

Part I: Introduction

It is widely accepted that “the criminal law generally assumes the existence of free will” and assumes that those who freely choose to commit crimes are morally responsible for their actions [1]. The concepts of “free will” and “moral responsibility” are pre-requisites for retributive punishment, and judges often refer to the offender’s moral responsibility when sentencing [e.g.2,3,4]. A rich philosophical literature exists on the meaning of these concepts – “nearly every major figure in the history of philosophy has had something ...to say about free will” [5:5]. Recently, a growing number of theorists working on free will and moral responsibility have begun to focus on the practical implications of their views for the punishment of offenders [e.g. 6,7,8,9,10]. However, much more work needs to be done to enable this literature to have an impact on the criminal justice system [11].² Experimental philosophy has the potential to help bridge the gap between philosophical debate and penal policy. This relatively new approach seeks to investigate empirically how lay people actually use terms that have long interested philosophers. In particular, many of these studies have investigated whether lay people consider free will and moral responsibility to be compatible or incompatible with determinism. Theorists would be in a better position to communicate arguments for penal reform if they had a better understanding of how lay people – upon whose support the criminal justice system depends – currently interpret terms like “moral responsibility” and “free will”, which figure prominently in reformers’ arguments. Studies on “motivated reasoning” suggest that lay people’s use of such terms is driven by social motives, such as punitive desires, so that they use these terms differently in different situations, depending on the goal they are seeking to achieve in that situation [12]. Using the same term in different ways gives rise to the risk of arguing at cross-purposes. A better understanding of the concerns that drive lay people’s use of the terms “free will” and “moral responsibility”, could help to avoid confusion and enable theorists to focus on addressing those concerns directly.

This article has three aims. Firstly, it aims to identify certain questions that have not been sufficiently explored in previous studies on the compatibility of moral responsibility and determinism. These underexplored questions will first be briefly summarised in this introductory section and then will be explained in more detail in Parts II-V, which will define the key terms involved and will discuss the relevant background literature, and Part VI, which will outline the aims and hypotheses of an exploratory study (involving 602 participants) designed to investigate these questions.

¹ We are grateful to the Carnegie Trust for funding this research and to the journal editor and reviewers for their comments.

² Although important philosophical writings on free will have occasionally been mentioned by the judiciary. See *R v Smith (Morgan)* [2001] 1 A.C. 146, per Lord Hoffman at p167, citing Strawson, P. 2008. *Freedom and resentment and other essays*. New York: Routledge. For a discussion of how applied ethics might have more impact on society see [11].

Secondly, this article will discuss the methodology and findings of our exploratory study (Parts VII- IX). Thirdly, in Part X we will map out some directions for future research that would examine alternative explanations for our findings and would shed more light on the questions we investigated.

The first underexplored issue that we identified concerns whether lay people interpret the term “moral responsibility” in a retributive way and whether they are motivated to hold offenders responsible for pre-determined behaviour by considerations other than retributivism, e.g. the desires to condemn the action (as opposed to the actor) and to protect society. Specifically, we maintain that the following question deserves greater attention: are participants less likely to apply the term “morally responsible” to an offender whose conduct was predetermined if participants are given the (non-retributive) options of denying that the offender is morally responsible while simultaneously a) condemning the action (as opposed to the actor) and/or b) supporting measures to protect society? One novel contribution of our exploratory study was the way in which we aimed to shed light on the above-mentioned question concerning lay people’s use of the term “moral responsibility”. Our preliminary findings were that the majority of participants gave incompatibilist responses about moral responsibility and determinism (even when the non-retributive options were not mentioned) and that incompatibilist responses increased when the non-retributive options were explicitly mentioned. This raises the possibility that a significant proportion of participants either i) may not interpret “moral responsibility” in the basic, retributive sense of the term, which is at issue in the determinism debate, or ii) may be unconsciously motivated by non-retributive considerations to judge that the offender is morally responsible, in the basic, retributive sense. Failure to distinguish between different senses of moral responsibility and to consider non-retributive motives for holding offenders responsible might have led certain previous researchers to overestimate the number of folk compatibilists. If our findings are confirmed by future research, a wider implication would be that theorists’ arguments against retributivism are more likely to affect public attitudes to punishment when non-retributive ways of achieving important punishment goals are emphasised.

The second underexplored issue that we identified concerns whether explicitly rejecting moral responsibility and retributivism, after reading about determinism, would have any impact on “implicit” retributivism when recommending a sentence for a hypothetical offender. Certain types of aggravating/mitigating factors seem more relevant to the offenders’ future dangerousness (a non-retributive consideration), while other types of aggravating/mitigating factor seem more relevant to the offender’s blameworthiness for the crime he has committed (a retributive consideration). Participants who place greater weight on factors relevant to blameworthiness are more “implicitly retributive” than those who place greater weight on the offender’s dangerousness. Our preliminary findings

could be interpreted as suggesting that explicit retributivism did not correlate with implicit retributivism, in this sense, (although it seems that explicit rejection of retributivism did correlate with more *lenient* sentencing).³ If this interpretation is confirmed in future research, this would imply that free will theorists who wish to affect public attitudes toward punishment should, when communicating their research to the public, give detailed consideration to the implications for sentencing.

Part II: Retributivism and the challenge of determinism

Among penal theorists, “a broadly retributive theory of punishment...remains in the ascendancy in the Anglo-American world” [13:145]. The majority of legal systems contain retributive elements and there is a growing body of empirical evidence that most lay people have strong and persistent retributive attitudes [14]. At the heart of retributivism is the following claim: if an individual, acting freely, commits a serious wrong, then the offender deserves to be punished because of the blameworthy nature of this free action, whether or not punishment promotes further beneficial consequences for society [15,16].

However, for millennia it has been debated whether humans possess the kind of free will required for retributive punishment to be fair. [17]. One traditional challenge to the idea that humans possess the relevant kind of free will arises from determinism - which implies that all our decisions and actions are determined by factors beyond our control, such as our biological constitution and environment. If determinism is true, then these factors determine how our characters develop and which reasons we see fit to act upon. Given the presence of these factors, all of our decisions *had to occur* exactly as they did. This challenge to retributivism has received new impetus from neuroscience [see e.g., 18: 8-9, 19: 22-36, 20]. Neuroscientists currently adopt the philosophical assumption that “all behaviour is caused by our brains” [21:161]. If everyone’s motivations, personality and deliberations are entirely determined by their brains’ development (and the way the brain responds to their situation), which in turn was caused to develop in that exact way by genetic and environmental factors beyond their control, the question arises whether it could be fair to inflict retributive punishment on anyone.⁴

³ This is in line with previous research on “Part V: Explicit versus implicit retributivism” [38]. However, this previous research had not examined the effect of reading about determinism on explicit and implicit retributivism.

