Testing the intuitive retributivism hypothesis

Paul Rehren[a]* and Valerij Zisman[b]

[a]MAD Lab, Duke University; [b]Department of Philosophy, Bielefeld University

Email for correspondence: paul.rehren@duke.edu

Declaration of interest statement: The authors declare no conflict of interest.

Testing the intuitive retributivism hypothesis

Research on the question of what motivates individuals to punish criminal offenders suggests that punitive reactions are primarily responsive to retributive, but not to utilitarian, factors. Several authors have as an explanation suggested what we will call the *intuitive retributivism hypothesis*. According to this hypothesis, punitive reactions are the product of two distinct types of processing (type-I and type-II) which differentially support retributive vs. utilitarian punishment motives. When confronted with a case of criminal wrongdoing, type-I processing swiftly outputs a retributive reaction. In contrast, for utilitarian motives to play a role, this reaction has to be overridden by type-II processing, which rarely happens.

Here, we revisit the case for the intuitive retributivism hypotheses. We review several arguments in support of it but argue that they are either unconvincing or provide only very limited support. We conclude that despite its popularity, little in the way of concrete evidence for the hypothesis exists. In light of this, the research described in this preregistration hopes to provide the first direct test of the intuitive retributivism hypothesis. To this end, we propose to investigate the effect of increased processing effort on retributive vs. utilitarian punitive reactions. Along the way, we plan to conceptually replicate Keller et al. (2010 Exp. 2).

Introduction

When confronted with a case of criminal wrongdoing, most people will have the reaction that the culprit needs to be punished in some way, shape or form (Henrich et al. 2006; Hoffman and Goldsmith 2004).[1] Despite its ubiquity, however, this punitive reaction may be supported by a number of different, and sometimes conflicting, motives. Over the past 20 years, psychologists have increasingly become interested in understanding what motivates people to punish criminal offenders. For inspiration, they have looked to a much longer-standing debate in philosophy. Since the dawn of their discipline, philosophers have been trying to identify and question the reasons (if any) that justify the imposition of punishment on offenders (for an overview, Duff and Hoskins 2019). To date, psychologists have primarily

1 While there is some debate on the issue, we for the purposes of this paper understand punishment as "the imposition of consequences generally believed to be painful or burdensome on someone found to have violated the law [...] by persons vested with legal authority to impose these consequences" (sometimes called the Benn-Flew-Hart definition; Dolinko 2011, 405).

focused on the two major theories: retributivism and utilitarianism.[2]

The core claim of retributivism, perhaps most famously associated with the work of Immanuel Kant (Kant [1785] 1998), is that "punishment is justified as an intrinsically appropriate, because deserved, response to wrongdoing" (Duff and Hoskins 2019). In particular, this entails that whether punishment is justified does not depend on any beneficial consequences it may or may not have.

In contrast, utilitarianism holds that punishment can only be justified if its beneficial consequences outweigh the cost of making the criminal offender suffer (Bentham [1830] 1998; Duff and Hoskins 2019). While utilitarians have put forward a number of different proposals for what precisely the beneficial consequences of punishment are (for an overview, Wood 2010), psychologists have tended to focus on two (Carlsmith and Darley 2008, 200–2): deterrence—punishment deters the offender or other would-be criminals from committing similar offenses in the future; incapacitation—while the offender is undergoing punishment (e.g. incarceration), he or she will not be able to commit further crimes.

Review of relevant scholarship

Previous literature paints an intriguing picture of why people punish. When asked directly, people by and large report that both retributive and utilitarian motives matter for punishment (for reviews, Cullen, Fisher, and Applegate 2000; Roberts and Stalans 2004). However, when their punitive reactions are assessed not by self-report, but instead by behavioral measures, retributivism largely dominates utilitarianism (for a review, Goodwin and Gromet 2014; but van Doorn and Brouwers 2017).

Most existing behavioral research into the motives underlying punitive reactions uses a similar approach. First, researchers isolate factors of criminal offenses which are uniquely relevant either from within the retributive or from within the utilitarian framework.[3] For instance, several authors (see table 1) have put forward the magnitude of the harm caused by the offender as a factor that should only play a role for people's punitive reactions if their underlying motives are retributive. Likewise, factors such as the general frequency of the crime and the risk of offender recidivism have been argued to primarily matter from a utilitarian point of view. Table 1 shows an overview of factors that have been used in previous literature.

2 To be sure, some authors have investigated other motives, such communication (e.g. Funk, McGeer, and Gollwitzer 2014; Gollwitzer, Meder, and Schmitt 2011; Nadelhoffer et al. 2013) and restorative justice (for a review, van Doorn and Brouwers 2017).

3 A few authors (Keller et al. 2010, Exp. 3; van Prooijen 2010) have instead (or in addition) worked with factors not of criminal offenses, but of the proposed punishment.

Retributivism

- Seriousness of the offense (Darley, Carlsmith, and Robinson 2000; Roberts and Gebotys 1989)
- Magnitude of harm (Carlsmith, Darley, and Robinson 2002; Carlsmith 2006; Keller et al. 2010)
- Extenuating circumstances (Carlsmith, Darley, and Robinson 2002; Carlsmith 2006)
- Moral offensiveness of the offense (Carlsmith 2008)
- Offender intent (Aharoni and Fridlund 2012; Carlsmith 2006, 2008; Keller et al. 2010)
- Blameworthiness of the offender (Carlsmith 2008)

Utilitarianism

Deterrence

- Detection rate (Carlsmith, Darley, and Robinson 2002; Carlsmith 2006, 2008; Keller et al. 2010)
- Publicity of offense and trial (Aharoni and Fridlund 2012; Carlsmith, Darley, and Robinson 2002; Carlsmith 2006, 2008; Keller et al. 2010)
- Frequency of the offense (Carlsmith 2006, 2008; Keller et al. 2010; Roberts and Gebotys 1989)
- Frequency trend (Keller et al. 2010; Roberts and Gebotys 1989)

Incapacitation

- Risk of offender recidivism (Carlsmith 2006; Darley, Carlsmith, and Robinson 2000; Keller et al. 2010; Roberts and Gebotys 1989)
- Dangerousness of the offender (Aharoni and Fridlund 2012; Carlsmith 2006; Goodwin and Benforado 2015; Keller et al. 2010)
- Self-control (Carlsmith 2006; Keller et al. 2010)
- Prior record (Carlsmith 2006)

Table 1. Retributive and utilitarian factors that have been used in previous research.

After identifying such factors, their role in people's punitive reactions is then investigated.

