and even hard to manipulate (Fischborn, 2018, pp. 50–51; see, e.g., Nadelhoffer et al., 2014, p. 38; Schooler et al., 2015). Also, there is a tendency for immoral behavior to be seen as more free (Clark et al., 2014) and for immoral outcomes to be more easily perceived as intentional (Knobe, 2003) than neutral or morally positive behavior and outcomes. It must be noted that the reality (replicability) and the strength of the effect of immoral behavior on beliefs about free will (in general or specifically related to the authors of immoral actions) hypothesized by Clark (2014) have been disputed (Monroe & Ysidron, 2021). Even so, a conservative assessment of the available results suggests that *general* beliefs about free will are sometimes at least slightly reinforced after exposure to immoral behavior (Monroe & Ysidron, 2021, results of study 3) and that the *authors* of immoral behavior are usually seen as more free than morally neutral behavior that

is not exceptional in some other aspect (Monroe & Ysidron, 2021, results of studies 2a and 3).<sup>5</sup> Thus, there is not just a robust belief in moral responsibility generally in place, but it is also likely to be reinforced every time a response from the criminal justice system (or any alternative system) is called for.

I take the studies just mentioned to provide at least *prima facie* evidence against the implementability of a skeptical reform of responsibility practices. But the fact that people strongly believe in moral responsibility and free will does not automatically support the viability of non-skeptical *reform*. It might well be that people who believe in moral responsibility and free will want to keep not just their beliefs, but also their practices. For that reason, it is also worth exploring the viability of reformative proposals that are not motivated by a denial of moral responsibility and free will, such as the ones suggested by Mele and Strawson. The remainder of the paper presents and discusses the results from an empirical study that supports the hypothesis that a reduction of punishment severity can be achieved without skepticism.

### **3. An experiment on folk attitudes about moral responsibility and punishment**

If skeptical reform were to be implemented in a very direct way, one could expect that convincing people that no one is morally responsible in the basic desert sense would make them support less severe forms of punishment or even non-punitive responses to crime. In order to assess the viability of *non-skeptical* reform, the general goal of the present study was to investigate how people respond to an incentive to be less punitive that is unrelated to skepticism. If people can support less severe forms of punishment without reducing their naturally strong belief in moral responsibility, that counts as some evidence of the viability of non-skeptical reform.

The design of the study took into account two main points from previous studies. First, previous research revealed correlation and association between support for punishment and belief in the efficacy of punishment for crime prevention (Miller & Vidmar, 1981; Nadelhoffer et al., 2014, table 1, item 12; Thomas & Cage, 1974). Hence, one specific goal of the present study was to test the prediction that manipulating information about the effects of punishment on reoffending would causally affect the severity of punishment recommendations. Second, Clark et al. (2014, p. 504) suggested, based on a mediation analysis, that the desire to punish causally influences belief in free will (for disputes, see Monroe & Ysidron (2021); but see also Clark et al. (2021)). The hypothesis that the desire to punish causally affects belief in free will (or moral responsibility) challenges the viability of non-skeptical reform: even if people could recommend less severe punishment when presented with information that is conceptually unrelated to skepticism, that might still causally make them less confident in free will and moral responsibility, which would amount

---

5 This is not to say that Clark et al.'s (2014) motivated account of belief in free will is devoid of problems. In addition to the observation that general and specific beliefs about free will tend to be stronger after exposure to immoral behavior, Clark et al. also argue that the explanation of the observation resides in the stronger desire to punish that arises when one is exposed to immoral behavior. This issue is further addressed in subsequent sections of this paper (see also Clark et al., 2021; Monroe & Ysidron, 2021).

to skepticism entering through the backdoor. Thus, the present study included measures of beliefs about free will and moral responsibility in order to monitor whether they would be affected by the attempt to manipulate the desire to punish.