⁴ The study reported in this article investigated whether lay people considered moral responsibility to be incompatible with “biopsychosocial determinism”. This idea implies that every decision a person makes is the result of a set of biological, psychological and social factors that were sufficient to guarantee that the person would make exactly that decision. When describing biological factors, our study explicitly mentioned the brain’s role in decision-making, because in recent years, evidence from brain science has been among the main reasons that theorists have cited for taking determinism seriously and for questioning our current practices of holding offenders responsible and subjecting them to retributive punishment.

There are two ways of attempting to defend retributivism from this kind of challenge – compatibilism and libertarianism. Compatibilism is the idea that moral responsibility and determinism are compatible. (For overviews of different varieties of compatibilism see [22,23].) According to an influential version of compatibilist retributivism, all that is required for free will and retribution, are capacities, such as the ability to respond to reasons. If a person is rational, understands the reasons against breaking the law and does not care enough about those reasons to refrain from offending, then she is a suitable candidate for retributive punishment: it is irrelevant that her genes and environment made it inevitable that she would develop into the kind of person who would not care enough about the reasons for obeying the law, see, e.g. [24,25], drawing on the account of reasons-responsiveness developed in [26].⁵

⁶ Libertarian incompatibilists maintain that determinism and free will are incompatible. Furthermore, although libertarians accept that factors beyond our control may *influence* us, they deny that our decisions and actions are completely determined. According to libertarians, it is still possible, up to the moment of choosing, for a free agent to decide differently. (For an overview of libertarian theories see [27]; for an account which links this conception of free will to retributive punishment see, e.g., [28].)

Our criminal justice system seems to presuppose that offenders have “free will” in some sense of the word. However, it is disputed whether current legal doctrines are more consistent with a compatibilist or a libertarian conception of free will [29]. As brain science breathes new life into the free will debate, theorists have increasingly speculated about whether growing acceptance of determinism is capable of “radically transforming the criminal justice system” [20:1784]. In an influential article [20], Green and Cohen predict that this transformation will come about through a shift in *public attitudes*, rather than through legal arguments. They claim that most lay people intuitively believe that retributive punishment is only fair if offenders have libertarian free will. They predict that greater public understanding of neuroscience will undermine belief in libertarian free will and that, as a result, retributive sentencing will lose public support and we will have to seek alternative, non-retributive responses to criminal behaviour.

Recently, researchers have begun to empirically investigate lay people’s intuitions about free will and determinism [e.g.30,31]. These studies have yielded important insights into the factors that can affect lay peoples’ intuitions about

⁵ However, in those rare cases where someone cannot care *at all* about the reasons against committing serious crimes, like murder, some compatibilists argue that this suggests she does not fully understand those reasons and might relieve her from responsibility on that basis, see, e.g. [25].

⁶ Some compatibilists, have defended “historical” pre-conditions for moral responsibility, which require that the agential characteristics in virtue of which a person is responsible (e.g. one’s rational capacities and values) must not have the “wrong type” of causal history. For example, Haji’s account would rule out a history that involves extreme manipulation by others, such as indoctrination or interference by a malevolent neuroscientist: Haji, I., 1998. *Moral appraisability: Puzzles, proposals, and perplexities*. Oxford: Oxford University Press. However, causation by “normal” factors, not involving manipulation (e.g. genetic factors), would not undermine responsibility on this account, even if these factors make it inevitable that one will develop into the kind of person who behaves immorally.

this topic - for a discussion of the value of this research see [32]. However, studies in this area should pay more attention to clarifying certain key terms, such as “moral responsibility”. Participants are typically presented with a scenario involving someone in a deterministic universe performing an immoral action. Participants are then asked whether the wrongdoer is morally responsible. Unfortunately, it is unclear how participants interpret the term “moral responsibility”, and, specifically, whether they understand it in the sense that is relevant to retributive punishment. To be morally responsible in the retributive sense is to be deserving of punishment in virtue of the blameworthy nature of the action, rather than to promote further beneficial consequences for society. Retributive responsibility involves “*basic*” desert, in that one deserves blame/punishment *just because* “one has performed the [morally wrongful] action, given sensitivity to its moral status, and not by virtue of consequentialist or contractualist considerations” [33:122]. This particular type of responsibility is not only of utmost significance to the institution of punishment [13, 14], but is actually the sense of responsibility that is really at issue in the debate about free will and determinism [34:123]. There are other senses of responsibility, which are much more easily reconciled with determinism. For instance, individuals might be regarded as responsible in a forward-looking, consequentialist, sense, if blaming/punishing them is likely to bring about the good consequence of weakening their disposition to commit wrongful acts in the future [35].

Part III: Motivated reasoning about responsibility

In studies that did not distinguish responsibility in the retributive sense from forward-looking considerations like social protection [e.g.30] participants’ responsibility judgements might have been motivated (consciously or unconsciously) by the desire to punish the offender to stop such crimes happening in future. In other words, participants may have judged offenders “responsible”, in order to justify punishing offenders to protect society. If so, it would be unwarranted to draw the conclusion from such studies that lay people are “compatibilists” or “incompatibilists”, because (a proportion of) participants would not have interpreted responsibility in the retributive sense - the sense that is at issue in the debate between compatibilists and incompatibilists. There are good reasons for suspecting that judgments about moral responsibility may be influenced by the desire to punish,⁷ since existing research on motivated reasoning strongly suggests that the use of responsibility-related concepts such as “free will” [12] and “intention” [36,

⁷ This article uses the phrase “desire to punish” in a broad sense that is intended to be neutral on the issues such as whether a) participants were motivated by the thought that the offender ought to be punished and by the perception that holding him responsible was a necessary step before that could happen or b) whether participants’ desire to punish was more like an affective “gut reaction” that motivated their use of responsibility-related terms without any process of conscious reasoning. For a discussion of the different possibilities see [37].

37] are driven by this desire.⁸ For example, in one study, two groups of participants were presented with one of two scenarios – a man performing a morally neutral action (foraging for discarded aluminum cans and selling them to a recycling company) or a man performing a wrongful action (burglary) [12]. Participants were then asked to what extent they believed the man exercised free will. In both scenarios the man was described as being subject to exactly the same economic pressures – he was unemployed and had two children to support. Yet participants attributed significantly more free will to the robber than to the forager.⁹ A persuasive explanation for this result is that people desire to blame and punish the robber (but not the forager), and so they are motivated to attribute free will only to the robber in order to justify these desires.

So far, studies on motivated reasoning have focused on the effect of the *general desire to punish* on lay people's use of responsibility-related concepts. These studies have not investigated the effect of the *specific purposes of punishment* (e.g. retribution, or consequentialist goals such as protecting society from crime). The study reported here begins to explore this question by investigating the extent to which consequentialist (as opposed to retributive), motives affected lay people's use of the term "moral responsibility".