This can be done in several different ways. For instance, on the so-called *policy-capturing approach* (Cooksey 1996; for an overview, Carlsmith and Darley 2008), participants are asked to assign punishment to descriptions of criminal wrongdoing for which the levels of a subset of retributive and utilitarian factors are manipulated. For example, Darley, Carlsmith, and Robinson (2000, Exp. 1) manipulate the seriousness of the offense (a retributive factor) and the risk of offender recidivism (a utilitarian factor); they find a much larger effect on punishment severity of the former manipulation than of the latter. Similar results are reported by Carlsmith et al. (2002), Carlsmith (2008), and Aharoni and Fridlund (2012).

Along similar lines, Roberts and Gebotys (1989) ask participants to rate the seriousness of the offense (a retributive factor), the risk of offender recidivism, the likelihood that the offender could be rehabilitated, the general frequency of the offense and whether the frequency of offenses of this type was increasing or decreasing (four utilitarian factors). These ratings were then partially correlated with the severity of the punishment assigned by participants. A significant correlation was observed only for the retributive factor.

Another approach is *behavioral process-tracing* (Jacoby et al. 1987). The core idea of this approach is to learn something about people's motives for punishment by investigating their information-search behavior. Participants are given a choice between different items of information regarding a crime that they are put in charge of determining the punishment for. The items correspond to a subset of the factors in table 1. From the order in which participants request information, researchers make inferences about the importance of the corresponding punishment motive; the earlier an item is chosen, the more important that motive is thought to be for people's punitive reactions. For instance, Carlsmith (2006, Exp. 2) had participants choose from a list of nine items: magnitude of harm, offender intent, extenuating circumstances (three retributive items), dangerousness of the offender, prior record, offender self-control, frequency of the offense, detection rate, and publicity of offense and trial (six utilitarian items). He finds that participants overwhelmingly choose retributive items early on in the process. Similar results are reported by Keller et al. (2010).

The intuitive retributivism hypothesis

The body of behavioural findings we reviewed above suggests that retributivist motives largely dominate utilitarian motives in people's punitive reactions. This has led several authors to propose what we will call the *intuitive retributivism hypothesis*. The core idea is that when confronted with cases of criminal wrongdoing, people tend to have an intuitive reaction to punish which is motivated primarily by retributive factors. In contrast, utilitarian motives only play a subordinate role in this reaction, if they play any role in it at all. According to Aharoni and Fridlund (2012), for example, what "we call retribution may better be explained by heuristic processes rather than by abstract moral principles" (18). Keller et

al. (2010) suggest “an intuitive preference for retribution” (113). And Darley (2009) maintains that “the information processed when punishments are being determined [...] is information about what the transgressor justly deserves for the offense committed” (14).

The most detailed discussion of the intuitive retributivism hypothesis has been given by Carlsmith and Darley (2008). According to Carlsmith and Darley (2008), “punishment reactions are the product of a dual-process system in which the retributive desire is automatic, and the reasoning process that might override it is only selectively brought online” (215). Dual process models have been exceedingly popular in a variety of domains of psychological research. In the context of higher cognition, the central assumption of such models is that “cognitive tasks evoke two forms of processing that contribute to observed behavior” (Evans and Stanovich 2013, 225), with the two forms—type-I processing and type-II processing—being qualitatively distinct. Typically, the distinction is thought to line up roughly with the more familiar distinction between intuition and reflection or deliberation (cf. Carlsmith and Darley 2008, 217; also Darley 2009, 3-4). Different researchers differ considerably in how they spell out the distinctiveness of type-I and type-II processing (Evans 2008; Evans and Stanovich 2013). However, common attributes of type-I processing are that it is fast, parallel, automatic and does not require working memory; in contrast, type-II processing is often thought to be slow, serial, controlled, and to require working memory (Evans and Stanovich 2013, tab. 1; cf. Carlsmith and Darley 2008, 211-17; also Darley 2009, 4).

Carlsmith and Darley (2008) suggest that type-I and type-II processing differentially support different punishment motives. More specifically, they hypothesize that retributive reactions (punitive reactions responsive to retributive factors) are primarily the output of type-I processing. In contrast, utilitarian reactions (punitive reactions responsive to utilitarian factors) require type-II processing. When people are confronted with a case of criminal wrongdoing, type-I processing engages and swiftly outputs a retributive reaction. This reaction may sometimes be overridden by type-II processing, allowing for utilitarian motives to play a role (cf. Oswald and Stucki 2009). However, according to Carlsmith and Darley, this rarely happens. Thus, people’s punitive reactions are usually being determined by an initial intuition which skews heavily retributive.

Problem

The intuitive retributivism hypothesis with its dual-process framework provides a neat explanation for the findings we reviewed in the last section: If punitive reactions tend to be the output of type-I processing, which is primarily responsive to retributive factors and is only infrequently overwritten by more utilitarian type-II processing, then people’s punitive

reactions should primarily be retributive. This is indeed what the data suggests.

However, it is not the *only* explanation of the data. For one thing, a single-process model, for example along the lines of the rule-based account proposed by Kruglanski and Gigerenzer (2011), may too be able to capture the results. Even if we accept that a dual-process framework is helpful for understanding people's punitive reactions, however, alternative explanations to the intuitive retributivism hypothesis remain open. For instance, people may have both retributive and utilitarian intuitions, but the former simply tend to prevail. Alternatively, people's initial type-I punitive reactions may primarily be utilitarian but are routinely being overwritten by retributive type-II processing.

As far as we can see, the results we reviewed in the last section by themselves do not rule out any of these alternatives in favor of the intuitive retributivism hypothesis. Again, what these results suggest is that people's punitive reactions are primarily responsive to retributive, but not to utilitarian, factors. This by itself, however, does not imply anything about the details of the psychological mechanisms underlying this pattern. In order to establish the intuitive retributivism hypothesis, then, additional work is required.

In fact, a number of authors seem to be confident that this work has already been done. Darley (2009), for instance, advertises the intuitive retributivism hypothesis as "a relatively clear picture of the naive psychology of punishment" (2; also cf. Saulnier and Sivasubramaniam 2018, 195–96). Likewise, both Carlsmith and Darley (2008) and Robinson and Darley (2007) draw a number of policy implications from the hypothesis, suggesting that they, too, believe it to be reasonably securely established.

One argument to this effect is proposed by Carlsmith and Darley (2008, 211-17). Carlsmith and Darley point to research on dual-process models of moral cognition (e.g. Greene 2014; Haidt 2001; for a review, Guglielmo 2015), which they take to suggest that people's moral judgments are typically intuitive in nature (the output of type-I processing) and are only seldomly overwritten by deliberation or reasoning (type-II processing). This, they argue, also lends support to the intuitive retributivism hypothesis (also Darley 2009).