Given the considerations above, the present experiment was designed to test the effects of beliefs about the effects of punishment on the desire to punish and the effects of the desire to punish on beliefs about free will and moral responsibility. The experiment included three conditions, all of which began with the description of a robbery. Then, Condition A (Less Effective Punishment) stated that the author of the robbery would have a very high chance of reoffending if subjected to traditional punishment, but a very low chance of doing so if subjected exclusively to an alternative treatment program. Condition B (Similar Efficacy) described a situation in which both punishment and a treatment program would have a similar weak/moderate effect in preventing reoffending. Condition C (Punishment Only) was a control condition providing no information about the effects of punishment nor about alternative programs. The main hypothesis before running the experiment was that the desire to punish would be weaker in A than in C. Another interest the study originally had was to test whether, as a consequence of a change in the disposition to punish the offender, belief in free will would also be weaker in A than in C, in line with Clark et al.'s (2014, pp. 504–505) hypothesis. However, as I will further discuss in what follows, the studies previously mentioned (Clark et al., 2021; Monroe & Ysidron, 2021), which were published after the present study was run, reveal that the present sample size was insufficient to properly address this issue. Condition B was designed to check whether either punishment or treatment would be preferred when both were available and thought to have similar effects on recidivism; it was also less artificial because actual interventions to prevent criminal behavior are unlikely to be as effective as suggested in condition A.

### **3.1 Method**

#### *Participants*

Taking into account results from Clark et al. (2014, study 2), the sample size for the present study was fixed at 60 valid responses per condition ( $N = 180$ ). Thus, the study had an 80% power to detect an effect size of  $\eta^2 = 0.052$ . That means the present study was able to detect, for example, variations in general beliefs about free will similar to those reported in Clark et al.'s study 2 ( $d = 0.47$ ), but not smaller variations such as those found in some subsequent studies (e.g.,  $\eta^2 = 0.009$  in one of the analyzes in Monroe & Ysidron's, 2021). Desired responses were obtained by inviting participants from Facebook groups related to Brazilian universities. Participants were balanced on sex (51% male) and from 15 different Brazilian states (27% from Rio Grande do Sul). Most of them were young adults (58% were 21–30 years-old) who had attended undergraduate-level education (70%) in several fields (22% from Exact and Earth Sciences).

#### *Procedure*

The study was conducted online. Participants were told that the study was about the relation between philosophical beliefs and social attitudes. After giving informed consent, participants were randomly assigned to one of the three conditions. In all conditions, participants read

about a robbery committed by a 32-year-old man of initials M.C.D. M.C.D. was said to have used a gun to threaten a person who was preparing to leave a market's parking area in order to steal the victim's motorcycle. Conditions differed in the following aspect, which was designed to manipulate participants' desire to punish M.C.D.:

- Condition A (Less Effective Punishment): participants were told that, according to a group of experts, M.C.D. satisfied the conditions for participation in a social reintegration program that would make him very unlikely to repeat the crime (10%), but that the program would not be effective if accompanied by punishment (80% of chance of repeating the crime). Participants were asked whether M.C.D. should receive the treatment program.
- Condition B (Similar Efficacy): participants were told that, according to a group of experts, the social reintegration program and traditional punishment would have similar effects on the probability of reoffending (40% chance of repeating the crime either participating in the program or being punished). Participants were also asked whether M.C.D. should receive the treatment program.
- Condition C (Control): no alternative to punishment was mentioned.

Henceforth, all participants indicated the amount of punishment (in years of imprisonment) M.C.D. should receive (0 to 15 years; intervals of one year); this was used as the measure of desire to punish. In a 7-point Likert-scale (1 = totally disagree; 7 = totally agree), participants indicated whether M.C.D.'s action was free, and whether he was blameworthy and responsible for the action.<sup>6</sup> Participants also indicated their general beliefs about free will in a Brazilian version of the free will subscale of the Free Will Inventory (Nadelhoffer et al., 2014; Santin et al., 2018).<sup>7</sup> Finally, participants answered some demographic questions and were informed, upon conclusion, of the fictitious nature of the material they previously read.

### 3.2 Results

An analysis of variance revealed significant variation in the level of punishment considered appropriate across conditions,  $F(2, 177) = 16.28, p < 0.001, \eta^2 = 0.155$ . Post-hoc TukeyHSD comparisons indicated that punishment differed across all conditions ( $ps < 0.05$ ). Punishment in condition A ( $M = 3.25; SD = 4.89$ ) differed from condition B ( $M = 5.63, SD = 4.84$ ), and both differed from the control condition ( $M = 8.12, SD = 4.26$ )—see Figure 1.