Part IV: Punishment and communication

Apart from social protection, another possible non-retributive motive for holding an offender responsible is the desire to condemn his wrongful action (as opposed to condemning the actor). Previous empirical research has attempted to determine the strength of people's desire to communicate disapproval of wrongful actions through (informal) punishment. One study revealed that people still desire to punish wrongdoers even if the wrongdoer is not aware he is being punished [14]. The researchers interpreted this as casting doubt on the idea that the desire to communicate censure is a major motivation lay people's decisions about punishment. However, this study did not rule out the possibility that the desire to condemn the offender's action had some motivational force. The researchers seemed to assume that the point of communicating condemnation is to express censure *to the offender*. Although this is typically one of the main purposes of condemnation, participants might have been motivated to punish the offender by the desire to express condemnation to relieve *their own feelings*, or to let *the researcher* know that the participants condemned the wrongdoing, even when they could not communicate condemnation to the offender. Likewise, in

⁸ Research by Joshua Knobe [36] in particular has been very widely discussed and has proved highly replicable: Cova, F. et al. 2018. Estimating the reproducibility of experimental philosophy. *Review of Philosophy and Psychology*. Although there have been various explanations of Knobe's findings, explanations in terms of motivation by the desire to punish or blame have been especially influential.

⁹ Interestingly, members of the group who read about the robber were, after reading this scenario, also significantly more likely to report belief that free will exists in general, compared to members of the group who read about the forager, suggesting that the desire to punish also promotes belief in free will in general.

previous studies on whether lay people consider determinism and moral responsibility to be compatible, participants may have been motivated to hold the offender morally responsible in order to condemn the offender's action (as opposed to the actor) – either to relieve their own feelings, or to let the researcher know that they disapproved of the crime. The exploratory study reported in this article aimed to shed some light on whether lay people's judgements about responsibility for pre-determined offending are influenced by the desire to condemn the crime.

Part V: Explicit versus implicit retributivism

Greene and Cohen predict that public acceptance of neurobiological determinism will transform the criminal justice system by leading to the rejection of retribution (a prediction that has been empirically investigated e.g. by [42] and [46]). For this transformation to take place, it would not be enough for the public to have theoretical doubts about whether offenders are truly responsible for their actions. These doubts must be translated into political support, in practice, for non-retributive sentencing. In other words, "explicit" non-retributivism (i.e. saying that one rejects retributivism, when asked about which general principles should inform sentencing) needs to translate into "implicit" non-retributivism (i.e. only giving weight to non-retributive considerations, when deciding which sentences are appropriate in particular cases). For example, if someone explicitly claims to believe that the goal of sentencing should be preventing recidivism, rather than retribution, this seems to imply that they will give more weight to factors such as the offender's likelihood of reoffending than to considerations which would matter more to retributivists, such as whether the offender's reasons for committing the original offence indicated a blameworthy state of mind. However, in reality, there seems to be a discrepancy between what people say about the goals of sentencing and how they actually behave when faced with specific sentencing decisions. In fact, some studies suggest that the majority of explicit non-retributivists are actually implicitly retributive [38]. Therefore, in addition to asking whether the offender was responsible, we also asked some questions about what sentence would be appropriate and about the relevance of the offender's blameworthiness versus his dangerousness to the sentencing decision.

Part VI: Aims and hypotheses

The research discussed above gives rise to the following questions and hypotheses, which we investigated in this study:

1. *Will participants be less motivated to judge that an offender is responsible for pre-determined behaviour when participants realize that they can achieve non-retributive goals (e.g. social protection or condemning wrongful actions) without holding the offender responsible?*

We suspected that in previous studies that left the term “moral responsibility” undefined a proportion of the participants who appeared to give compatibilist responses were not interpreting the term “moral responsibility” in the sense that is relevant to retributive punishment. It seems plausible that when confronted with a description of serious wrongdoing, participants’ responses are affected by a) the desire to condemn the wrongful action (as distinct from condemning the wrongdoer) and b) the desire that something should be done to prevent such wrongdoing happening in future. These two considerations are separate from the question of whether the wrongdoer deserves retributive punishment. When participants are simply presented with two stark alternatives – to state that the offender either is or is not morally responsible – a proportion of participants may have feared that denying that the offender is morally responsible might suggest that the participants condone the action and are not concerned about preventing such actions in future. If we are correct, then a proportion of participants in previous studies who appeared to consider moral responsibility to be compatible with determinism, were in fact motivated by non-retributive considerations. The following hypotheses were proposed: **H1** When participants are invited to either a) hold the offender morally responsible or b) to deny that he is morally responsible *while simultaneously condemning the wrongful action*, more participants will be willing to deny that the offender is morally responsible than when they are simply asked about the offender’s responsibility status without being given the opportunity to condemn the action. **H2** When participants are invited to either a) hold the offender morally responsible or b) to deny that he is morally responsible *while simultaneously proposing that measures are taken to prevent such behaviour in future*, more participants will be willing to deny that the offender is morally responsible than when they are simply asked about the offender’s responsibility status without being given the opportunity to propose preventive measures.

2. *Do participants who explicitly state that responsibility and retribution are incompatible with retributivism also implicitly reject retributive considerations when recommending an appropriate sentence for a hypothetical offender whose behaviour was predetermined?*

In view of the evidence on explicit versus implicit retributivism discussed above, we formulated a further hypothesis, **H3**: a significant proportion of participants who deny that the offender is morally responsible will nevertheless be influenced by retributive considerations when recommending an appropriate sentence.

Part VII: Methodology

929 participants were recruited via Prolific Academic (a crowdsourcing platform supported by Oxford University Innovation). The study was conducted online using Qualtrics research software. The data was analysed using SPSS statistical software.

The scenario

The study began with the statement below:

“The following scenario takes place in 2080. The scenario involves scientific evidence being presented in a criminal court. Please imagine that science in 2080 has advanced far beyond science today and that scientists know far more about how we make conscious decisions.”

We included this statement in order to make it easier for participants to accept our stipulation that, for the purposes of the experiment, they should assume that human decision-making was subject to determinism. We feared that participants would be more reluctant to accept this stipulation if the scenario were set in the present, given that there is still much that scientists do not know about the human brain. To prevent the scenario from seeming too remote from the real world, it was not set in the very distant future. Earlier studies have been criticized for involving unrealistic, science fiction scenarios, e.g. events taking place on another planet. There is a risk that people’s reactions to such scenarios might not correspond to the intuitions that they would have about the real world [39]. The present study aimed to reduce this risk, by setting the scenario in the relatively familiar environment of a courtroom.

Next, participants read the following statement about an offender’s wrongful action (burglary):

“Mr Williams has been found guilty of breaking into the house of his neighbour, Mr Samuels, and stealing some money from inside the house.”