However, this strikes us as unpersuasive. Criticisms of dual-process accounts of moral judgment aside (e.g. Kahane 2012; Sauer 2012; Pizarro and Bloom 2003), the argument assumes that punitive reactions can straightforwardly be treated as moral judgments. Yet it is not clear that this is the case. Many legal scholars argue that while there is considerable overlap between the law and morality, there are nevertheless significant conceptual differences between the two domains (for discussion, Peczenik 2005, chap. 4). More importantly, there is research suggesting psychological differences between judgments of punishment and paradigmatic moral judgments such as rightness/wrongness and permissibility/impermissibility (Barbosa and Jiménez-Leal 2017; Cushman 2008). On the other hand, even though some researchers do sometimes seem to lump the two together

(e.g. Greene 2014, 705–6), we know of no evidence which would justify this. In light of these points, we believe that at this time, the extent to which dual-process accounts of moral judgment lend support to the intuitive retributivism hypothesis is questionable.

A second argument is suggested by Darley (2009, 3-8). Unfortunately, Darley does not state the argument quite as clearly as perhaps one would have liked; we understand his point to be this: Punitive reactions have been shown to be associated with emotion, particularly anger and outrage (e.g. Goldberg, Lerner, and Tetlock 1999; Nelissen and Zeelenberg 2009; Seip, van Dijk, and Rotteveel 2014). An association with emotion is a common feature of type-I processing (Evans 2008, 556-57). Therefore, there is some reason to believe that retributive punitive reactions are the output of type-I processing.

Note that this conclusion is fairly weak. The reason for this is that while emotion is indeed commonly associated with type-I processing, it is not a defining feature (Evans and Stanovich 2013). Clearly, not every process associated with emotion is type-I of some description (and conversely, not every type-I process is emotionally charged). Thus, even if sound, the argument would only lend limited support to the intuitive retributivism hypothesis.

As it stands, however, the argument is not sound. To support the claim that *retributive* (as opposed to utilitarian) punitive reactions are the primary output of type-I processing, it would need to be shown that retributive, but not utilitarian, punitive reactions are associated with emotion. Yet most research on the role of the emotions in punitive reactions that we know of does not meet this requirement. This is because such research tends to investigate punitive reactions generically, without controlling for or looking at what motivated individual participants to punish. Therefore, while it is possible that the connection between punitive reactions and anger/outrage is produced by a unique association of retributive punishment with emotion, at this time, there is no way to tell. The pattern may instead be due to a strong and unique connection of utilitarian punishment with anger/outrage, or by an association with anger/outrage of both retributive and utilitarian punitive reactions. In other words: From the fact that people's punitive reactions are emotionally charged, little can be inferred about the psychology of the underlying punishment motives as long as we do not know which motives were driving those reactions.

What is needed, then, is data on role of emotion in retributive vs. in utilitarian punitive reactions. We are aware of three results along those lines. Darley et al. (2000) report that self-reported moral outrage partially mediated the effect of both seriousness of the offense (a retributive factor) and risk of offender recidivism (a utilitarian factor) on punishment severity. However, the former relationship was stronger than the latter. Carlsmith et al. (2002) carried out a similar analysis; they found that self-reported moral outrage partially mediated only the influence of the two retributive factors manipulated in their study (seriousness of the offense, mitigating circumstances) on punishment severity. Lastly, Aharoni, Weintraub and Fridlund

(2007) report that punitive reactions of high psychopathy individuals are insensitive to manipulation of offender intent (a retributive factor), but not to manipulations of risk of offender recidivism (a utilitarian factor). Since high psychopathy individuals are often thought to be impaired with respect to moral emotion, this suggests a unique role of emotion in retributive punitive reactions (Aharoni, Weintraub and Fridlund 2007, 880).

Thus, while the majority of research linking emotion and punitive reactions does not bear on the intuitive retributivism hypothesis one way or the other, some evidence suggests an association of emotion and retributive (vs. utilitarian) punitive reactions. We agree that this finding is suggestive. Nevertheless, as we pointed out above, the extent to which it directly supports the intuitive retributivism hypothesis is quite limited. In particular, we believe it does not quite provide sufficient grounds to advertise the hypothesis as "a relatively clear picture of the naive psychology of punishment" (Darley 2009, 2).

There are other results which proponents of the intuitive retributivism hypothesis might similarly be tempted to point to for support (though we are not aware of anyone who has done so yet). First, Need for Cognition (Cacioppo and Petty 1982), an individual difference measure of cognitive style sometimes used to assess the tendency of individuals to engage in type-II processing (Petty et al. 2009), has been found to be negatively associated with punitiveness (Sargent 2004). Second, punitive reactions become more severe with cognitive load (van Knippenberg, Dijksterhuis, and Vermeulen 1999; Oswald and Stucki 2009; Gollwitzer et al. 2016 Exp. 1). Since type-II processing requires cognitive resources (in particular, working memory capacity) to a much greater extent than type-I processing, burdening those resources by inducing cognitive load is commonly used in experiments to inhibit type-II processing (Stanovich and Evans 2013, 232). Third, punitive reactions become less severe when participants are induced to think more carefully about their decision (Oswald and Stucki 2009; Gollwitzer et al. 2016, Exp. 2)—a manipulation which is thought to increase type-II processing effort (Stanovich and Evans 2013, 232).

While at first glance, these studies may certainly appear to bear on the intuitive retributivism hypothesis, we believe that this appearance is largely illusory. This is because they all share one crucial limitation: In all of them, people's punitive reactions were investigated only generically—that is, without controlling for or looking at the underlying punishment motives. Therefore, by the same argument we raised in our discussion of Darley (2009), little can be inferred about the mechanisms underlying retributive vs. utilitarian punitive reactions from these results.

In our view, the strongest evidence in favor of the intuitive retributivism hypothesis comes from Aharoni and Fridlund (2012). Aharoni and Fridlund (Exp. 2) show that punitive reactions are susceptible to dumbfounding. Participants read a description of a crime for which the efficacy of common utilitarian motives (cf. table 1) for punishment had been

minimized and were asked to recommend a punishment. Participants who did punish were then challenged to justify this decision. If a participant cited common utilitarian reasons, they were reminded that those considerations did not apply to the crime at hand. A majority of participants continued to recommend punishment even while admitting that no utilitarian reasons applied and while not being able to articulate other reasons for their decision. Aharoni and Fridlund take this to suggest that people's punitive reactions are "shaped more by heuristic than rational processes" (17), the former of which are primarily responsive to retributive factors.

