COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 39 (2015) 619–646

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12175

Judging the Goring Ox: Retribution Directed Toward Animals

Geoffrey P. Goodwin,[a] Adam Benforado[b]

[a]*Department of Psychology, University of Pennsylvania*
[b]*Drexel University School of Law*

Received 11 September 2013; received in revised form 26 January 2014; accepted 3 February 2014

Abstract

Prior research on the psychology of retribution is complicated by the difficulty of separating retributive and general deterrence motives when studying human offenders (Study 1). We isolate retribution by investigating judgments about punishing animals, which allows us to remove general deterrence from consideration. Studies 2 and 3 document a "victim identity" effect, such that the greater the perceived loss from a violent animal attack, the greater the belief that the culprit deserves to be killed. Study 3 documents a "targeted punishment" effect, such that the responsive killing of the actual "guilty" culprit is seen as more deserved than the killing of an almost identical yet "innocent" animal from the same species. Studies 4 and 5 extend both effects to participants' acceptance of inflicting pain and suffering on the offending animal at the time of its death, and show that both effects are mediated by measures of retributive sentiment, and not by consequentialist concerns.

Keywords: Retribution; Punishment; Moral judgment; Anthropomorphization; Mental states; Causation

1. Introduction

1.1. Theories of punishment

Across the ages, scholars have devoted considerable attention to the issue of punishment not only because it is pervasive in human societies but also because of the enormous impact punishment has at the individual and community level. Philosophers have been primarily concerned with developing *normative* theories of punishment. The justifications they have developed for why we should punish have tended to fall into two major

Correspondence should be sent to Geoffrey P. Goodwin, Department of Psychology, University of Pennsylvania, 3720 Walnut Street, Philadelphia, PA 19104-6241. E-mail: ggoodwin@psych.upenn.edu

categories. On the one side are backward-looking retributive justifications, associated most notably with the work of Immanuel Kant (1790/1952), who asserted that transgressors deserved punishment because they committed morally wrong actions that harmed others. For Kant, it was the corrupted disposition of the offender that required the meting out of just deserts in proportion to the bad act. Accordingly, under a retributive theory, the greater the moral outrage that an action elicits, the greater the punishment ought to be (Carlsmith & Darley, 2008).

On the other side are forward-looking consequentialist justifications that focus on punishment's value in reducing transgressions in the future (Beccaria, 1764/1963; Jeremy Bentham, 1843/1962; John Stuart Mill, 1871/1998). Consequentialist perspectives have characterized the benefits of punishment with respect to (a) incapacitation, aimed at physically preventing the actual malefactor from repeat offending; (b) specific deterrence, aimed at deterring the malefactor from repeat offending; and (c) general deterrence, aimed at deterring other would-be wrongdoers from committing similar crimes. In addition, theorists have asserted that punishment can provide restitution to victims (e.g., Darley & Pittman, 2003), and serve as a means of rehabilitation, among other functions (Baron & Ritov, 2009).

However, separate from the normative question of how to justify punishment is a *descriptive* question about what actually motivates people's desire for punishment. We may report that a law is designed primarily to deter and incapacitate dangerous individuals, or that we support the death penalty for similar reasons, but what actually drives our desires for punishment?

1.2. Existing psychological research

While individuals frequently offer utilitarian rationales to justify inflicting penalties on perceived transgressors, several researchers have argued that people's behavior is largely motivated by retributive concerns (Carlsmith, 2008; Carlsmith, Darley, & Robinson, 2002). For example, Carlsmith et al. (2002) presented participants with descriptions of different offenses and asked them to provide corresponding punishments. Across different offense descriptions, the researchers manipulated factors that were pre-tested as differentially relevant either to retribution (e.g., offense seriousness, the presence of extenuating circumstances) or to deterrence (e.g., probability of detection, publicity of punishment). The punishments that participants selected were found to be far more sensitive to factors related to retribution than to factors related to deterrence; indeed, the punishment choices displayed very little sensitivity to deterrence related factors. Yet, when asked explicitly, participants expressed high endorsement of both retributive and deterrence justifications of punishment. In later work, Carlsmith (2006) showed further that, when making punishment decisions, individuals tend to *seek* information that is arguably more relevant to retribution than to deterrence.

These studies have been valuable in advancing our understanding of punishment motives, but they do not provide definitive evidence of the existence of a purely retributive motive in punishment. The seemingly insurmountable problem for all research in this area has been that it is difficult to isolate the variables that pertain solely to retribution in

the context of punishing human offenders. In Carlsmith et al.'s (2002) study, participants were more inclined to classify factors such as crime severity, intent, and extenuating circumstances as relevant to retribution than to deterrence (although see Study 1, presently, for countervailing evidence), and they were highly sensitive to them in their punishment decisions. But this neglects the fact that crime severity, intent, and extenuating circumstances are also highly relevant to deterrence and incapacitation: It makes sense that members of society should want to both deter and incapacitate individuals who have intentionally committed serious offenses, without extenuating circumstances, and to deter similar would-be offenders. Indeed, as we document in our first study, factors such as the magnitude of the harm caused and the perpetrator's motivations—which have been taken by past researchers as unambiguously reflecting retributive motives—are rated just as relevant to deterrence concerns as they are to retributive concerns. As a consequence, it is unclear whether participants' sensitivity to information concerning the egregiousness of the crime and mental state of the perpetrator reflects a concern with delivering just deserts, or whether it instead reflects a concern with effectively deterring other would-be offenders. Is a participant's focus on such factors *while making punishment decisions* driven by retributive or more utilitarian considerations, or by some combination of the two (see also Baron & Ritov, 2009)? The existing evidence does not yield a definitive answer.