It was important, given the aims of the study, to describe the scenario in terms of a named offender, a named victim and a specific offence. The study aimed to examine the effect of the participants’ desire to punish the offender on the participants’ judgments about responsibility. Earlier research suggests that this desire to punish is more salient when the participants are asked to consider a concrete case of wrongdoing, rather than only being asked about the responsibility-status of pre-determined human beings in the abstract [40].

Participants then read the testimony of an expert witness – a neuroscientist giving evidence at Mr Williams’s trial - who argued that Mr Williams’s actions were the result of neurobiological determinism. The participants were asked, when answering the subsequent questions, to assume that everything the expert said accurately represented how human decision-making worked. Here is the expert’s description of neuro-determinism:

“When we make a conscious choice to act in a particular way, we make this particular choice (rather than any other choice) because of the way our character has developed and because of the reasons that we are motivated to act upon. There is strong agreement within the scientific community that brain processes determine the way our character develops and determine which reasons we are motivated to act upon in a particular moment.

Every unconscious and conscious thought (including all the thoughts that make up our character and our reasons for acting) is linked to particular brain processes, so that every time a specific brain-process occurs, the corresponding thought-process must also occur.

There is also agreement in the scientific community that our genes and our environment determine the way our brain, our character and our thought processes develop. By 'environment', we mean all of our life experiences and the circumstances of our current situation.

It is also agreed that our environment and our genes are completely caused by earlier events that eventually go back to events that occurred before we were born. Therefore, factors beyond our control - our genes and events that occurred in the distant past - are the ultimate cause of every decision we make in life. Given the presence of these factors, all of our choices had to happen exactly as they did.”

This initial description of determinism made reference to brain processes (rather than, for example, describing determinism in purely psychological terms, or simply in terms of causation by “prior events”). The role of the brain was emphasized, because the study aimed to shed light on whether Greene and Cohen were right to think that public acceptance of determinism and rejection of retribution would come about *as a result of increased public awareness of brain science*. However, placing emphasis on the brain when describing determinism gives rise to the risk that participants will assume that the brain brings about behaviour in a way that is unresponsive to what is going on in the person’s environment. There is also the risk that participants will make the “by-passing error” and incorrectly assume that the brain influences our actions in a way that necessarily *bypasses* conscious reasoning [31]. In order to reduce

such risks, we explicitly stated that brain-processes were, in turn, affected by environmental (as well as genetic) factors and stated that decision-making involves “conscious choice[s]” and “thought-process[es]”. In addition, in keeping with the courtroom scenario, we described a lawyer cross-examining the expert witness in order to seek clarification on matters that could potentially confuse a layperson. Here is an example:

“Lawyer: If Mr Williams’s genes and environment caused his decisions, does that mean reasons played no role in his decision-making?”

Expert: Mr Williams was influenced by reasons. Your ‘environment’ includes the reasons that you come across for behaving one way rather than another. For example, a parent or teacher might sit down and explain to a child why it is important to obey the law. That explanation is part of the child’s environment. When the child grows up he might read a newspaper article about the impact of crime on its victims. That article is also part of his environment and will provide a reason to obey the law.

Lawyer: Did Mr Williams understand the reasons why burgling his neighbour’s house was wrong? For example, did Mr Williams understand that the burglary would distress his neighbour?

Expert: Yes he did. But his genes and environment *guaranteed* that, on the night of the crime, Mr Williams would not be sufficiently motivated by that reason against burgling his neighbour’s house (the distress it would cause his neighbour). Given his genes and environment, it was impossible for Mr Williams to be motivated by that reason on this occasion.”

We included a number of manipulation checks in the study to check that participants had read and understood the descriptions of the hypothetical crime and of determinism, e.g. “So, according to the expert, is every thought process linked to a brain process?”.

Explicit retributivism part 1: The responsibility question

After reading the scenario, participants were divided into four groups. Members of all groups were presented with two statements and asked which statement best reflected their own opinion of Mr Williams (on the assumption that the expert witness’s evidence is accurate). The options presented to each group are shown in table 1:

	Group 1	Group 2	Group 3	Group 4
	Responsible	Responsible	Responsible	Responsible
	Versus	Versus	Versus	Versus
	Non-responsible	Non-responsible	Non-responsible	Non-Responsible
		+ condemn action	+ protect society	+ protect society and condemn action
Options	<p>a) "Mr Williams is morally responsible for his decision to commit burglary</p> <p>or</p> <p>b) "Mr Williams is not morally responsible for his decision to commit burglary."</p>	<p>a) "Mr Williams is morally responsible for his decision to commit burglary"</p> <p>or</p> <p>b) "Mr Williams is not morally responsible for his decision to commit burglary, but it should still be recognised that Mr Williams's actions were morally wrong."</p>	<p>a) "Mr Williams is morally responsible for his decision to commit burglary</p> <p>or</p> <p>b) Mr Williams is not morally responsible for his decision to commit burglary, but Mr Williams should receive a sentence that aims to protect society from behaviour like this happening in the future."</p>	<p>a) "Mr Williams is morally responsible for his decision to commit burglary</p> <p>or</p> <p>b) Mr Williams is not morally responsible for his decision to commit burglary, but it should still be recognised that Mr Williams's actions were morally wrong and Mr Williams should receive a sentence that aims to protect society from behaviour like this happening in the future."</p>

Explicit retributivism part 2: Justifying coercive measures

After asking the questions about whether a hypothetical offender was responsible for pre-determined behaviour, we asked the participants about their level of agreement (ranging from 1 = "strongly disagree" to 6 "strongly agree") with a number of statements about the justification for imposing coercive measures on the hypothetical offender. One of these statements contained the following retributive justification: the offender "deserves to be punished because he made a free choice and is blameworthy". The other justifications referred to the following non-retributive considerations: rehabilitation, compensation, incapacitation, specific deterrence and general deterrence.

Implicit retributivism

Next, participants were given two additional pieces of information about the offender: information about whether his motive for offending was good or bad (reflecting retributive concerns) and whether his risk of reoffending was low or high (reflecting consequentialist concerns). Participants were shown one of the following 4 combinations: i) good motive + high risk, ii) good motive + low risk, iii) bad motive + high risk, iv) bad motive + low risk. The “good motive” was described in the following terms: “Mr Williams wanted the money from the burglary in order to protect his son from a drug dealer. The drug dealer was threatening to harm his son if he did not pay the money”. The “bad motive” was described as: “Mr Williams wanted the money from the burglary in order to pay a drug dealer”. The “high risk” scenario was described as follows: “A reliable and expert psychologist conducts a rigorous assessment of the social, psychological and biological factors affecting Mr Williams. The psychologist concludes that Mr Williams has a very high (87%) likelihood of committing further crimes”. In the “low risk” scenario, it was stated that “...The psychologist concludes that Mr Williams has a very low (3%) likelihood of committing further crimes”. Participants were then asked to recommend a prison sentence for the hypothetical offender between 0 and 24 months. Finally, participants were asked to rate how much the offender’s motive and likelihood of reoffending had influenced their sentencing decision using the following scale: 1 = very unimportant, 2 = unimportant, 3 = slightly unimportant, 4 = slightly important, 5 = important, 6 = very important.