This result certainly is suggestive. Nevertheless, there are a number of caveats. First, the study was only exploratory, and quite small (n = 47). Second, as Aharoni and Fridlund themselves point out, not all of their participants were dumbfounded, leaving open the possibility that a subset of punitive reactions was responsive to utilitarian factors (16-17). Third, a very similar argument has recently come under severe fire: Haidt, Björklund, and Murphy (2000) used semi-structured interviews to investigate people's reactions to harmless taboo violations. They report that for some violations, a majority of participants continued to maintain that the violation was wrong, even though they were unable to provide any reasons for this. Several authors have cited these results in support of a dual-process model of moral judgment (e.g. Haidt 2001; Haidt and Bjorklund 2007; Prinz 2006). However, this move has repeatedly been challenged, both on methodological (Royzman, Kim, and Leeman 2015) and on conceptual grounds (Hindriks 2015; Stanley, Yin, and Sinnott-Armstrong 2019). Due to the similarity of the designs of Aharoni and Fridlund (2012) and Haidt et al. (2000), however, such objections would also seem to call into question the extent to which dumbfounding can be used to support the intuitive retributivism hypothesis.

To summarize, while the intuitive retributivism hypothesis is an appealing explanation for the result that people's punitive reactions are primarily responsive to retributive (as opposed to utilitarian) factors, this alone does not establish it as psychological fact. Instead, additional work is required. We have reviewed a number of arguments to this effect, but have argued that they are either unconvincing or provide only limited support for the intuitive retributivism hypothesis. We conclude that previous work has overstated the evidential basis of the intuitive retributivism hypothesis.

Hypothesis, aims and objectives

In the last section, we argued that the evidential basis of the intuitive retributivism hypothesis is less secure than several authors had previously suggested. Thus, there is a need for direct empirical investigation. This is what this paper hopes to provide (or at least to get rolling).

The intuitive retributivism hypothesis is a claim about the processes driving retributive vs. utilitarian punitive reactions. Therefore, in order to test it, we will need two ingredients. The first is a method of investigating people's punitive reactions in a way that measures the underlying punishment motives. As we discussed in our review of the relevant scholarship, the existing literature provides several options. Here, we will use the information search task (IST) approach of Keller et al. (2010, Exp. 2). Recall that Keller et al. put participants in charge of assigning punishment to an offender guilty of a crime. To inform their decision, participants were asked to pick five pieces of information about the crime from a list. Some items were retributivism-related; others were related to deterrence or incapacitation, two main concerns of utilitarian theories of punishment (see table 1). The earlier an item was selected, the more important the corresponding motive (retributivism, deterrence, incapacitation) was interpreted to be in participants' punitive reactions.

The second ingredient is a way of manipulating the type of processing producing participants' punitive reactions. Again, there are a number of existing approaches to choose from (for a review, Horstmann, Hausmann, and Ryf 2010). Here, we increase type-II processing effort by inducing participants to think carefully about their decisions. This manipulation is frequently used in this way in research on dual-process models, both outside of (e.g. Evans et al. 2010; Shenhav, Rand, and Greene 2012) and within moral psychology (e.g. Oswald and Stucki 2009; Rand, Greene, and Nowak 2012).

Putting the two ingredients together, all participants will complete the IST from Keller et al. (2010, Exp. 2) while being randomly assigned to one of two conditions. In the treatment condition, participants will be induced to think carefully about each of their item selections. To this end, we will explicitly instruct participants to only make selections after thorough deliberation and to take their time. In contrast, in the treatment condition, participants will not be given any special instruction or motivation. Thus, the control condition amounts to a conceptual replication of Keller et al. (2010, Exp. 2).

The main measure of interest is the order in which retributive and utilitarian items are selected. Following Keller et al. (2010), to capture this order, we will calculate a rank-preference score for each participant and punishment motive (retributivism, deterrence, incapacitation). Each item selection trial will be weighted. The first trial will receive a weight of 5; the second trial will receive a weight of 4; and so on. For a given punishment motive, its rank-preference score will then be calculated as the sum of the trial weights for which an item related to that motivation was selected. For example, if a participant chooses retributive items on the first, third and fourth trials, the retributivism rank-preference score will be 5 + 3 + 2 = 10.

Now for predictions. Recall that according to the intuitive retributivism hypothesis, two different types of processing contribute to punitive reactions. Type-I processing (intuition)

primarily outputs retributive punitive reactions; in contrast, utilitarian punitive reactions require type-II processing (reflection/deliberation). Furthermore, in normal circumstances, people largely rely on type-I processing, and so their punitive reactions skew heavily retributive. We will follow Keller et al. (2010; also Carlsmith 2006) in assuming that our control condition (a replication of their Exp. 2) represents more or less normal judgment conditions. Thus, the intuitive retributivism hypothesis makes the following two predictions:

h1a. In the control condition, participants will have higher retributivism rank-preference scores than deterrence rank-preference scores.

h1b. In the control condition, participants will have higher retributivism rank-preference scores than incapacitation rank-preference scores.

To the extent that our manipulation is successful in increasing type-II processing effort, the intuitive retributivism hypothesis suggests that more participants in the treatment than in the control condition will override their initial retributive intuitions in favour of more utilitarian punitive reactions. Therefore, the importance of retributive items should decrease relative to the control condition, while the importance of deterrence and incapacitation items should increase. In other words:

h2a. In the control condition, participants will have higher retributivism rank-preference scores than in the treatment condition.

h2b. In the control condition, participants will have lower deterrence rank-preference scores than in the treatment condition.

h2c. In the control condition, participants will have lower incapacitation rank-preference scores than in the treatment condition.

Study

Materials and Methods

Data collection

Data collection will take place online. We are planning to use the ZPID's PsychLab online (https://leibniz-psychology.org/en/services/data-collection/psychlab-online/) for data collection, which purchases samples for online studies from commercial panel providers, such as respondi (https://www.respondi.com) or Cint (https://www.cint.com/).

Sample size

We plan to recruit a total sample of n = 559 participants. Our analysis is going to be carried

out using linear mixed-models. Thus, we choose a simulation-based approach to power analysis, applied to the model in (2) (Brysbaert and Stevens 2018). The smallest mean difference between retributivism rank-preference score and one of the two utilitarian rank-preference scores reported by Keller et al. (2010, Exp. 2) was 3.18; we thus conservatively choose fixed effect sizes of $\beta = 2.0$ for the two contrasts corresponding to h1a and h1b. For each of the contrasts corresponding to the remaining three hypotheses, we estimated the smallest effect sizes of interest (Albers and Lakens 2018) to be $\beta = 1.5$.