Regarding the other variables, no significant difference was observed in the level of general belief in free will in each condition;  $F(2, 177) = 2.277, p = 0.106, \eta^2 = 0.025$ . Differences were also absent in specific belief in free will— $F(2, 177) = 0.61, p = 0.544, \eta^2 = 0.007$ —and in specific attributions of responsibility— $F(2, 177) = 1.838, p = 0.162, \eta^2 = 0.020$ —and blame— $F(2, 177) = 0.979, p = 0.378, \eta^2 = 0.011$ ; see also Figure 1. The treatment

6 Following Clark et al. (2014, p. 504), three statements assessed participants' beliefs about M.C.D.'s free will during the action: "M.C.D. exercised his free will during the robbery", "M.C.D. could have decided not to rob", and "M.C.D. decided to rob freely". The means for these three variables were used in subsequent analyses.

7 The use of the free will subscale alone may count as a limitation of the present study in that it was validated together with two other subscales that compose part 1 of the Free Will Inventory. I followed previous studies that have also sometimes used only some of the subscales (see, e.g., Monroe et al., 2017).



program was readily endorsed by most subjects when it was mentioned: 86.7% of participants in condition A, and 81.7% of participants in condition B said M.C.D should receive the treatment program. The difference between the proportions was not significant ( $\chi^2 = 0.250$ ; 95% confidence interval: -0.097, 0.197;  $p = 0.617$ ).

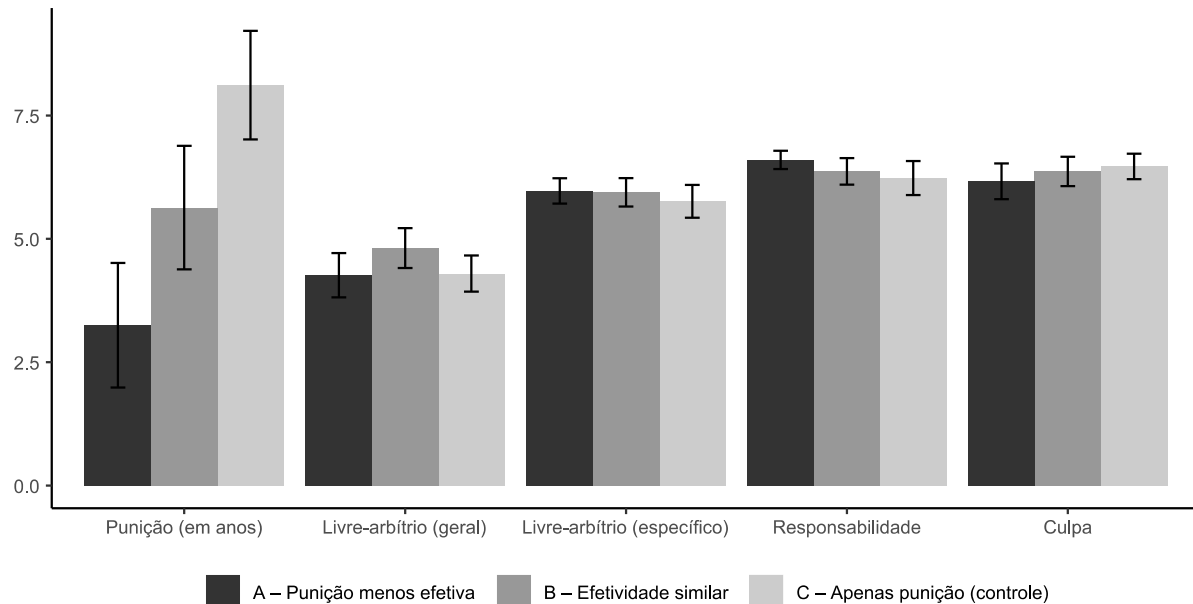


Figure 1 – Mean scores of dependent variables by condition. Error bars represent 95% confidence intervals. Note that the range for the punishment variable (0–15) was different from the other variables (1–7).

In sum, in the present study manipulations of belief in the efficacy of punishment relative to a treatment program led to significant changes in the desire to punish the fictitious criminal. Given the experimental design, one can conclude that considerations about the consequences of punishment causally influence the amount of punishment considered appropriate. At the same time, no variation was observed in general or specific beliefs about free will nor in attributions of responsibility and blame. The relevance of these findings is discussed in the next section in light of the overall goal of the paper.