Part VIII: Results

Participants

After excluding the 327 participants who answered at least one of the five comprehension checks incorrectly, 602 participants remained. Feltz and Millan observe that “Having a large number of comprehension failures is common with experiments concerning free will. For example, Nahmias et al. (2007) excluded 22% of their sample for failing comprehension checks. Sommers (2010) notes that people routinely exclude 10–30% of participants in these types of studies. However, in many studies, there were no appreciable differences between those who failed the comprehension checks and those who passed.”^[41] Furthermore, the relatively high number of comprehension checks probably led to a larger proportion of participants being excluded from the sample than would otherwise be the case. The comprehension checks were introduced in order to reduce the risk that participants would be influenced by various misconceptions that have been identified in the existing literature, e.g. the assumption that determinism bypasses conscious deliberation [47] and to reduce the risk that participants (despite our instructions to the contrary) would revert to their default assumption that people (including the hypothetical offender in the scenario) possess indeterministic free will [48].

Of the 602 remaining participants, the following proportions of the sample had or were studying the following subjects at university: 5.81% law, 14.85% psychology, 3.12% neuroscience, 12.06% biology, 18.84% any other science, 9.58% sociology, 4.09% criminology, 10.33% any other social science, 9.26% philosophy, and 19.38% any other of the arts or humanities. In respect to gender, 49.41% percent of the sample identified as female, 49.30% identified as male and .97% identified as other. The following proportions of the sample had completed the following highest levels of education: 1.08% less than secondary/senior/high school, 20.02% secondary/senior/high school, 26.70% education obtained between school and university, 38.32% bachelor’s degree, 11.30% master’s degree and 2.15% doctorate degree. While 49 nationalities were represented in the sample, by far the most common nationalities were American (36.06%) and British (35.41%). The mean age of the participants was 32.26 (SD = 11.00). The present study was larger than many previous studies and the age-range of participants was wider than many previous studies, in which participants were typically undergraduates.

The responsibility question

Table 2 indicates the proportion of participants who attributed responsibility to the predetermined offender in each condition.

Table 2: Providing the options of condemning the crime and protecting society without holding the offender responsible: The effect on compatibilist responsibility judgments				
Question	Group 1	Group 2	Group 3	Group 4
Wording	Responsible	Responsible	Responsible	Responsible
	Versus	Versus	Versus	Versus
	Non-responsible	Non-responsible	Non-responsible +	Non-
		+ condemn action	protect society	

				Responsible + protect society and condemn action
Proportion (%) who say he is responsible	41.50	31.33	26.75	23.65

When participants are given the option of achieving the (non-retributive) goals of social protection and condemning the action (as opposed to the actor) without holding the offender responsible, the proportion of compatibilist responsibility judgments decreases. Specifically, a significantly smaller proportion of participants held the offender responsible when given the independent opportunity to condemn the offending behaviour ($\chi^2(1, N = 464) = 4.03, p = .045$), protect society from further offending ($\chi^2(1, N = 464) = 5.76, p = .016$) or both ($\chi^2(1, N = 465) = 8.59, p = .003$). These results appear to confirm our first two hypotheses.

Public support for incompatibilism?

In addition to showing that the question wording affected responsibility judgements, it is clear from Table 1 that the majority of participants were “incompatibilists” about responsibility and neuro-determinism, i.e they did not consider responsibility to be compatible with neuro-determinism. This was found across all 4 variations of the question wording – even the first variation, which involved a simple choice between responsibility and non-responsibility. Previous studies that presented participants with this simple choice yielded conflicting results about public support for compatibilism versus incompatibilism [e.g. **Error! Bookmark not defined.** and **Error! Bookmark not defined.**]. We had not made a prediction about whether the majority of participants would be compatibilists or incompatibilists. We had only hypothesized that our alterations to the question wording would affect these results.

The “fragility” of incompatibilist responses

The question about whether retributive considerations (which cited the offender’s “free choice” and “blameworthiness”) provided good reasons for imposing coercive measures on the pre-determined offender gave rise to the following interesting finding: A significant proportion (43.7%) of participants who had given incompatibilist responses when asked about responsibility (i.e. judged that the offender was “not responsible” for neurobiologically determined behaviour), nevertheless, later in the survey, expressed some level of agreement with the retributive idea

that the same offender “deserved to be punished because he had made a free choice and was blameworthy”. Participants were asked to rate their level of agreement with this statement about deserved punishment on the following scale: 1 = strongly disagree, 2 = disagree, 3 = slightly disagree, 4 = slightly agree, 5 = agree, 6 = strongly agree. Of those participants who had denied that the offender was morally responsible, but who also expressed some agreement with the statement about deserved punishment, the mean level of agreement was 4.53 (SD 0.66).

Explicit versus implicit retributivism

Explicit judgments about responsibility had no significant effect on the extent to which participants were implicitly influenced by the retributive consideration (motive) as opposed to the consequentialist consideration (risk of reoffending). We had merely hypothesised (H3) that there would be *some disparity* between explicit responsibility judgments and implicit retributivism and so the finding of *no* significant relationship went beyond what we had hypothesised.

Sentencing and the correlation between compatibilist retributivism and punitiveness

Participants who endorsed the idea that the offender was responsible and deserved retributive punishment for neuro-determined behaviour recommended a slightly longer prison sentence than participants who considered the offender non-responsible and who endorsed only non-retributive justifications for the sentence. Attitude Towards Retributivism had a significant ($p = .048$) effect, with retributivists recommending a mean sentence of 12.39 months and non-retributivists recommending a mean sentence of 10.17 months.

Part IX: Discussion

This section will compare the aims and preliminary findings of our exploratory study with some earlier studies in this area and will discuss the ways in which the questions we have begun to investigate advance beyond the existing research. We will also consider some of the limitations of our exploratory study and alternative explanations for our preliminary findings, which raise possible directions for future research (discussed further in the next section). *The*

Responsibility Question

Previous studies investigated lay people’s intuitions about the compatibility of determinism and “responsibility”. In this article we have discussed the importance of clarifying different senses of the term “moral responsibility” and of investigating lay people’s motives for holding offenders morally responsible. Our preliminary findings raise the possibility that that a significant proportion of participants either 1) may not interpret “moral responsibility” in the

basic, retributive sense of the term, which is at issue in the determinism debate, or 2) may be unconsciously motivated by non-retributive considerations to judge that the offender is morally responsible, in the basic, retributive sense. Failure to distinguish between different senses of moral responsibility and to consider non-retributive motives for holding offenders responsible might have led certain previous researchers [e.g.30] to overestimate the number of folk compatibilists. If our preliminary findings are confirmed by future research, a wider implication would be that theorists' arguments against retributivism are more likely to affect public attitudes to punishment when non-retributive ways of achieving important punishment goals are emphasised. One limitation of our exploratory study was that, when asking the responsibility question, the word length was not identical for each option. If participants preferred the options that were described using more words, this might partly explain our preliminary findings. Future follow-up studies could be designed to address this issue.