The power analysis was run using the simr package (Green and MacLeod 2016, nsim = 2000). Using the estimates $\sigma^2_u = 1.0$, $\sigma^2_v = 1.0$ and $\sigma^2_\varepsilon = 4.0$, it indicates that in order to detect the fixed effects specified above at a level of significance of $\alpha = 0.01$ (Bonferroni corrected) with power of at least 90% (Chambers et al. 2019), a sample of size $n > 485$ participants is required. The code used for the power analysis is available at: [Link omitted for anonymous review]. We estimate that 15% of participants will be excluded due to attention check failure (see Data exclusion criteria). To account for this, we plan to recruit a total sample of $n = 559$.

Stopping rule

Data collection will be stopped once 559 participants have completed all study materials.

Participant characteristics

Participants will be recruited from the pool of whichever panel provider the ZPID's PsychLab online is going to use for data collection. The exact panel characteristics may differ between providers; however, all providers must adhere to the standards for online access panels determined by ISO 20252:2009 or ISO 20252:2019. Individuals will be considered eligible for participation if their first language is English and they have at least a 95% approval rate on previous submission (provided the panel provider in question records this information).

Procedure

Eligible participants will receive an invitation email containing a link to the survey. Upon accepting to participate, participants will be redirected to the study, which is going to be hosted on LimeSurvey (https://www.limesurvey.org). After giving informed consent, participants will receive the study instructions. Following instructions, participants will complete the study materials. At the end of this, participants will be thanked for and compensated for their participation. Incentives for participation are provided in the form of tokens or bonus points which, after a certain amount has been accumulated, participants can either have paid out to them or donate.

Materials

All materials described in this section are available online at the following link: [Link omitted for anonymous review].

Items of information

We extend Keller et al. (2010, Exp. 2) by including most of the items for each of the factors listed in table 1 (this includes all of the items used by Keller et al.). We hope that this will help us provide a more general test of the intuitive retributivism hypothesis (cf. Keller et al. 2010, Exp. 3). Below are the item descriptions out of which participants will make their selections. Items (r1)-(r4) are retributive; the remaining items are utilitarian, with items (d1)-(d4) relating to deterrence and items (i1)-(i4) relating to incapacitation.

(r1) *Seriousness of the offense*: How serious is this particular offense?

(r2) *Magnitude of harm*: How severe is the financial, physical or psychological harm that the offender has caused?

(r3) *Offender intent*: Did the offender act with intention?

(r4) *Extenuating circumstances*: Are there aspects of this particular offense that make the offender less than fully responsible?

(d1) *Publicity of offense and trial*: Will this particular offense and its trial attract a lot of public attention?

(d2) *Detection rate*: How frequently are offenses like this detected and brought to trial?

(d3) *Frequency of the offense*: How frequently do offenses like this occur in society?

(d4) *Frequency trend*: Is the frequency of offenses like this in society increasing or decreasing?

(i1) *Risk of offender recidivism*: How likely is it that the offender will commit further offenses?

(i2) *Dangerousness of the offender*: How dangerous is the offender?

(i3) *Self-control*: Does the offender normally have good self-control or does the offender frequently act on impulse?

(i4) *Prior record*: Does the offender have a prior criminal record?

After having made their selection and before making their punishment decision, participants will receive a brief answer to each item of information they requested. For example, if a participant requests offender intent, they will be informed that the offender had been planning their offense for several days. If a participant requests information about the risk of offender recidivism, they will be informed that the offender has publicly stated that he or she would repeat their offense if given the chance. As these answers only impact participants' punitive reactions, which themselves will not be analyzed in this study (see

Analysis Plan), we do not include all answers here (they will, however, be available online).

Conditions and design

The study is an experimental study using randomized control trials. Participants will be randomly assigned to either the control or the treatment condition. All participants will be asked to read a short prompt informing them that a crime has been committed, that the offender has been caught and that their task will be to assign a punishment. The type of crime will be randomly chosen from a set of five crime types (blackmail, stolen property, arson, aggravated assault, murder). This extension of Keller et al. (2010, Exp. 2) was included to improve the generalizability of the results. This, as well as all other randomizations used in this study, will be performed by the survey software, LimeSurvey.

Following this prompt, all participants will be presented with a list of descriptions of pieces of information about the crime (see Materials). The ordering of this list will be randomized. Participants will be asked to select the items of information from this list which they consider most relevant for making their punishment decision one at a time. In order to select an item, participants will click on it, followed by clicking on a button labeled “Select item” below the list to confirm the selection. Participants will be instructed to select items in order of priority and will not be made aware about how many items they will be able to select in total. This selection procedure will be repeated six times, with the first trial being an attention check. In addition, participants in the treatment condition will be instructed to think carefully before each of their selections, and to take their time. In contrast, participants in the control condition will not receive any further instructions.

After having selected their five items, all participants will receive the information they requested and will indicate their punishment decision. After completing this task, participants will be asked to provide demographic information. At the end of the survey, participants will be thanked for their participation, and will exit the survey.

Variables

All instruments described in this section are available online at the following link: [Link omitted for anonymous review].

Punishment

To measure participant’s punitive reaction, we will use an instrument commonly employed in the literature for this purpose (e.g. Carlsmith et al. 2002; Aharoni and Fridlund 2012). It consists of three items:

(p1) How severe a punishment should be given for this particular offense? (“Not at all severe” to “Very severe”)

(p2) How much should [offender] suffer for this particular offense? (“Not at all” to “Very much”)

(p3) What would be an appropriate sentence for this particular offense?

For the first two items, participants will indicate their answer on continuous sliding scales, ranging from 0 to 100, labeled at both endpoints. Endpoint labels are shown in parenthesis after each item. For the third item, participants will pick their answer from an ordered list of eleven options: “0 days”, “1 day”, “2 weeks”, “2 months”, “6 months”, “1 year”, “3 years”, “7 years”, “15 years”, “30 years”, and “Life”.

Experimental manipulation

Our manipulation seeks to increase type-II processing effort by inducing participants to think carefully about their decisions. To this end, as part of the study instructions, participants will be instructed to only select each item after thorough deliberation and to take their time for each selection (cf. Horstmann, Hausmann, and Ryf 2010). In contrast, participants in the control condition will not receive any further instructions.

Manipulation check

For each item selection, we will record the time from first viewing the list of items to clicking “Select item”. This will be implemented in LimeSurvey. If our manipulation is successful, we expect longer item selection timings in the treatment condition than in the control condition (Horstmann, Hausmann, and Ryf 2010).

Demographic questions

We will include a series of standard demographic questions. Participants will be asked to report their age, gender, ethnicity, religiosity and political attitudes. Moreover, participants will be asked to indicate whether they have ever taken an ethics course and a law course.