#### 4. Discussion

The results just presented suggest that people can support milder forms of punishment without becoming skeptics about moral responsibility or free will. If that conclusion is right, it is evidence of the viability of non-skeptical reform. This section discusses the present results and some of their limitations in light of some additional studies.

As I mentioned earlier, Clark et al. (2014) found experimental evidence that people who read about immoral behavior have increased willingness to punish and stronger belief in free will than those who read about some morally neutral behavior. Because all of the conditions of the present study included the same robbery case, that finding alone provides no reason to assume a difference in beliefs about free will should emerge here as well. However, Clark et al. (2014, p. 504) also claim, on the basis of a mediation analysis, that “a heightened

desire to punish accounts for the heightened levels of both specific free will attributions and general free will belief". The present results run counter to that claim to some degree. When people's responses to a single type of immoral action are considered, the present study found that stronger desire to punish (as measured by the length of the sentence recommended) was not followed by stronger belief in free will, moral responsibility, or blame. Thus, the present results are consistent with the hypothesis that the variations in beliefs about free will found in Clark et al.'s study may be due to how morally bad (i.e., how serious) the actions considered are rather than simply to the intensity of the desire to punish. Under normal conditions, the desire to punish can be expected to be approximately proportional to the seriousness of the action considered. But when there is a reason to dissociate punishment from seriousness (as in conditions A and B of the present study) belief in free will may not follow the desire to punish.

The comments on Clark et al. (2014) just made need to be taken together with at least three cautionary notes, which reveal some of the limitations of the present study, ones that also apply to the studies discussed below. First, as mentioned earlier (see 3.1) more recent studies suggested that the relevant effect sizes are substantially smaller than those reported in Clark et al.'s original study, which renders the present study insufficiently powered to properly address the issue. For that reason, the present results do not exclude the possibility of smaller variations in beliefs about free will. Second, the present results also cannot exclude the possibility of cultural variation, in such a way that variations in beliefs about free will would be found if the present study were run with participants from the same cultural background of those taking part in Clark et al.'s studies. Even though the validation of the Free Will Inventory in a Brazilian sample provides at least some evidence of cultural similarity (Santin et al., 2018), the present study was not designed to make cross-cultural comparisons, especially about the relation between general belief in free will and attitudes regarding punishment. Finally, the present study did not assess the impact of the emotional intensity associated with different types of crimes on people's attitudes regarding punishment (see, e.g., Krueger et al., 2014). For that reason, it is also possible that the relation between beliefs about moral responsibility and free will, on the one hand, and punishment recommendations, on the other, could be different for other types of actions. That limitation notwithstanding, a large part of the prison population in Brazil was involved in crimes against property (approximately 45%), including robbery cases (Moura, 2019, p. 44), which adds to the ecological relevance of the present study.

Carlsmith, Darley, and Robinson's (2002) claim that laypeople's punishment recommendations are retributivistic rather than consequentialist also raises questions related to the present results. Their findings indicate that factors such as seriousness of the offense and presence of mitigating circumstances impact punishment decisions more than factors such as difficulty of detecting the offense and the publicity a given sentence will receive (288-289). The authors interpret factors of the first group as representative of retributivistic theories of punishment and factors of the second group as representative of deterrence theory. Despite their claim to the contrary (2002, p. 296), I think it is problematic to dissociate the seriousness of an offense from consequentialist theories of punishment. Under consequentialist theories, for example, how much harm you aim to prevent is critical for assessing how much harm you

are justified in employing as a means to that end. This is explicitly true even of Bentham's theory, which the authors take as a paradigm for defining deterrence theory (see, e.g., Bentham, 2007, Chapter XIV). In addition to this interpretive issue, the present results suggest that considerations related to *specific* deterrence (i.e., how likely the target of a punishment is to reoffend) do play a role in punishment decisions. This suggestion tells us nothing about how relevant those considerations are in comparison to other factors, be they associated with retributivism or not. But the suggestion is consistent with a more complicated picture of punishment recommendations according to which they can be affected by consequentialist considerations even if judgments of moral responsibility and blameworthiness are not similarly affected.