In the present paper, in addition to documenting the fundamental ambiguity in previous studies, we propose a novel solution: investigating the existence of a purely retributive motive by examining people's retributive desires toward nonhuman actors, namely animals that have attacked and killed humans, pets, or livestock. One significant advantage of this approach is that it allows us largely to remove general deterrence as a possible punishment motive—general deterrence requires mechanisms, at both the individual and group level, that facilitate the dissemination and understanding of punishment information (e.g., individual X was later subjected to punishment Y for previously acting in Z manner and if I act in Z manner, in the future, I, too, will be subjected to punishment Y). As animals, like sharks, clearly lack these complex individual and social mechanisms, people are highly unlikely to think that punishing one animal will deter other animals from violent attacks. Moreover, if the punishment of the animal in question entails killing it, which is the case in our studies, this means that the motive to deter the specific animal is also eliminated (or, put differently, this motive devolves to the motive to incapacitate). Thus, only incapacitation and retribution are left as possible punishment motives, making the task of separating them more tractable.

1.3. Punishing animals

Given their relatively lower mental capacities, animals seem to be less natural targets for retribution than humans. Consequently, if evidence for retributive motives directed toward animals can be demonstrated, it would be reasonable to conclude that similar motives can be directed toward humans. Moreover, the existence of retributive motives directed toward animals is not as far-fetched as it may seem at first. In fact, humans have

been punishing animals implicated in harms for thousands of years, often in ways that are difficult to distinguish from the punishments of human transgressors deemed guilty of comparable misdeeds. There is evidence that the ancient Greeks tried animals that had caused human deaths in the Prytaneum of ancient Athens (Girgen, 2003; Hyde, 1916). And later European records from the medieval period all the way up into the nineteenth and even twentieth centuries show that many animals that had spilt blood while committing an offense—goring oxen, violent dogs and goats, bucking horses, and ravenous pigs—were tried in secular courts and put to death (see e.g., Berman, 1994; Bondeson, 1999; Evans, 1906; Humphrey, 2003). The punishments handed out in these cases often suggest a retributive motivation, reflecting, among other things, an adherence to retributive principles of proportionality. In one particularly compelling example, in 1386, when a sow bit a child in the face and arms, it was sentenced to be given matching injuries to its head and forelegs, garroted, and then hanged in the public square in Falaise (Bondeson, 1999). Over the centuries, ecclesiastical courts carried out similar, although often more elaborate, judicial proceedings against vermin—rats, locusts, and the like—that had destroyed crops, brought disease, or infested villages (Bondeson, 1999; Evans, 1906). In addition, the history of animal punishment is not confined to the European subcontinent, nor is it limited to Western or Judeo-Christian cultures (Berman, 1994; Girgen, 2003).

Even in the modern United States, animal punishment continues to be practiced, particularly with respect to dogs that have attacked humans (Girgen, 2003). Indeed, there is recorded evidence that in 1926, in the State of Kentucky, a stray German shepherd was subjected to the electric chair after being condemned to death for the attempted murder of a child (Bondeson, 1999). Although numerous states have enacted laws requiring that "vicious" dogs be put to death, formal animal trials are uncommon today (Girgen, 2003). Yet killing animals that have attacked humans is not—private citizens regularly take matters into their own hands to execute transgressing animals, often with the tacit approval of authorities (Girgen, 2003).

Legal codes and religious texts also reference animal punishment as an expected part of enforcing the moral order. The Old Testament, for example, notes the righteousness of animal punishment at various points, as do Mesopotamian sources from centuries earlier (Finkelstein, 1981). According to Exodus 21:28 (King James), for instance, "If an ox gore a man or a woman, that they die: then the ox shall be surely stoned, and his flesh shall not be eaten; but the owner of the ox shall be quit." As a punishment, death by stoning was a special sentence, meant to convey the particular outrage of the public at the crime (Finkelstein, 1981).

Although many of these historical examples quite strongly suggest that retributive motives might underlie animal punishment practices and that retribution directed at animals may not be fundamentally different from that directed at humans, none of them provide clear-cut evidence that retribution underlies animal punishment. Moreover, to date, no experimental studies appear to have been conducted that investigate individuals' desire to inflict retributive punishment on animals. We set out to address this gap by conducting a series of novel experimental investigations.

In sum, investigating retributive motives toward animals offers the promise of both (a) revealing more precise evidence for the role of retribution in punishment decisions, given that previous experiments have not adequately separated out retributive concerns from deterrence or incapacitation concerns owing to the methodological problems with studying human perpetrators, and (b) demonstrating that retributive motives can be directed much more broadly than has previously been understood. The first concern is focused on isolating a purely retributive motive; the second is focused on identifying the scope of such a motive (i.e., identifying to which entities it might apply). Data on whether lay individuals experience retributive feelings toward entities (e.g., nonhuman animals) that are generally understood by our current legal system to lack the mental (and moral) capacities necessary for just punishment would be of significant value to our understanding of the necessary mental state variables and thresholds for retributive motives to be engaged. Thus, achieving either or both of these aims would advance the current understanding of retribution.

2. Study 1: Validation study

The first step in this project was to document the ambiguity of current evidence in support of the existence of retributive motives. To this end, we conducted a methodologically revised version of Carlsmith et al.'s (2002) initial "validation" study, which investigated which factors of criminal offenses individuals regard as relevant to the retributive and deterrence-based goals of punishment.

In their initial study, Carlsmith et al. (2002) asked participants to indicate whether four separate factors—the magnitude of harm caused by a crime, the perpetrator's motivation, the detection rate for a crime, and the publicity that a crime and its attendant punishment attract—were relevant either to the goal of retribution or to the goal of deterrence. The study found that 76% of participants classified magnitude of harm as most relevant to retribution, while only 16% classified it as most relevant to deterrence. Similarly, 69% of participants classified perpetrator motivation as most relevant to retribution, while only 14% classified motivation as most relevant to deterrence. This study was used as the basis for interpreting Carlsmith et al.'s (2002) subsequent results—namely, that participants were most sensitive to magnitude of harm and perpetrator motivation in their punishment decisions—as evidence of a retributive motive.