Analysis Plan

Preprocessing

For participant i, their retributivism rank-preference score (RPS-R_i) will be defined as

$$\text{RPS-}R_i = 5\delta_{i1} + 4\delta_{i2} + 3\delta_{i3} + 2\delta_{i4} + 1\delta_{i5} ,$$

where $\delta_{ij} = 1$ if the j^{th} item participant i requested was a retributive item, and $\delta_{ij} = 0$ otherwise.

Deterrence (RPS-D_i) and incapacitation rank-preference scores (RPS-I_i) will be calculated in the same way.

Data exclusion criteria

Our design includes an instructional attention check (Oppenheimer, Meyvis, and Davidenko 2009). As part of the study instructions, participants will be told to select a specific item on the first trial. This is done to make it less likely that participants are included in the analysis who do not read the instructions conscientiously. Failure to select the specified item will result in the exclusion of that participant's data from all analysis. Moreover, any participants with missing data for any of the measured variables will be excluded from all analysis. There will be no further exclusion criteria.

Tests

Analysis is going to be carried out in R (R Core Team 2020) using linear mixed-effects models (Bates et al. 2015). P-values will be computed using Satterthwaite's approximation of denominator degrees of freedom (Kuznetsova, Brockhoff, and Christensen 2017). Models will be fit using ML. Following the recommendation of Baayen, Davidson, and Bates (2008), all models will include random intercepts both for participant and for type of crime (blackmail, stolen property, arson, aggravated assault, murder).

To check our manipulation, we will enter condition (control, treatment) into a model predicting item selection time:

$$Time_{ij} = \beta_0 + \beta_1 Condition_{ij} + u_{0i} + v_{0j} + \varepsilon_{ij}$$

Here, i indexes participant, j type of crime, $u_{0i} \sim N(0, \sigma^2_u)$, $u_{0j} \sim N(0, \sigma^2_v)$, $\varepsilon_{ij} \sim N(0, \sigma^2_\varepsilon)$.

For our main analysis, we will perform a planned contrast analysis. We closely follow the recommendations of Schad et al. (2020). We first combine condition and motive (retributivism, deterrence, incapacitation) into one factor with six levels. We call this factor ConditionxMotive. We label the combination of condition = control and motive = retributivism "CR"; the combination of condition = treatment and motive = deterrence "TD"; etc. The nulls corresponding to our five hypotheses (h1a-h2c) can then be expressed as contrasts of group means indexed by the levels of ConditionxMotive. For example, the null corresponding to h1a can be expressed as:

$$1 \cdot \mu_{CR} + (-1) \cdot \mu_{CD} + 0 \cdot \mu_{CI} + 0 \cdot \mu_{TR} + 0 \cdot \mu_{TD} + 0 \cdot \mu_{TI} = 0 \quad (1)$$

The μ_x are the mean rank-preferences scores of participants in group x. For example, μ_{CR} is the mean retributive rank-preference score in the control condition; μ_{TD} is the mean deterrence rank-preference score in the treatment condition; etc. In other words, then, (1) states that in the control condition, there will be no difference between the mean retributivism and mean deterrence rank-preference score.

Once expressed in this way, we want to test the contrasts. However, if we were to simply enter ConditionxMotive into a linear mixed-effects model predicting RPS, then because R uses treatment contrasts by default, the model would not evaluate the comparisons that we are after. Therefore, we need to tell the model which contrasts we would like it use. To do this, we first extract all the contrast coefficients and combine them in the following matrix:

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix}$$

Each column of this matrix contains the contrast coefficients of one null hypothesis. For example, comparison with (1) shows that the first column contains the coefficients corresponding to h1a. The second column contains the coefficients corresponding to h1b; the third column contains the coefficients corresponding to h2a; etc. This matrix is then converted into what Schad et al. (2020) call a contrast matrix by applying the generalized matrix inverse. This contrast matrix has the correct format to specify the desired contrasts for a factor in R.

$$\begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} \\ -\frac{2}{3} & \frac{1}{3} & \frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} \\ \frac{1}{3} & -\frac{2}{3} & \frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} \\ \frac{1}{3} & \frac{1}{3} & -\frac{5}{6} & -\frac{1}{6} & -\frac{1}{6} \\ -\frac{2}{3} & \frac{1}{3} & \frac{1}{6} & \frac{5}{6} & -\frac{1}{6} \\ \frac{1}{3} & -\frac{2}{3} & \frac{1}{6} & -\frac{1}{6} & \frac{5}{6} \end{pmatrix}$$

Since the contrast matrix has full rank, we can test all five contrasts in the same model. To this end, we enter ConditionxMotive into a model predicting RPS and specify our contrast matrix in the process:

$$RPS_{ij} = \beta_0 + \beta_1 ConditionxMotive_{h1a,ij} + \beta_2 ConditionxMotive_{h1b,ij} + \beta_3 ConditionxMotive_{h2a,ij} + \beta_4 ConditionxMotive_{h2b,ij} + \beta_5 ConditionxMotive_{h2c,ij} + u_{0i} + v_{0j} + \varepsilon_{ij} \quad (2)$$

Again, i indexes participant, j type of crime, $u_{0i} \sim N(0, \sigma^2_u)$, $u_{0j} \sim N(0, \sigma^2_v)$, $\varepsilon_{ij} \sim N(0, \sigma^2_\varepsilon)$. Moreover, h1a-h2c indicate the contrast being evaluated. Thus, for example, estimating β_1 will provide a test of h1a; estimating β_2 will provide a test of h1b; etc.

Exploratory analysis

To explore the role of individual retributive and utilitarian factors in people's punitive reactions, and in particular how they might change depending on the type of processing primarily underlying them, we will calculate overall rank-preference scores for each item in both conditions, and then compare them. No further exploratory analyses are planned.

References

Aharoni, Eyal, and Alan J. Fridlund. 2012. "Punishment without Reason: Isolating Retribution in Lay Punishment of Criminal Offenders." *Psychology, Public Policy, and Law* 18 (4): 599–625.

Aharoni, Eyal, Lisa L. Weintraub, and Alan J. Fridlund. 2007. "No Skin off My Back: Retribution Deficits in Psychopathic Motives for Punishment." *Behavioral Sciences & the Law* 25 (6): 869–89.

Albers, Casper, and Daniël Lakens. 2018. "When Power Analyses Based on Pilot Data Are Biased: Inaccurate Effect Size Estimators and Follow-up Bias." *Journal of Experimental Social Psychology* 74 (January): 187–95.