In contrast to the picture just described, Spitzley (2021) found that people take agents who are likely to repeat some morally significant kind of behavior to be more morally responsible than agents who are unlikely to repeat the behavior. According to one interpretation Spitzley offers, moral responsibility is not just about desert and "some forward-looking considerations affect the extent to which agents are morally responsible for their actions". The interpretation is, as stated, consistent with the present results. One way to conciliate the results of both studies is to say that some forward-looking considerations (like the ones considered in the present study) do not affect moral responsibility judgments, while some other forward-looking considerations (like to ones Spitzley considers) do. The conciliation would not work, however, if both studies involved the same kind of forward-looking consideration. Despite the fact that the present study also manipulated information about how likely an agent was of repeating a (criminal) behavior, that manipulation of likelihood was more uncontrolled than Spitzley's. For example, condition A informed participants that the criminal was *unlikely* to reoffend *if* subjected to the treatment offered, but very *likely* to reoffend if subjected to regular punishment. Thus, even though both studies considered the likelihood of future behavior, the present study addressed that in connection to an assessment of different *responses* to a crime rather than in connection to how the agent would behave in a more natural course of his life.

The discussion so far is relevant to assess the specifics of the present study and its results. Shariff et al. (2014), in contrast, present results that are more directly relevant for the overall goal of comparing the viability of skeptical and non-skeptical reformative proposals. They showed that people recommend less severe punishments after being exposed to manipulations that reduce belief in free will or that inform about neuroscientific findings. The part of their findings in which less severe punishment was recommended because of reduced belief in free will is evidence of the viability of skeptical reform. Given Shariff et al.'s and the present results, one could claim that skeptical and non-skeptical reform look equally viable. However, there are additional considerations that may favor the plausibility of non-skeptical reform. These considerations concern how easily results from controlled studies can transfer to more ordinary contexts. A first observation about the studies by Shariff et al. (2014) is that participants decided on how long an offender should stay in prison after spending two years in a rehabilitation facility. Similarly, conditions A and B of the present study mentioned a program that would interfere with the likelihood of reoffence. But participants in Shariff et al.'s experiments were exposed, in addition, to information against free will. In this regard, the

skeptical way to reform is more complex and demanding. A second observation relates to the robustness of belief in free will and moral responsibility (discussed in section 2), which are likely to provide a source of resistance against skeptical reform. Even if people can, under experimental conditions, have their beliefs in free will reduced (which does not mean they become skeptics about free will) other studies have attested that it is not easy for manipulations of free will beliefs to result in behavioral changes (Crone & Levy, 2019; Nadelhoffer et al., 2020). As such, even if many questions still remain to be investigated, a non-skeptical reform of punishment practices may be easier to achieve by not requiring an additional, potentially ineffective effort directed at reducing belief in moral responsibility and free will.

## **5. Conclusion**

Skeptics should be praised for adding to the philosophical agenda more practical questions about how to improve responsibility practices. But, in line with the suggestions by Mele and Strawson quoted at the beginning, the skepticism underlying their projects may be unnecessary or even detrimental to the pursuit of the relevant reform. Empirical studies, including the one reported here, show that human moral psychology includes a robust commitment to the view that human agents deserve some form of response for their morally significant behavior. When deciding how to respond, however, people can also be sensitive to considerations about the consequences of different types of responses. Hence, there is wide room for exploring, within a non-skeptical framework, alternative ways of responding to morally significant behavior (including behavior that is relevant for the criminal justice system) and also for articulating how one can combine considerations related to desert and consequences into a unified evaluation of responsibility practices. That kind of evaluative work is necessary for the identification of problematic aspects of current responsibility practices. Similarly, there is also room for empirical studies that continue to explore the effects of alternative ways of responding to morally significant behavior. That kind of study helps both to inform which alternatives are likely to represent improvements in relation to current practices and to assess which alternatives have better prospects of being actually implemented in human societies. To quote Strawson again, there are reasons to believe a “redirection and modification of our human attitudes in the light of these studies” is viable even if the common view of ourselves as free and morally responsible agents is left unchanged. There are reasons to believe we can reform responsibility practices without skepticism.