However, a major issue in interpreting these results as revealing a purely retributive motive stems from Carlsmith et al.'s (2002) use of a forced choice methodology in their initial validation study. Participants were not able to indicate that any single variable was relevant to *both* retribution and deterrence. It therefore remains possible that while a majority of participants classified both harm severity/magnitude and perpetrator motivation as more relevant to retribution than deterrence, they may only have regarded these factors as slightly more relevant (i.e., as highly relevant to deterrence, even though slightly more relevant to retribution).[1] In order to investigate this possibility, we conducted a new validation study, in which participants were able to express the

importance of a range of crime variables to both retribution and deterrence in a nonexclusive fashion.

2.1. Method

2.1.1. Participants

A total of 121 participants from the United States (74 male, 47 female, M_{age} = 33.30 *SD* = 11.73) participated in the study, which was posted on Amazon.com's Mechanical Turk interface. Participants received payment for their participation.

2.1.2. Materials, design, and procedure

Participants acted as their own controls and judged the importance of five different crime variables from the perspective of two different punishment philosophies—retribution and general deterrence. The passages describing retributive and deterrence-based punishment, as well as the exact wording of the dependent variables, are provided in the Supporting Information (which also contain the scenario and dependent measure wordings for the remaining studies).

Participants made five judgments about deterrence or retribution, with the overall order in which they considered these punishment perspectives counter-balanced. For each punishment perspective, participants judged, on a 5-point scale, how important it was to punish offenders based on the severity of the harm they had committed (*magnitude*), the goodness or badness of their underlying motivation (*motivation*), the difficulty of detecting and successfully prosecuting the particular type of crime committed (*detection rate*), the amount of publicity the crime attracted (*publicity*), and the prevalence of the crime in society (*frequency*).[2] The specific order in which these five offense-related factors were judged was varied randomly (both across participants, and within participants across the two punishment perspective blocks). Following completion of these measures, participants responded to demographic items and were then thanked and debriefed. No other measures were collected.

2.2. Results

As Table 1 shows, participants judged both the magnitude of harm and the perpetrator's motivation as being highly important when punishing to serve both retributive and general deterrence goals, such that there was no significant difference between these ratings. Perpetrator motivation was rated as slightly more relevant to retribution than to deterrence, although the difference was nonsignificant. In contrast to these null results, detection rate was rated as being more important to deterrence than to retribution, as were publicity and crime frequency.

Moreover, as Table 1 shows, harm magnitude and motivation were rated as being the *most important* variables to attend to from *both* punishment perspectives. Harm magnitude and motivation together (averaged) were considered more important than the remaining

Table 1
The mean rated importance (on a 5-point scale) of five different variables to retributive and general deterrence punishment philosophies, in Study 1. Standard deviations are in parentheses

	Rated Importance for Retribution	Rated Importance for General Deterrence	*t*-value	Cohen's *d*	*p*-value (two-tailed)
Magnitude of harm	4.43 (0.83)	4.46 (0.78)	.41	.03	.68
Perpetrator's motivation	3.81 (1.04)	3.67 (1.15)	1.36	.12	.19
Detection rate of crime	2.85 (1.33)	3.27 (1.30)	3.75	.34	< .001
Publicity of crime and its punishment	2.53 (1.27)	2.91 (1.31)	3.25	.30	.001
Crime frequency	3.28 (1.36)	3.52 (1.27)	2.24	.21	.03

variables (averaged) for retribution, $t(120) = 12.87$, $p < .001$, $d = 1.20$, as well as for general deterrence, $t(120) = 12.87$, $p < .001$, $d = 0.75$.

2.3. Discussion

These results show that when participants are not constrained by a forced choice procedure, they regard magnitude of harm and perpetrator motivation as highly (and equally) relevant to both retribution and deterrence. This suggests that Carlsmith et al.'s (2002) assumption that harm severity and perpetrator motivation are more relevant to retribution than to general deterrence is incorrect, which substantially complicates the interpretation Carlsmith et al. (2002) made of their ensuing punishment findings. Participants' documented sensitivity to both magnitude of harm and motivation, in the context of crimes committed by humans, cannot unambiguously be interpreted as stemming from a retributive motive. In addition, recent research that has attempted more conclusively to isolate the retributive motive (e.g., Aharoni & Fridlund, 2012) faces an analogous problem in relying on intent manipulations that do not effectively rule out incapacitation (as opposed to deterrence) motives. When it comes to isolating the retributive motive, we therefore argue that a different approach is required. Accordingly, we now turn to studies of people's desire to punish animal attackers in order more clearly to disentangle retributive from other motivations.

To document the existence of retributive motives toward animals, in Studies 2 and 3, we first sought to find evidence for a retributive principle of proportionality: that punishment is owed to an offender according to the offender's just deserts (Kant, 1797/1991), which is commonly defined in terms of an offender's blameworthiness or in terms of the severity of the wrong committed (Ristroph, 2005).[3] To this end, we manipulated the victim of an animal attack, such that in some cases an animal attacked and killed a relatively unsympathetic victim (or caused some other lesser harm), and in other cases, an animal attacked and killed a highly sympathetic victim. The killing of a more sympathetic victim, by virtue of engendering a greater sense of loss, should be encoded as a more severe wrong, which should therefore produce a greater desire to put the offending animal to death and a stronger belief that the animal *deserves* to be killed.

3. Study 2: The "victim identity" effect

In Study 2, we described an animal attack as causing (a) significant property damage, (b) the death of a German shepherd, (c) the death of a 53-year-old homeless man, or (d) the death of a 10-year-old girl. Our prediction was that the perceived degree of loss caused by the attack should predict individuals' support for the killing of the implicated perpetrator. If this pattern of results (what we refer to as a "victim identity" effect) were to arise, we argue that it would indicate that people desire the retributive punishment of transgressing animals, in accordance with the retributive principle of proportional punishment.

An alternative explanation based on general deterrence motivations would seem implausible. As described in the Introduction, general deterrence depends on a sophisticated communication system, in which the threat of punishment is both conveyed to the wider society and internalized by would-be offenders—which we presume almost all participants will regard as infeasible for the relevant animal communities. And, indeed, in a preliminary study, we asked participants to indicate on a 9-point scale their agreement with the idea that "catching and killing a shark that has attacked a human will deter or dissuade other sharks from attacking humans." Very few of these participants (n = 204) rated the possibility of deterring other sharks from attacking humans through the punishment of one,[4] as remotely plausible (M = 1.46, on a 1–9 scale, where 1 = "disagree completely").