Baayen, Harald, Doug Davidson, and Douglas Bates. 2008. "Mixed-Effects Modeling with Crossed Random Effects for Subjects and Items." *Journal of Memory and Language* 59 (4): 390–412.

Barbosa, Sergio, and William Jiménez-Leal. 2017. "It's Not Right but It's Permitted: Wording Effects in Moral Judgement." *Judgment and Decision Making* 12 (3): 308–313.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using Lme4." *Journal of Statistical Software* 67 (1).

Bentham, Jeremy. (1830) 1998. "The Rationale of Punishment." In *Strafrechtsdenker Der Neuzeit*, edited by Thomas Vormbaum, 387–94. Berliner Wissenschafts-Verlag.

Brysbaert, Marc, and Michaël Stevens. 2018. "Power Analysis and Effect Size in Mixed Effects Models: A Tutorial." *Journal of Cognition* 1 (1).

Cacioppo, John T., and Richard E. Petty. 1982. "The Need for Cognition." *Journal of Personality and Social Psychology* 42 (1): 116–31.

Carlsmith, Kevin M. 2006. "The Roles of Retribution and Utility in Determining Punishment." *Journal of Experimental Social Psychology* 42 (4): 437–51.

———. 2008. "On Justifying Punishment: The Discrepancy Between Words and Actions." *Social Justice Research* 21 (2): 119–37.

Carlsmith, Kevin M., and John M. Darley. 2008. "Psychological Aspects of Retributive Justice." In *Advances in Experimental Social Psychology*, 40:193–236.

Carlsmith, Kevin M., John M. Darley, and Paul H. Robinson. 2002. "Why Do We Punish?: Deterrence and Just Deserts as Motives for Punishment." *Journal of Personality and Social Psychology* 83 (2): 284–99.

Chambers, Chris, George Christopher Banks, Dorothy Bishop, Sara Bowman, Kate Button,

Molly Crockett, Zoltan Dienes, Tim Errington, Agneta Fischer, and Alex O. Holcombe. 2019. “Registered Reports.” *OSF*.
Cooksey, Ray W., ed. 1996. *Judgment Analysis: Theory, Methods, and Applications*. Emerald Publishing.
Cullen, Francis T., Bonnie S. Fisher, and Brandon K. Applegate. 2000. “Public Opinion about Punishment and Corrections.” *Crime and Justice* 27 (January): 1–79.
Cushman, Fiery. 2008. “Crime and Punishment: Distinguishing the Roles of Causal and Intentional Analyses in Moral Judgment.” *Cognition* 108 (2): 353–80.
Darley, John M. 2009. “Morality in the Law: The Psychological Foundations of Citizens’ Desires to Punish Transgressions.” *Annual Review of Law and Social Science* 5 (1): 1–23.
Darley, John M., Kevin M. Carlsmith, and Paul H. Robinson. 2000. “Incapacitation and Just Deserts as Motives for Punishment.” *Law and Human Behavior* 24 (6): 659–83.
Dolinko, David. 2011. “Punishment.” In *The Oxford Handbook of the Philosophy of the Criminal Law*, edited by John Deigh and David Dolinko, 403–40. Oxford University Press.
Doorn, Janne van, and Lieve Brouwers. 2017. “Third-Party Responses to Injustice: A Review on the Preference for Compensation.” *Crime Psychology Review* 3 (1): 59–77.
Duff, Antony, and Zachary Hoskins. 2019. “Legal Punishment.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2019. Metaphysics Research Lab, Stanford University.
Evans, Jonathan. 2008. “Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition.” *Annual Review of Psychology* 59: 255–78.
Evans, Jonathan, Simon J. Handley, Helen Neilens, and David Over. 2010. “The Influence of Cognitive Ability and Instructional Set on Causal Conditional Inference.” *Quarterly Journal of Experimental Psychology* 63 (5): 892–909.
Evans, Jonathan, and Keith E. Stanovich. 2013. “Dual-Process Theories of Higher Cognition: Advancing the Debate.” *Perspectives on Psychological Science* 8 (3): 223–41.
Funk, Friederike, Victoria McGeer, and Mario Gollwitzer. 2014. “Get the Message: Punishment Is Satisfying If the Transgressor Responds to Its Communicative Intent.” *Personality and Social Psychology Bulletin* 40 (8): 986–997.
Goldberg, Julie H., Jennifer S. Lerner, and Philip E. Tetlock. 1999. “Rage and Reason: The Psychology of the Intuitive Prosecutor.” *European Journal of Social Psychology* 29 (5–6): 781–95.
Gollwitzer, Mario, Judith Braun, Friederike Funk, and Philipp Süssenbach. 2016. “People as Intuitive Retaliators: Spontaneous and Deliberate Reactions to Observed Retaliation.” *Social Psychological and Personality Science* 7 (6): 521–29.
Gollwitzer, Mario, Milena Meder, and Manfred Schmitt. 2011. “What Gives Victims Satisfaction When They Seek Revenge?” *European Journal of Social Psychology* 41 (3): 364–74.
Goodwin, Geoffrey P., and Dena M. Gromet. 2014. “Punishment.” *Wiley Interdisciplinary Reviews: Cognitive Science* 5 (5): 561–72.
Green, Peter, and Catriona J. MacLeod. 2016. “SIMR : An R Package for Power Analysis of Generalized Linear Mixed Models by Simulation.” *Methods in Ecology and Evolution* 7 (4): 493–98.
Greene, Joshua D. 2014. “Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics.” *Ethics* 124 (4): 695–726.
Guglielmo, Steve. 2015. “Moral Judgment as Information Processing: An Integrative Review.” *Frontiers in Psychology* 6 (October): 1637.
Haidt, Jonathan. 2001. “The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment.” *Psychological Review* 108 (4): 814–34.
Haidt, Jonathan, and Fredrik Bjorklund. 2007. “Social Intuitionists Answer Six Questions about Moral Psychology.” In *Moral Psychology: The Cognitive Science of Morality: Intuition and Diversity*, edited by Walter Sinnott-Armstrong, 2:181–217. MIT Press.