Previous studies [e.g. 42, 43] have drawn a distinction between retributive and *certain* non-retributive reasons for punishment, but they did not distinguish retribution from condemning a criminal action (as opposed to the actor) as we did in the present exploratory study. Furthermore, these earlier studies focused on retributive versus non-retributive *punishment*, but did not investigate the specific questions raised in this article concerning the use of the term "moral responsibility". One novel contribution of our study was the approach it took to investigating the use of this particular term. The specific question we investigated was: would participants be less likely to apply the term "morally responsible" to an offender whose conduct was predetermined if participants were given the (non-retributive) options of denying that the offender is morally responsible while simultaneously a) condemning the action (as opposed to the actor) and/or b) supporting measures to protect society? Shariff et al conducted a series of ingenious studies that appear to suggest that reading "mechanistic" explanations of human behaviour reduces belief in free will and responsibility and thereby reduces support for retributive punishment - which is a very significant finding. However, they did not aim to examine the question investigated in the present study concerning whether participants' use of the term "moral responsibility", in the context of a deterministic scenario, would be affected by reading that it was possible to achieve non-retributive goals without holding the offender responsible. Furthermore, the conception of "mechanism" used by Shariff et al (while worth investigating) was very different from the concept of determinism employed in the present study.¹⁰ For example, consider the following extracts from their study [44]:

¹⁰ Shariff et al also used certain items from the FAD and FAD plus tools for measuring participants' belief in free will, but they did not select any of the items designed to measure belief in determinism. For an illuminating discussion of the limitations of these tools see Nadelhoffer et al 2014. One of the four studies contained a reference to determinism in a manipulation check: "to what degree do you believe that humans have free will in the sense that we can consciously generate spontaneous choices and actions, not fully determined by prior events?". However, this statement is very brief and participants may not have fully understood/paid attention to the phrase "determined by prior events" and may have focused on the word "consciously", given the focus of the main passage they were asked to read. Furthermore, for this item to distinguish between compatibilists and

“Three assumptions can be made about free will. The first assumption is that part of one’s brain is concerned with making plans for future actions...The second assumption is that one is not conscious of the “computations” done by this part of the brain but only of the “decisions” it makes...The third assumption is that the decision to act on one’s plan...is also subject to the same limitations in that one has immediate recall of what is decided, but not of the computations that went into the decision...One’s self can attempt to explain why it made a certain choice. Sometimes we may reach the correct conclusion. At other times, we will either not know or, more likely, will confabulate, because there is no conscious knowledge of the ‘reason’ for the choice. This implies that there must be a mechanism for confabulation, meaning that given a certain amount of evidence, which may or may not be misleading, part of the brain will jump to the simplest conclusion.”

“...Your decisions are strongly prepared by brain activity. By the time you become aware of what you want to do, your brain waves have often already given you away...”

“...scientists could force the subjects to choose to move their left hands. But despite their choice being electronically directed, these patients continued to report that they were freely choosing which hand to move!”

These passages emphasise the ideas that behaviour is caused by unconscious, rather than conscious processes and that people are often mistaken about the real reasons for their own behaviour. As Sie and Wouters have pointed out, determinism can be distinguished from “...the thesis that consciousness does not influence our behaviour....Such a lack of influence (if true) would threaten compatibilist and incompatibilist positions alike.”[45:121]

Nadelhoffer et al have developed the “free will inventory” (FWI) a sophisticated tool for measuring beliefs about free will, determinism and dualism and the relationship between these beliefs and attitudes toward responsibility and punishment. The FWI overcomes many of the problems with previous tools of this kind. One virtue of the FWI is that it distinguishes between retributive and non-retributive reasons for *punishment*. However, it does not contain any items that are framed in terms of retributive versus non-retributive senses of “*moral responsibility*”. The FWI could, however, be used in future research in an adapted form, or alongside other materials in order to get another perspective on the questions raised in the present article concerning the interpretation of the term “*moral responsibility*”. Possible alterations to the FWI might include modifying the wording of the statements about responsibility in order to make a clearer distinction between compatibilists and libertarians. Currently, the FWI

incompatibilists it would have to be made clearer this item does not distinguish clearly between compatibilists and incompatibilists, as both views are consistent with believing that human actions are, in fact, not determined by prior events.

contains two items that mention “responsibility”. One item states: “To be responsible for our present decisions and actions we must also be responsible for all of our prior decisions and actions that led up to the present moment”. The phrase “led up to” could be interpreted as meaning “preceded in time” as opposed to “pre-determined”. Neither compatibilists nor libertarians would endorse the idea that to be responsible for any of one’s actions one must also have been responsible for all of one’s *preceding* actions (e.g. even childhood actions). The second item states: “People could be morally responsible even if scientists discovered all of the laws that govern human behavior.” This statement does not rule out the possibility that a) the laws governing human behaviour are indeterministic (e.g. merely probabilistic) or b) even after all the laws that do apply to humans have been discovered, it will turn out that some aspects of human behaviour are not law-governed. Therefore both compatibilists and libertarians might agree with this item. This is not a criticism of the FWI – this particular item was intended to measure intuitions about the compatibility between responsibility and a “law governed universe” (which seems to leave it open whether the universe is governed by deterministic laws). However, the questions we investigated concerned the compatibility between responsibility and determinism. If we were to use the FWI in future research to investigate the questions raised in this article, we would need to alter the wording of the responsibility items.

Explicit versus Implicit Retributivism

Our finding that explicit rejection of retributivism correlated with more lenient sentencing was in line with previous research by Carlsmith, and Shariff et al [3838; 44]. It was not in line with research by Scurich and Appelbaum that genetic evidence had neither an aggravating nor a mitigating effect on sentencing (but that might be due to the possibility that participants in Scurich and Appelbaum’s study might not have interpreted the genetic evidence as being deterministic)[46]. Our study aimed to go beyond the existing research by investigating whether, after reading about determinism, participants who explicitly rejected retributivism would also be implicitly retributive, in that they would be more influenced by evidence of the offender’s blameworthiness than by evidence of his dangerousness when recommending a sentence. Carlsmith had not investigated the impact of *reading about determinism* on implicit versus explicit retributivism. Shariff et al investigated the effect of reading about mechanistic explanations of behavior on the length of sentence participants recommended, where it was stipulated that the sentence would not help to prevent recidivism or deter other offenders. Our study took a different approach – we compared the influence of different combinations of evidence about the offenders’ degree of blameworthiness versus his dangerousness on participants’ sentencing recommendations and also enabled participants to rate the influence that they believed this information had on their sentencing decision. One possible limitation of our study was our reliance on a description of the offender’s motive as a way of indicating his blameworthiness. It is possible that participants might have

considered his motive to be relevant to *both* his dangerousness and his blameworthiness and thus to be relevant from *both* a consequentialist and a retributive perspective. We attempted to reduce this problem, by stipulating that despite having a bad motive, the offender had been assessed as having a very low risk of reoffending and that when he had a good motive, he had been assessed as having a very high risk of reoffending. The problem is that most aggravating and mitigating factors could have *some* relevance from both a retributive and a consequentialist perspective – the difference in the relative weight accorded to the each factor by different penal theories is usually a matter of degree.