It is also implausible that the predicted victim identity effect might reflect incapacitationist motivations. Individuals might see an animal as differentially dangerous depending on the consequences of its attack, which might then lead to a greater desire to incapacitate it (e.g., by killing it). Indeed, dangerousness is the key variable from a forward-looking incapacitationist perspective because it is the future dangerousness of the offender that determines the likelihood of future harm. However, while plausible in general, this alternative account would not easily explain a monotonic increase across the four selected outcome conditions, since, for instance, an animal that kills a little girl should not be seen as more dangerous than one that kills an adult male or a German shepherd. If anything, an opposing trend with respect to perceptions of dangerousness would be expected (i.e., a shark that is able to kill a grown man or a German shepherd is also very likely able to kill a little girl, but a shark that is able to kill a little girl is not necessarily able to kill a grown man). Thus, we chose these particular victims in part to render this alternative interpretation highly unlikely. To rule it out more decisively, we also measured and statistically accounted for the perceived dangerousness of the offending animal in each case.

3.1. Method

3.1.1. Participants

A total of 267 participants from the United States participated in the study, which was posted on Amazon.com's Mechanical Turk interface. Of these, 14 did not complete the study, leaving data for 253 remaining participants (91 male, 162 female, M_{age} = 34.48, SD = 12.08). Participants were each paid for their participation.

3.1.2. Materials, design, and procedure

Each participant read about four different attacks and then responded to a variety of questions concerning each attack. We chose three possible animal perpetrators of the attack: a shark, a bull, and a pit bull. For the sake of comparison, we also included one human attacker—a political activist who detonated a small bomb in a subway station. We manipulated the consequences of the human's attack in the same way as for the animal perpetrators.

The three different victims were described as "a 10-year-old girl, Melissa," "a 53-year-old homeless man, John," and "a 10-year-old German shepherd, Buster." In the fourth, property damage case, the attacker was described as causing several thousand dollars worth of property damage. For the three animal perpetrators, this property damage resulted from the perpetrator's attempt to attack a 46-year-old man named Chris. The type of property damaged depended on the identity of the attacker (shark: kayak; bull: tractor; pit bull: furniture). In each case, Chris, the intended victim, was unscathed. The property damage in the human attacker case resulted from the destruction of two ticketing machines in the subway station. Just as with the animal attackers, this individual was not described as intending to kill any particular person.

The attacker and the consequences of the attacks were varied so that participants never read about the same victim or the same perpetrator. Participants were randomly assigned to one of four different versions of the study, each of which contained a different assignment of the four attackers to the four possible consequences (i.e., four such configurations were chosen out of the 24 possible configurations).[5]

After reading each scenario, participants then responded to a series of questions, all of which were on 9-point scales. As manipulation checks, they first indicated how tragic the victim's death was (this question was omitted for the property damage scenarios), how significant a loss the outcome was, how sad the victim's death made them feel (this was also omitted for the property damage cases), how angry the attack made them feel, and how frightening they found the description of the attack, in that order. As a control measure, participants next indicated how dangerous they thought the attacker was. Participants then responded to the two key dependent variables assessing how much they would be in favor of the relevant authorities putting the attacker to death, and how much they thought the attacker deserved to be put to death. These two measures were highly correlated for each of the four attackers (all Cronbach's $\alpha > .93$) so we averaged them together in the analyses below. Finally, participants were asked to consider a situation in which a citizen decided to kill the attacker without official sanction and to indicate how immoral or moral they thought this was, as well as how justified an animal (or human) rights group would be in protesting this vigilante killing. These auxiliary variables showed very similar overall patterns as did our main dependent variables, so we do not report on them further here. However, additional analyses, including all means for all measures are included in the Supporting Information. Upon completion of the survey, participants provided some demographic information and were thanked and debriefed. No other measures were collected.

3.2. Results

3.2.1. Victim identity manipulation check

To confirm the effect of the victim identity manipulation,[6] we examined participants' perceptions of the loss that each outcome constituted. As predicted, an increasing trend in terms of the degree of loss was observed across the outcomes: with property damage viewed as the least significant loss, followed by the killing of the dog, the killing of the man, and, finally, the killing of the little girl. This trend was reliable for all four attackers[7]: shark, Kendall's τ_b (251) = .53, $p < .001$; bull, Kendall's τ_b (251) = .41, $p < .001$; pit bull, Kendall's τ_b (251) = .38, $p < .001$; human, Kendall's τ_b (251) = .43, $p < .001$. Table S1 in the Supporting Information presents the effect of the victim identity manipulation on the degree of loss variable, as well as other variables of interest.

3.2.2. Victim identity effect on retributive responses

The victim identity effect was highly reliable in the predicted direction for our key retributive measure, which combined participants' beliefs about the extent to which the attacker deserved death and their support for killing the attacker (Fig. 1). To control for perceptions of dangerousness (and, thus, address the possibility of an alternative incapacitation-based explanation for the results), we regressed participants' punitive judgments on their assessments of the dangerousness of each attacker, in order to create residualized scores that reflect the component of the punitive responses that is not accounted for by dangerousness. Nonparametric correlations between victim identity and the residualized punitive responses were reliable for the bull attacker, Kendall's τ_b (251) = .33, $p < .001$, the shark attacker, Kendall's τ_b (251) = .18, $p = .002$, the pit bull attacker, Kendall's τ_b (251) = .16, $p = .002$, and the human attacker, Kendall's τ_b (251) = .22, $p < .001$.[8] Hence, participants' punitive responses toward the attackers were driven by victim identity in a way that is not explained by their perceptions of the attackers' dangerousness.