Haidt, Jonathan, Fredrik Björklund, and Scott Murphy. 2000. “Moral Dumbfounding: When Intuition Finds No Reason.” Unpublished manuscript.
Henrich, Joseph, Richard McElreath, Abigail Barr, Jean Ensminger, Clark Barrett, Juan Camilo Cardenas, Michael Gurven, et al. 2006. “Costly Punishment Across Human Societies.” *Science* 312 (5781): 1767–70.
Hindriks, Frank. 2015. “How Does Reasoning (Fail to) Contribute to Moral Judgment? Dumbfounding and Disengagement.” *Ethical Theory and Moral Practice* 18 (2): 237–50.
Hoffman, Morris B., and Timothy H. Goldsmith. 2004. “The Biological Roots of Punishment.” *Ohio State Journal of Criminal Law* 1 (2): 627–42.
Horstmann, Nina, Daniel Hausmann, and Stefan Ryf. 2010. “Methods for Inducing Intuitive and Deliberate Processing Modes.” In *Foundations for Tracing Intuition: Challenges and Methods*, 219–37. Psychology Press.
Jacoby, Jacob, James Jaccard, Alfred Kuss, Tracy Troutman, and David Mazursky. 1987. “New Directions in Behavioral Process Research: Implications for Social Psychology.” *Journal of Experimental Social Psychology* 23 (2): 146–75.
Kahane, Guy. 2012. “On the Wrong Track: Process and Content in Moral Psychology: Process and Content in Moral Psychology.” *Mind & Language* 27 (5): 519–45.
Kant, Immanuel. (1785) 1998. *Grundlegung zur Metaphysik der Sitten*. Edited by Frank-Peter Hansen. directmedia.
Keller, Livia B., Margit E. Oswald, Ingrid Stucki, and Mario Gollwitzer. 2010. “A Closer Look at an Eye for an Eye: Laypersons’ Punishment Decisions Are Primarily Driven by Retributive Motives.” *Social Justice Research* 23 (2–3): 99–116.
Knippenberg, Ad van, Ap Dijksterhuis, and Diane Vermeulen. 1999. “Judgement and Memory of a Criminal Act: The Effects of Stereotypes and Cognitive Load.” *European Journal of Social Psychology* 29 (2–3): 191–201.
Kruglanski, Arie W., and Gerd Gigerenzer. 2011. “Intuitive and Deliberate Judgments Are Based on Common Principles.” *Psychological Review* 118 (1): 97–109.
Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen. 2017. “LmerTest Package: Tests in Linear Mixed Effects Models.” *Journal of Statistical Software* 82 (13).
Nadelhoffer, Thomas, Saeideh Heshmati, Deanna Kaplan, and Shaun Nichols. 2013. “Folk Retributivism and the Communication Confound.” *Economics and Philosophy* 29 (2): 235–61.
Nelissen, Rob, and Marcel Zeelenberg. 2009. “Moral Emotions as Determinants of Third-Party Punishment: Anger, Guilt and the Functions of Altruistic Sanctions.” *Judgment and Decision Making* 4 (7): 543–53.
Oppenheimer, Daniel M., Tom Meyvis, and Nicolas Davidenko. 2009. “Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power.” *Journal of Experimental Social Psychology* 45 (4): 867–72.
Oswald, Margit E., and Ingrid Stucki. 2009. “A Two-Process Model of Punishment.” In *Social Psychology of Punishment of Crime*, edited by Margit E. Oswald, Steffen Bieneck, and Jorg Hupfeld-Heinemann. Wiley.
Peczenik, Aleksander. 2005. *Scientia Juris: Legal Doctrine as Knowledge of Law and as a Source of Law*. A Treatise of Legal Philosophy and General Jurisprudence. Springer.
Petty, Richard E., Pablo Briñol, Chris Loersch, and Michael J. McCaslin. 2009. “The Need for Cognition.” In *Handbook of Individual Differences in Social Behavior*, edited by Mark R. Leary and Rick H. Hoyle, 318–29. Guilford Press.
Pizarro, David A., and Paul Bloom. 2003. “The Intelligence of the Moral Intuitions: A Comment on Haidt (2001).” *Psychological Review* 110 (1): 193–96.
Prinz, Jesse. 2006. “The Emotional Basis of Moral Judgments.” *Philosophical Explorations* 9 (1): 29–43.
Prooijen, Jan-Willem van. 2010. “Retributive versus Compensatory Justice: Observers’ Preference for Punishing in Response to Criminal Offenses.” *European Journal of Social Psychology*, no. 40: 72–85.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.
Rand, David G., Joshua D. Greene, and Martin A. Nowak. 2012. “Spontaneous Giving and Calculated Greed.” *Nature* 489 (7416): 427–30.
Roberts, Julian V., and Robert J. Gebotys. 1989. “The Purposes of Sentencing: Public Support for Competing Aims.” *Behavioral Sciences & the Law* 7 (3): 387–402.
Roberts, Julian V., and Loretta J. Stalans. 2004. “Restorative Sentencing: Exploring the Views of the Public.” *Social Justice Research* 17 (3): 315–34.
Robinson, Paul H., and John M. Darley. 2007. “Intuitions of Justice: Implications for Criminal Law and Justice Policy.” *Southern California Law Review* 81 (1): 1–69.
Royzman, Edward B., Kwanwoo Kim, and Robert F. Leeman. 2015. “The Curious Tale of Julie and Mark: Unraveling the Moral Dumbfounding Effect.” *Judgment and Decision Making* 10 (4): 18.
Sargent, Michael J. 2004. “Less Thought, More Punishment: Need for Cognition Predicts Support for Punitive Responses to Crime.” *Personality and Social Psychology Bulletin* 30 (11): 1485–93.
Sauer, Hanno. 2012. “Morally Irrelevant Factors: What’s Left of the Dual Process-Model of Moral Cognition?” *Philosophical Psychology* 25 (6): 783–811.
Saulnier, Alana, and Diane Sivasubramaniam. 2018. “Restorative Justice: Reflections and the Retributive Impulse.” In *Advances in Psychology and Law*, edited by Monica K. Miller and Brian H. Bornstein, 3:177–210. Springer International Publishing.
Schad, Daniel J., Shravan Vasishth, Sven Hohenstein, and Reinhold Kliegl. 2020. “How to Capitalize on a Priori Contrasts in Linear (Mixed) Models: A Tutorial.” *Journal of Memory and Language* 110 (February): 104038.
Seip, Elise C., Wilco W. van Dijk, and Mark Rotteveel. 2014. “Anger Motivates Costly Punishment of Unfair Behavior.” *Motivation and Emotion* 38 (4): 578–588.
Shenhav, Amitai, David G. Rand, and Joshua D. Greene. 2012. “Divine Intuition: Cognitive Style Influences Belief in God.” *Journal of Experimental Psychology: General* 141 (3): 423–28.
Stanley, Matthew L., Siyuan Yin, and Walter Sinnott-Armstrong. 2019. “A Reason-Based Explanation for Moral Dumbfounding.” *Judgment and Decision Making* 14 (2): 120–29.
Wood, David. 2010. “Punishment: Consequentialism.” *Philosophy Compass* 5 (6): 455–69.