## **Acknowledgments**

I thank the Federal University of Santa Maria and especially Frank Thomas Sautter for supporting the realization of the experiment reported in this article. For comments and discussion of the whole or parts of earlier versions of the paper, I also thank the reviewers (for this and previous journals) as well as Rogério P. Severo, Gilberto Lourenço Gomes, Beatriz Sorrentino Marques, Róbson Ramos dos Reis, and the audiences of the V Conference of the

Brazilian Society for Analytic Philosophy (2018) and the VIII Meeting of the Anpof Workgroup on Philosophy of Neuroscience (2021).

### **Funding**

Financial support for the present work was provided by a doctoral fellowship sponsored by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and by a postdoctoral fellowship jointly sponsored by CAPES and the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS).

### **Supplemental online material**

Materials used in the present study are available at <https://osf.io/hrk3m/> .

### **ORCID**

Marcelo Fischborn: <https://orcid.org/0000-0002-0410-5251>

### **Contact information:**

Marcelo Fischborn – [marcelofischborn@gmail.com](mailto:marcelofischborn@gmail.com)

Instituto Federal Farroupilha, Rua Monteiro Lobato 4442, 97503-748, Uruguaiana, RS, Brazil.

### **Note on contributor:**

Marcelo Fischborn works at the Farroupilha Federal Institute of Education, Science, and Technology, in Uruguaiana, RS, Brazil, where he teaches philosophy mostly for high school students and does research on the practices of moral responsibility, free will, and moral psychology more broadly. He got his doctoral degree in Philosophy from the Federal University of Santa Maria, in Brazil, with a research visit to the Philosophy Department of the Florida State University sponsored by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and the Fulbright Program in Brazil.

### **References**

Bentham, J. (2007). *An Introduction to the principles of morals and legislation: The collected works of Jeremy Bentham*. Dover.

- Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology, 83*(2), 284–299.
- Caruso, G. D. (Ed.). (2013). *Exploring the illusion of free will and moral responsibility*. Lexington Books.
- Caruso, G. D. (2016). Free will skepticism and criminal behavior: A public health-quarantine model. *Southwest Philosophy Review, 32*(1), 25–48.
- Caruso, G. D. (2017). *Public health and safety: The social determinants of health and criminal behavior*. ResearchersLinks Books.
- Caruso, G. D., & Morris, S. G. (2017). Compatibilism and retributivist desert moral responsibility: On what is of central philosophical and practical importance. *Erkenntnis, 82*(4), 837–855.
- Caruso, G. D., & Pereboom, D. (2020). A non-punitive alternative to retributive punishment. In F. Focquaert, B. Waller, & E. Shaw (Eds.), *Routledge handbook of the philosophy and science of punishment* (pp. 355–365). Routledge.
- Clark, C. J., Luguri, J. B., Ditto, P. H., Knobe, J., Shariff, A. F., & Baumeister, R. F. (2014). Free to punish: A motivated account of free will belief. *Journal of Personality and Social Psychology, 106*(4), 501–513.
- Clark, C. J., Winegard, B. M., & Shariff, A. F. (2021). Motivated free will belief: The theory, new (preregistered) studies, and three meta-analyses. *Journal of Experimental Psychology. General, 150*(7), e22–e47. <https://doi.org/10.1037/xge0000993>
- Crone, D. L., & Levy, N. L. (2019). Are Free Will Believers Nicer People? (Four Studies Suggest Not). *Social Psychological and Personality Science, 10*(5), 612–619. <https://doi.org/10.1177/1948550618780732>
- Fischborn, M. (2018). How should free will skeptics pursue legal change? *Neuroethics, 11*(1), 47–54. <https://doi.org/10.1007/s12152-017-9333-8>
- Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis, 63*(279), 190–194. <https://doi.org/10.1111/1467-8284.00419>