3.3. Discussion

Study 2 demonstrated that individuals desire to punish animals that attack and kill humans to a greater degree as the perceived loss caused by the animal's violent offense increases. This aligns with the retributive principle of proportionality, which dictates that the severity of the punitive response track the severity of the infraction, and provides support for the existence of a specifically retributive motive directed toward animal attackers. Moreover, the data from this study suggest that this retributive motive directed toward animals does not differ significantly from that directed at a human attacker.

The results cannot easily be explained by the desire to deter or incapacitate animal offenders. Incapacitation is the more plausible of these two alternatives; yet it cannot explain the full pattern of our results. A desire to incapacitate should most closely track

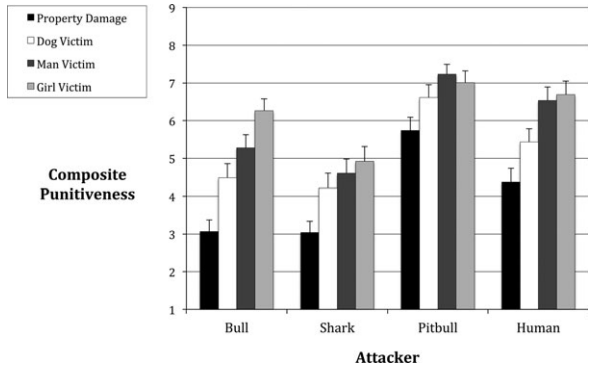

Fig. 1. Study 2 participants' punitive reactions (composite measure of support for death and belief that the attacker deserves death) as a function of attacker and victim. Error bars represent standard errors.

perceptions of an animal's dangerousness (i.e., the threat that the animal poses). Yet there is no good reason to think that an animal attacker is systematically more dangerous across the outcomes we manipulated in a way that matches our data. Indeed, the victim that prompted the most desire to kill the animal attacker—the 10-year-old girl—is surely the one that is easiest for an animal to kill. Our participants' attributions of dangerousness did not quite adhere to this rational account. As shown in the Supporting Information, participants sometimes (implausibly) judged the killers of the most significant victims as more dangerous, potentially reflecting belief-overkill (see Baron, 2009; Jervis, 1976). Nonetheless, the victim identity effect on punitive responses emerges clearly even when those seemingly biased attributions of dangerousness are statistically accounted for.

4. Study 3: The "targeted punishment" effect and replication of the "victim identity" effect

Study 3 aimed to replicate the victim identity effect observed in Study 2, while also providing a second means of establishing the existence of retribution directed at animal offenders. Specifically, we investigated whether evidence exists for what we refer to as "targeted punishment"—the idea that tracking down and killing the exact "guilty" animal that carried out an attack is more important than tracking down and killing an equally dangerous but "innocent" animal. Such a demonstration would provide corroborating evidence for the existence of retributive motives directed at animals since punishing only the guilty is a hallmark of retributive punishment (see e.g., Flew, 1954; Mabbott, 1939; Quinton, 1954; Rachels, 1977; but cf. Cottingham, 1979). By contrast, if individuals are motivated primarily by the desire to incapacitate dangerous animals, there should be relatively little difference in their responses toward the actual killer and an almost identical "stand-in" shark from the same species. Whatever difference does exist should correspond only to whatever difference in dangerousness might be inferred between the two animals.

4.1. Method

4.1.1. Participants

A total of 156 undergraduate students in the United States participated in the study for course credit. Three participants did not complete all relevant measures, and so their data were excluded, leaving a total sample of 153 participants (70 male, 82 female, 1 unspecified, M_{age} = 19.62 years, SD = 1.32).

4.1.2. Materials, design, and procedure

The study was carried out using paper and pencil, and was divided into two parts presented in a counter-balanced order and separated by approximately 15–20 min, during which time participants completed a series of unrelated questionnaires. Each of the two separate parts of the study presented subjects with a different victim of a shark attack: Melissa, a 10-year-old girl, in one part, and Rocky, a full-grown Rottweiler dog, in the complementary part, with the order of presentation of these two parts counter-balanced. Except for some minor wording differences, the description of each attack and the ensuing events was largely the same as in Study 2.

Following the description of each attack, participants indicated on 9-point scales how tragic the death was and how much of a loss to society it was (manipulation checks). Participants next read that authorities had decided to track down and kill the offending shark, and that they eventually killed a 15-foot white pointer shark. This shark was later determined to be the actual killer (*guilty shark* condition) or a highly similar white pointer shark that was not the actual killer (*innocent shark* condition). This variable was manipulated within subjects and within each of the two time-separated parts of the study. Thus, participants answered the same questions, with respect to each shark, for each of the two victims separately (again on 9-point scales). Participants then indicated how wrong they thought the killing of the shark was, and how much they thought the shark that was killed deserved to be killed (this measure of deserved killing was the key punitive measure). Following this, participants indicated how much the killing made amends for the victim's death. We expected that this measure of amends would mediate the targeted punishment effect—that is, that killing the guilty shark would be thought to make better amends than killing the innocent shark, thus explaining the greater desire to punish the guilty shark. Finally, as a control measure, participants indicated how dangerous the shark that was killed was. The order of presentation of the guilty shark and innocent shark was counter-balanced, although it was held constant within subjects across the two parts of the study. The counter-balancing of the order of the victims and the order of the sharks gave rise to four different versions of the study that participants were assigned to at random.

At the end of the study, participants responded to measures assessing their general retributive beliefs, including their support for the principle of an "eye for an eye," and related measures (see Supporting Information). We included these measures because we thought they might moderate some of the predicted effects. However, they did so only weakly and somewhat inconsistently, so we do not report on these measures further.

Finally, participants provided some demographic information and were thanked and debriefed. No other measures were collected.

4.2. Results

4.2.1. Victim identity manipulation check

Confirming the effect of the victim identity manipulation, participants saw the death of Melissa as more tragic than the death of Rocky ($M_{Melissa} = 8.30$, $M_{Rocky} = 6.92$, $t(153) = 10.34$, $p < .001$, $d = 1.00$), and as more of a loss to society ($M_{Melissa} = 5.60$, $M_{Rocky} = 2.71$, $t(155) = 17.75$, $p < .001$, $d = 1.46$).