Part X: Future directions

The results lend some support to our main hypotheses (H1 and H2) that when participants are presented with a dichotomous choice between judging an offender responsible or non-responsible for pre-determined behaviour, a proportion of participants will be motivated by non-retributive reasons to hold the offender responsible. These hypotheses generated the prediction that participants will be less motivated to hold the offender responsible when they realize that they can achieve their non-retributive goals without holding the offender responsible. Our hypotheses provide a possible explanation for our preliminary findings. However, follow-up studies could be designed to test alternative explanations, e.g. the possibility (mentioned in the previous section) that the variation in the question lengths might have partly influenced the results, and the possibility that some participants might have misinterpreted the question wording or the description of determinism in some way. We therefore intend to repeat the study with different ways of formulating the responsibility question and the description of determinism, to see if the effect persists.

We plan to discover whether the apparent public support for incompatibilism can be replicated in a future study and to investigate why our study generated this finding, whereas certain previous studies did not. We also aim to investigate whether participants are more likely to judge that responsibility is incompatible with determinism if the description of determinism mentions the brain (as it did in the present study). Nahmias et al [47] found some evidence that incompatibilist responses are higher if determinism is described in neurobiological rather than psychological terms. They suggested that participants are more likely to assume that the individual's conscious reasoning processes are *bypassed* by forces outside their control if the causes of the person's behaviour are described in neurobiological rather than psychological terms. There is a risk that participants might confuse this bypassing issue with the concept of determinism. We attempted to avoid this confusion by repeatedly emphasising in the study that the offender was

motivated by reasons and asking relevant comprehension questions. Follow-up studies could shed light on whether our attempts to avoid the bypassing problem were successful.

Another issue for future research is why some participants in the present study judged that the offender “not responsible” for determined behaviour, but, nevertheless agreed (to some extent) with the statement that the same offender had “free will” and was “blameworthy”. This might have been due to participants forgetting about the details of the description of determinism by the time the question about free will and blameworthiness was asked. Rose, Buckwalter and Nichols [48] suggest that a proportion of readers may be capable of reverting to their default assumption that people have indeterministic free will even when the description of determinism remains on the screen throughout. However, when participants need to recall the information about determinism, after some time has elapsed, there may be a greater risk that their initial indeterministic intuitions will lead them to misrecollect or ignore this information [49]. This could be tested by measuring the time gap between reading the description of determinism and answering the questions about free will, responsibility and blameworthiness and seeing whether compatibilist responses increase as the time gap increases. In future studies we would also investigate the concepts of “free will” and “blameworthiness” using two separate questions.

Finally, the reported study focused on the term “moral responsibility”, but a similar methodology could be adapted for studies on the use of the terms “intention” and “free will”. Future studies on the use of these latter terms could usefully investigate the following question: Does the effect of the “desire to punish” on the interpretation of the terms “intention” and “free will” disappear when participants are offered ways of achieving their punishment goals without altering their interpretation of these terms?

Conclusion

In this article, we identified two issues that deserve greater attention from those researching lay people’s attitudes to moral responsibility and determinism. The first issue concerns whether people interpret the term “moral responsibility” in a retributive way and whether they are motivated to hold offenders responsible for pre-determined behaviour by considerations other than retributivism, e.g. the desires to condemn the action (as opposed to the actor) and to protect society. The second issue concerns whether explicitly rejecting moral responsibility and retributivism, after reading about determinism, would have any impact on “implicit” retributivism when recommending a sentence for a hypothetical offender. We reported the results of an exploratory study that investigated these questions. Our

preliminary findings raise the possibility that a significant proportion of participants either i) may not interpret “moral responsibility” in the basic, retributive sense of the term, which is at issue in the determinism debate, or ii) may be unconsciously motivated by non-retributive considerations to judge that the offender is morally responsible, in the basic, retributive sense. If this is confirmed by future research, a wider implication would be that theorists’ arguments against retributivism are more likely to affect public attitudes to punishment when non-retributive ways of achieving important punishment goals are emphasised. Our preliminary findings also suggested that explicit retributivism did not correlate with implicit retributivism (although it seems that explicit rejection of retributivism did correlate with more *lenient* sentencing).¹¹ If this is confirmed in future research, it would imply that free will theorists who wish to affect public attitudes toward punishment should, when communicating their research to the public, give detailed consideration to the implications for sentencing.