- Krueger, F., Hoffman, M., Walter, H., & Grafman, J. (2014). An fMRI investigation of the effects of belief in free will on third-party punishment. *SCAN*, 9, 1143–1149.
- Mameli, M. (2013). Meat made us moral: A hypothesis on the nature and evolution of moral judgment. *Biology & Philosophy*, 28(6), 903–931. <https://doi.org/10.1007/s10539-013-9401-3>
- Mele, A. R. (2013). Free will, science, and punishment. In T. Nadelhoffer (Ed.), *The future of punishment* (pp. 177–191). Oxford University Press.
- Miller, D. T., & Vidmar, N. (1981). The justice motive in social behavior. In M. J. Lerner & S. C. Lerner (Eds.), *The justice motive in social behavior* (pp. 145–172). Springer.
- Monroe, A. E., Brady, G., & Malle, B. F. (2017). This isn't the free will worth looking for: General free will beliefs do not influence moral judgments; agent-specific choice ascriptions do. *Social Psychological and Personality Science*, 8(2), 191–199.
- Monroe, A. E., & Ysidron, D. W. (2021). Not so motivated after all? Three replication attempts and a theoretical challenge to a morally motivated belief in free will. *Journal of Experimental Psychology. General*, 150(1), e1–e12. <https://doi.org/10.1037/xge0000788>
- Morris, N., & Rothman, D. J. (1995). Introduction. In N. Morris & D. J. Rothman (Eds.), *The Oxford history of the prison* (pp. vii–xiv). Oxford University Press.
- Moura, M. V. (Ed.). (2019). *Levantamento nacional de informações penitenciárias, atualização junho de 2017*. Ministério da Justiça e Segurança Pública.
- Nadelhoffer, T., Shepard, J., Crone, D. L., Everett, J. A. C., Earp, B. D., & Levy, N. (2020). Does encouraging a belief in determinism increase cheating? Reconsidering the value of believing in free will. *Cognition*, 203, 104342. <https://doi.org/10.1016/j.cognition.2020.104342>
- Nadelhoffer, T., Shepard, J., Nahmias, E., Sripada, C., & Ross, L. T. (2014). The free will inventory: Measuring beliefs about agency and responsibility. *Consciousness and Cognition*, 25, 27–41.
- Pereboom, D. (2001). *Living without free will*. Cambridge University Press.

- Pereboom, D. (2014). *Free will, agency, and meaning in life*. Oxford University Press.
- Pereboom, D., & Caruso, G. D. (2018). Hard-Incompatibilism Existentialism: Neuroscience, Punishment, and Meaning in Life. In G. D. Caruso & O. Flanagan (Eds.), *Neuroexistentialism: Meaning, morals, and purpose in the age of neuroscience*. Oxford University Press.
- Santin, T. R., Vilanova, F., Costa, Â. B., Tocchetto, D. G., Nadelhoffer, T., & Koller, S. H. (2018). Evidências de validade do Inventário do Livre-Arbitrio (ILA) para a população brasileira. *Avaliação Psicológica*, 17(1), 1–11.  
<https://doi.org/10.15689/ap.2017.1701.01.13308>
- Schooler, J., Nadelhoffer, T., Nahmias, E., & Vohs, K. D. (2015). Measuring and manipulating beliefs and behaviors associated with free will: The good, the bad, and the ugly. In A. R. Mele (Ed.), *Surrounding freedom: Philosophy, psychology, neuroscience* (pp. 72–94). Oxford University Press.
- Shariff, A. F., Greene, J. D., Karremans, J. C., Luguri, J. B., Clark, C. J., Schooler, J. W., Baumeister, R. F., & Vohs, K. D. (2014). Free will and punishment: A mechanistic view of human nature reduces retribution. *Psychological Science*, 25(8), 1–8.  
<https://doi.org/10.1177/0956797614534693>
- Spitzley, J. (2021). The Future of Moral Responsibility and Desert. *Review of Philosophy and Psychology*, 12(4), 977–997. <https://doi.org/10.1007/s13164-020-00522-5>
- Stearns, A. W. (1936). Evolution of punishment. *Journal of Criminal Law and Criminology*, 27(2), 219–230.
- Strawson, P. F. (1962). Freedom and resentment. In D. Pereboom (Ed.), *Free will* (pp. 148–171). Hackett.
- Thomas, C. W., & Cage, R. J. (1974). Correlates of public attitudes toward legal sanctions. *Paper Presented at the Meeting of the Western Sociological and Anthropological Association, Banff, Canada*, 1–31.
- Tomasello, M. (2016). *A natural history of human morality*. Harvard University Press.