4.2.2. Victim identity and targeted punishment effects

A 2 (victim identity) × 2 (shark) within subjects ANOVA on the primary punitive measure—the extent to which the shark was thought to deserve death—corroborated our main predictions. There was a main effect of victim identity, such that participants thought the shark that killed Melissa, $M = 2.98$, was more deserving of death than the shark that killed Rocky, $M = 2.36$, $F(1, 152) = 38.73$, $p < .001$, partial $\eta^2 = .20$. There was also a main effect of shark identity (the targeted punishment effect), such that participants thought the guilty shark, $M = 3.64$, was more deserving of death than the innocent shark, $M = 1.70$, $F(1, 152) = 145.03$, $p < .001$, partial $\eta^2 = .49$. In addition, there was a significant interaction between these variables, $F(1, 152) = 22.79$, $p < .001$, partial $\eta^2 = .13$, which revealed that the victim identity effect was considerably larger in the guilty shark condition, $M_{Melissa} = 4.14$ versus $M_{Rocky} = 3.14$, $t(152) = 6.52$, $p < .001$, $d = 0.54$, than in the innocent shark condition, $M_{Melissa} = 1.82$ versus $M_{Rocky} = 1.57$, $t(152) = 2.68$, $p = .008$, $d = 0.21$. Similar results held when analyzing participants' beliefs about how wrong it would be to kill the shark, except that there was no significant interaction between the two independent variables (see the Supporting Information for these analyses).

It is not possible to control for perceptions of dangerousness in a standard ANCOVA analysis, since dangerousness was a time-varying covariate in our design. Accordingly, we ran mixed model analyses with perceived dangerousness entered as a time-varying covariate. On these analyses, the effect of victim identity remained highly significant once dangerousness was accounted for, $F(1, 465.89) = 17.22$, $p < .001$, as did the targeted punishment effect, $F(1, 528.50) = 144.44$, $p < .001$.

We next investigated whether the targeted punishment effect was mediated by the perception that killing the guilty shark made better amends for the victim's death than killing the innocent shark. Consistent with a mediation model, participants indicated that killing the guilty shark made better amends for the original killing of the victim ($M_{guilty} = 2.41$, $M_{innocent} = 1.42$, $t(152) = 8.15$, $p < .001$, $d = 0.70$). Following the procedure outlined by Judd, Kenny, and McClelland (2001), further regression analyses revealed that the difference in perceived amends between killing the guilty shark and killing the innocent shark partially mediated the targeted punishment effect with respect to

the deserving death variable, $t(150) = 4.17$, $p < .001$. Moreover, this result still held strongly while controlling for perceived dangerousness, $t(148) = 3.93$, $p < .001$.

4.3. Discussion

In sum, Study 3 accomplished three main aims. First, it replicated the victim identity effect in Study 2 by showing that animals that kill a more sympathetic victim are judged to be more deserving of death. Second, it provided evidence for a "targeted punishment" effect such that a shark that is responsible for killing a human is seen as more deserving of death than an almost identical and equally dangerous shark that was not implicated in the attack. Both of these effects remain significant after perceptions of the shark's dangerousness are accounted for—thus bolstering the claim that they reflect retributive rather than consequentialist considerations. Third, Study 3 showed that the targeted punishment effect is mediated by the sense that killing the actual guilty perpetrator makes amends (i.e., achieves justice) better than does killing an equivalently dangerous yet innocent animal, further reinforcing our interpretation of this effect as reflecting retributive sentiment.

The key dependent variables in Studies 2 and 3 pertained to the killing of an offending agent—whether people supported such a killing (Study 2) and whether they thought it was deserved (Studies 2 and 3). The findings that emerged from these studies strongly suggest that retributive motives can indeed be directed toward a nonhuman agent, such as an animal. However, because the death of a dangerous attacker is also critically relevant to incapacitating that attacker, it remains conceivable that some incapacitationist alternative could explain these data. We think this is implausible given the various challenges that such an account has in explaining the entire pattern of data—namely, that (a) we chose specific victims in a way that was designed to neutralize the incapacitation concern, (b) the degree to which a responsive killing was thought to make amends (a construct that is integral to the notion of justice) partially mediates the targeted punishment effect, and (c) both the victim identity and targeted punishment effects hold while controlling for perceived dangerousness. Nevertheless, in the remaining two studies we sought to rule out the incapacitationist account more decisively, for both the targeted punishment and victim identity effects.

5. Study 4: The targeted punishment effect extends to acceptance of inflicting pain during execution

Study 4 investigated whether the targeted punishment effect can be demonstrated in terms of participants' acceptance of inflicting suffering on an animal (here, a shark) that is certain to be killed following a fatal attack on a human. If individuals are more accepting of causing pain to the guilty shark as opposed to an innocent shark, it would more definitively show that the targeted punishment effect is motivated, at least in part, by a

purely retributive motive. Incapacitation has been removed from consideration in this situation because participants know that the animal is certain to be incapacitated (i.e., killed) regardless of whether it is guilty or innocent. The only remaining question is how much it should suffer while being killed. In this study, we also included a larger range of potentially mediating questions that were designed to assess participants' perceptions of the dangerousness of the shark and the threat it posed, their beliefs about how important it is to incapacitate the shark and to prevent it from carrying out further attacks, and their retributive sentiment toward the shark. Our aim was to discover whether the specifically retributive mediators played an underlying role in participants' punishment judgments.

5.1. Method

5.1.1. Participants

A total of 415 participants from the United States agreed to take part in the study, which was posted on Amazon.com's Mechanical Turk interface. Of these, 8 did not complete the study, leaving data for 407 participants (233 male, 174 female, M_{age} = 30.58, SD = 11.06). Participants were each compensated for their participation.