References

-
- ¹ R v Kennedy (No 2) [2008] 1 AC 269 per Lord Bingham at pp 275–276, para 14.
- ² R v Topasna [2006] EWCA Crim 1969 at para 19, Lord Justice Latham quoting the sentencing judge, Dame Heather Steel.
- ³ R v Kesic 2000 WL 389578 at para 6, Mr Justice Butterfield, quoting the sentencing judge.
- ⁴ R v Greenwood [2005] EWCA Crim 2686 at para 12, per Mr Justice Hughes.
- ⁵ Timpe, K., 2013 (2nd edition). Free will: Sourcehood and its alternatives. London: Bloomsbury.
- ⁶ Pereboom, D. 2014. Free will, agency and meaning in life. Oxford: OUP
- ⁷ Caruso, G. 2017. Public health and safety: The social determinants of health and criminal behaviour. UK: ResearchersLinks books.
- ⁸ Focquaert, F., Glenn, A.L. and Raine, A., 2013. Free will, responsibility, and the punishment of criminals. In Nadelhoffer, T. (ed). The future of punishment.
- ⁹ Levy, N. 2012. Skepticism and sanction: The benefits of rejecting moral responsibility. Law and Philosophy 31 (5): 477.
- ¹⁰ Waller, B. 2017. The Injustice of Punishment. New York: Routledge.
- ¹¹ Frodeman, R. and Briggie, A. 2016. Socrates tenured: The institutions of 21st-century philosophy. London: Rowman & Littlefield International.
- ¹² Clark, C. et al. 2014. Free to punish: A motivated account of free will belief. Journal of Personality and Social Psychology 106(4): 501–513.
- ¹³ Matravers, M. 2016. De-moralising retributivism: Agency, blame and humanity in criminal law theory and practice. In The Routledge handbook of criminal justice ethics, ed Jonathan Jacobs and Jonathan Jackson, pp144-163. London: Routledge.
- ¹⁴ Nadelhoffer, T., Heshmati, S., Kaplan, D. and Nichols, S. 2013. Folk retributivism and the communication confound. Economics and Philosophy 28 (2): 235.
- ¹⁵ Moore, M., 2010. Placing blame: A theory of the criminal law. New York: Oxford University Press.
- ¹⁶ Walen, A., 2014. Retributive justice. In Zalta, E (ed.). The Stanford encyclopaedia of philosophy (Summer 2018 Edition). URL = <https://plato.stanford.edu/entries/justice-retributive/>.
- ¹⁷ O’Connor, T. and Franklin, C. 2018. Free Will. In Zalta, E (ed.). The Stanford encyclopaedia of philosophy (Summer 2018 Edition). URL = <https://plato.stanford.edu/entries/freewill/>
- ¹⁸ Caruso, G. 2012. Free will and consciousness: A determinist account of the illusion of free will. Plymouth: Lexington Books.
- ¹⁹ Honderich, T. 2003 (6th edition). How free are you? The determinism problem. Oxford: OUP
- ²⁰ Greene, J. and Cohen, J. 2004. For the law, neuroscience changes nothing and everything. Philosophical Transactions of the Royal Society of London 359: 1775.
- ²¹ Raine, A. and Glenn, A. 2014. Psychopathy: An introduction to biological findings and their implications. New York: New York University Press.
- ²² McKenna, M. and Coates, J. Compatibilism. In Zalta, E (ed.). The Stanford encyclopaedia of philosophy (Summer 2018 Edition). URL = <https://plato.stanford.edu/entries/compatibilism/>

¹¹ This is in line with previous research on “Part V: Explicit versus implicit retributivism” [38]. However, this previous research had not examined the effect of reading about determinism on explicit and implicit retributivism.

-
- ²³ McKenna, M. and Coates, J. Compatibilism: The state of the art. In Zalta, E (ed.). The Stanford encyclopaedia of philosophy (Summer 2018 Edition). <https://plato.stanford.edu/entries/compatibilism/supplement.html>
- ²⁴ Morse, S.J., 2011. Protecting liberty and autonomy: Desert/disease jurisprudence. *San Diego L. Rev.*, 48: 1077.
- ²⁵ Morse, S.J., 2008. Psychopathy and criminal responsibility. *Neuroethics*, 1(3): 205-212.
- ²⁶ Fischer, J.M. and Ravizza, M., 2000. Responsibility and control: A theory of moral responsibility. Cambridge University Press.
- ²⁷ Clarke, R. and Capes, J. 2017. Incompatibilist (non-deterministic) theories of free will. In Zalta, E (ed.). The Stanford encyclopaedia of philosophy (Summer 2018 Edition). <https://plato.stanford.edu/entries/incompatibilism-theories/>
- ²⁸ Hodgson, D., 2012. *Rationality+ consciousness= free will*. Oxford University Press.
- ²⁹ Kolber, A. 2016. Free will as a matter of law. In *Philosophical Foundations of Law and Neuroscience*, ed Michael Pardo and Dennis Patterson, pp9-29. Oxford: OUP.
- ³⁰ Nahmias, Eddy et al. 2006. Is incompatibilism intuitive? *Philosophy and Phenomenological Research* 73(1): 28-53.
- ³¹ Nichols, S. and Knobe, J. 2007. Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous* 41(4): 663.
- ³² Björnsson, G. and Pereboom, D. 2016. Traditional and experimental approaches to free will. In *The Blackwell Companion to Experimental Philosophy*, ed Wesley Buckwalter and Justin Sytsma, pp142-57. Oxford: Blackwell Publishers.
- ³³ Björnsson, G. and Pereboom, D. 2016. Traditional and experimental approaches to free will. In *The Blackwell Companion to Experimental Philosophy*, ed Wesley Buckwalter and Justin Sytsma, pp142-57. Oxford: Blackwell Publishers.
- ³⁴ Björnsson, G. and Pereboom, D. 2016. Traditional and experimental approaches to free will. In *The Blackwell Companion to Experimental Philosophy*, ed Wesley Buckwalter and Justin Sytsma, pp142-57. Oxford: Blackwell Publishers.
- ³⁵ Smart, J. 1961. Free-will, praise and blame. *Mind*, 70(279): 291-306.
- ³⁶ Knobe, J. 2003. Intentional action and side effects in ordinary language. *Analysis* 63(279): 190.
- ³⁷ Alicke, M. 2008. Blaming badly. *Journal of Cognition and Culture* 8(1): 179.
- ³⁸ Carlsmith, K. 2008. On justifying punishment: The discrepancy between words and actions. *Social Justice Research* 21: 119-137.
- ³⁹ Sommers, T. 2015. Free will and experimental philosophy: An intervention. In *The Springer Handbook of Neuroethics*, ed Clausen Jens and Levy, Neil, pp273-286. Dordrecht: Springer.
- ⁴⁰ Nichols, S. 2011. Experimental philosophy and the problem of free will. *Science* 331: 1401.
- ⁴¹ Feltz, A. and Millan, M., 2015. An error theory for compatibilist intuitions. *Philosophical Psychology*, 28(4):529.
- ⁴² Shariff, A.F et al. 2014. Free will and punishment: A mechanistic view of human nature reduces retribution. *Psychological science*, 25(8): 1563.
- ⁴³ Nadelhoffer, T. et al. 2014. The free will inventory: Measuring beliefs about agency and responsibility. *Consciousness and Cognition*, 25: 27.
- ⁴⁴ Shariff et al. 2014. Supplemental Material: <https://doi.org/10.1177/0956797614534693>.
- ⁴⁵ Sie, M. and Wouters, A., 2010. The BCN challenge to compatibilist free will and personal responsibility. *Neuroethics*, 3(2): 121.
- ⁴⁶ Scurich, N. and Appelbaum, P., 2015. The blunt-edged sword: genetic explanations of misbehavior neither mitigate nor aggravate punishment. *Journal of Law and the Biosciences*, 3(1), pp.140-157.
- ⁴⁷ Nahmias, E., Coates, D.J. and Kvaran, T., 2007. Free will, moral responsibility, and mechanism: Experiments on folk intuitions. *Midwest studies in Philosophy*, 31(1): 214.
- ⁴⁸ Rose, D., Buckwalter, W. and Nichols, S., 2017. Neuroscientific prediction and the intrusion of intuitive metaphysics. *Cognitive science*, 41(2): 482.
- ⁴⁹ Barrett, J.L. and Keil, F.C., 1996. Conceptualizing a nonnatural entity: Anthropomorphism in God concepts. *Cognitive psychology*, 31(3): 219.