5.1.2. Materials, design, and procedure

Before starting the survey, participants answered a Captcha question to verify that they were not automated bot programs. They then read about an attack by a white pointer shark that killed a 10-year-old girl named Melissa. Participants were randomly assigned to one of two conditions, which varied the identity of the shark that authorities captured nearby in the wake of the attack. Roughly half of the participants (n = 209) were randomly assigned to a condition in which they learned that the captured shark was determined to be the shark that had killed Melissa (*guilty shark*), whereas the remaining participants (n = 198) learned that the captured shark was *not* the guilty shark but was instead was "an almost identical white pointer shark of similar size, age, health, and coloration to the actual killer" (*innocent shark*). In both conditions, all participants learned that the authorities had decided to kill the shark.

Our main dependent variables were three questions designed to capture participants' acceptance of inflicting pain on the offending shark (9-point scales): the extent to which they thought the shark deserves a painful death, how painful a death they thought the shark should receive, and how much of a limited supply of a costly anesthetic they thought the shark deserved to receive to alleviate its pain during its killing. Prior to being presented with these main dependent variables, participants responded to a range of potentially mediating questions (9-point scales). These included two questions pertaining to the dangerousness of the shark and the threat it posed, which were averaged together, α = .94; two questions about the importance of preventing the shark from carrying out future attacks and the importance of incapacitating the shark so that it is no longer a threat, which were also averaged together, α = .87; and three questions pertaining directly to retributive sentiment—the importance of killing the shark to avenge Melissa's death, the importance of killing the shark to make amends for Melissa's death, and how blame-

worthy the shark was for killing Melissa—which were also averaged together, $\alpha = .89$. These mediators were presented in a new random order for each participant.

After responding to the main dependent measures, participants provided demographic information and were then thanked and debriefed. No other measures were collected.

5.2. Results and discussion

To simplify the analysis, the three pain measures were averaged together after we transposed the anesthetic question so that higher numbers reflected more acceptance of a painful killing, $\alpha = .74$. As predicted, participants favored the more painful method of killing to a greater extent for the guilty shark as opposed to the innocent shark, $M_{\text{guilty}} = 2.75$, $M_{\text{innocent}} = 1.74$, $t(342.81$, equal variances not assumed$) = 6.43$, $p < .001$, $d = 0.63$. Given that the assignment of pain is not relevant to a desire to incapacitate, it would seem that this result reflects retributive sentiment. Indeed, multiple mediator bootstrap analyses run with 5,000 samples (Preacher & Hayes, 2008a) showed clearly that the aggregated retribution mediator was the only significant mediator of this effect. As Table 2 shows, the shark identity manipulation had a significant effect on all three of the potential mediators. However, only the retribution mediator had a significant effect on the dependent variable. As a consequence, whereas the 95% confidence interval for the retributive variable did not include zero, the combined dangerousness and desire to incapacitate mediators had 95% confidence intervals that included zero, indicating that they had no significant mediating effects. These analyses simultaneously control for each of the other mediators entered into the model, thereby showing that the retributive mediator exerts its effect even while controlling for perceptions of dangerousness and the desire to incapacitate the offending shark.

In sum, Study 4 shows that the targeted punishment effect extends to the acceptance of inflicting suffering on an animal in cases where its death is assured and that variables

Table 2
Coefficients of mediation models when all mediators are entered simultaneously as predictors of the aggregated pain assignment variable, in Study 4 (targeted punishment effect)

Mediator	a path	b path	ab path	Lower bound	Upper bound
Retribution (*amends*, *avenge*, *blameworthiness*)	2.03***	.42***	**.85**	.60	1.15
Danger (*dangerousness*, *threat*)	1.27***	−.01	−.01	−.10	.07
Desire to incapacitate (*importance of future prevention*, *importance of incapacitation*)	1.37***	.05	.07	−.01	.16

Note. "a path" denotes the direct effect of shark identity on each mediator, "b path" denotes the direct effect of each mediator on punitiveness (pain), and "ab path" denotes bootstrap estimates of the indirect effect of shark identity on punitiveness (pain) through each mediator. Exact *p* values cannot be computed for the coefficients of the ab paths; we therefore indicate only whether the 95% bootstrap confidence interval does not contain zero, in bold.
*$p < .05$; **$p < .01$; ***$p < .001$.

targeting retribution—but not perceptions of dangerousness or the desire to incapacitate—mediate this effect. It therefore provides the most direct evidence that retributive sentiment underlies the targeted punishment effect.

6. Study 5: The victim identity effect extends to acceptance of inflicting pain during execution

Study 5 investigated whether the victim identity effect extends to participants' acceptance of inflicting pain on an offending shark. Here, we also used a range of potentially mediating variables—the same variables for dangerousness and desire to incapacitate, but a slightly wider range of relevant retribution variables—in order to examine more directly whether retributive sentiment underlies the victim identity effect.

6.1. Method

6.1.1. Participants

A total of 514 participants from the United States agreed to take part in the study, which was posted on Amazon.com's Mechanical Turk interface. Fourteen participants did not complete the study, leaving a final sample of exactly 500 participants (317 male, 183 female, M_{age} = 30.60, SD = 9.94), who were each paid for their participation.

6.1.2. Materials, design, and procedure

Before starting the survey, each participant answered a Captcha question to verify that he or she was not an automated bot program. Each participant then read about a violent attack by a great white shark. Roughly half of the participants (n = 248) were randomly assigned to a condition in which the shark killed a 10-year-old girl named Melissa, described as being in fourth grade and loving sports. The remaining participants (n = 252) were assigned to a condition in which the shark killed a 48-year-old pedophile named Dale. Dale was described as having raped and sexually abused numerous young girls while serving as their music teacher, without ever having been caught. This is a stronger victim identity manipulation than in the previous experiments, but, as in those experiments, the specific identities that were selected should not affect perceptions of the dangerousness of the shark in a way that compromises the interpretation of the results. If anything, the shark that killed the full-grown man should be seen as more dangerous than the one that killed the young girl, since the man should be more difficult to kill.

The same three dependent variables from Study 4 were used again (9-point scales) and averaged together to form a composite measure of pain assignment, α = .70. Prior to answering the dependent measures, participants answered a set of potential mediating questions that was slightly more extensive than the set used in Study 4 (all on 9-point scales and presented in a random order). Except for some very slight wording differences, the dangerousness and threat questions were as they were in Study 4, and were averaged together, α = .94. The same was true for the importance of preventing future attacks and

importance of incapacitating questions, $\alpha = .84$. The list of retribution questions was slightly more extensive, and it included questions about how important it was to avenge the shark's killing of its victim, how important it was to kill the shark to make amends for its killing of the victim, how important it was to make the shark pay for its killing, how important it was to seek retribution for the shark's killing of its victim, how important it was to seek justice for the shark's killing of its victim, and how blameworthy the shark was, $\alpha = .93$.

At the end of the survey, participants provided some demographic information before being thanked and debriefed. No other measures were collected.

6.2. Results and discussion

As predicted, on the composite dependent measure, participants were more inclined to inflict pain on the shark that had killed Melissa than the shark that had killed Dale, $M_{\text{Melissa}} = 2.89$, $M_{\text{Dale}} = 2.08$, $t(444.25$, equal variances not assumed$) = 5.47$, $p < .001$, $d = 0.49$). Moreover, multiple mediator bootstrap analyses run with 5,000 samples showed that, just as with the targeted punishment effect, the aggregated retributive mediator was the only significant mediator of the victim identity effect. As Table 3 shows, the victim identity manipulation had a significant effect on the retribution mediator, but not on the importance of incapacitation or dangerousness mediators. Moreover, only the retribution mediator had a significant effect on the dependent variable. Consequently, the 95% confidence interval for the retributive variable did not include zero, whereas the combined dangerousness and desire to incapacitate mediators had 95% confidence intervals that included zero, indicating that they had no significant mediating effects. Thus, because the analyses simultaneously control for each of the other mediators entered into the model, they show that the retributive mediator exerts its effect even while controlling for perceptions of dangerousness and the desire to incapacitate the offending shark.

Table 3
Coefficients of mediation models when all mediators are entered simultaneously as predictors of the aggregated pain assignment variable, in Study 5 (victim identity effect)

Mediator	a path	b path	ab path	Lower bound	Upper bound
Retribution (*amends*, *avenge*, *justice*, *make pay*, *retribution*, *blameworthiness*)	.57**	.41***	**.23**	.07	.42
Danger (*dangerousness*, *threat*)	−.10	.06	−.01	−.05	.01
Desire to Incapacitate (*importance of future prevention*, *importance of incapacitation*)	.01	−.01	.07	−.02	.01

Note. "a path" denotes the direct effect of shark identity on each mediator, "b path" denotes the direct effect of each mediator on punitiveness (pain), and "ab path" denotes bootstrap estimates of the indirect effect of victim identity on punitiveness (pain) through each mediator. Exact p values cannot be computed for the coefficients of the ab paths; we therefore indicate only whether the 95% bootstrap confidence interval does not contain zero, in bold.
*$p < .05$; **$p < .01$; ***$p < .001$.

In sum, Study 5 shows that the victim identity effect also extends to the acceptance of inflicting suffering on an animal in cases where its death is assured, and that variables targeting retribution—but not perceptions of dangerousness or the desire to incapacitate—mediate this effect. These results therefore provide the most decisive evidence that retribution underlies the victim identity effect.

7. General discussion

The present studies were designed to elucidate more precisely the role of retribution in punishment decisions by investigating individuals' motives to punish animals that have perpetrated violent attacks. In so doing, they were also able to investigate whether the retributive motive extends to nonhuman actors, such as animals. Together, the studies reported here demonstrate clear evidence for the existence of retributive motives and for a broader conception of the viable targets of retribution. Indeed, the present research shows that retributive motives can extend to entities generally seen as lacking the requisite characteristics to be worthy of punishment. In what follows, we synthesize the main empirical evidence for these claims, address potential objections to our interpretations, and consider the implications of the research.

7.1. Clearer evidence for retributivism

Past studies of the punishment of human actors have argued that of the three main motivations for punishment—deterrence, incapacitation, and retribution—retribution appears to dominate punishment decisions (e.g., Carlsmith, 2006; Carlsmith et al., 2002). But this research has not adequately isolated retribution as a motive for punishment. Participants in prior studies were shown to be quite sensitive to factors that are relevant to retribution—among them, crime severity, intent, and extenuating circumstances—and tended to indicate that such factors were more relevant to retribution than to deterrence or incapacitation (Carlsmith, 2008; Carlsmith et al., 2002). However, crime severity, intent, and extenuating circumstances are also extremely relevant to deterrence and incapacitation, even if they are perceived to be slightly more relevant to retribution. Indeed, as Study 1 showed, our adult participants rated the magnitude of the harm caused by a crime and the motivation underlying it as no more relevant to retribution than to general deterrence. Participants also rated these two crime features as the two most important features from a general deterrence perspective (out of a larger set of five features that also included detection rate, publicity, and crime frequency). Thus, the use of these more sensitive, continuous scales revealed what Carlsmith et al.'s (2002) forced choice procedure could not, and substantially complicates the interpretation of Carlsmith et al.'s past research on the retributive motive, as well as other similar research (e.g., Aharoni & Fridlund, 2012). In light of the present results, that earlier research cannot be interpreted as providing unambiguous evidence for the existence and strength of retributive motives.

The present studies, which focus on animal offenders, are able to avoid some of the pitfalls that are inextricably involved when studying the punishment of human actors. In particular, general deterrence is essentially removed from consideration as a relevant punishment motive since deterrence is predicated on complex systems of communication which are lacking in most, if not all, nonhuman animal species. Thus, the task of isolating retribution for animal offenders is simpler than it is for human offenders because it involves separating the retributive motive only from the motive to incapacitate.

In order to highlight the role of retribution, we manipulated the victim (if any) of an animal's attack (Studies 2 and 3). We found that individuals were more supportive of killing an animal, and more likely to think that an animal deserved to be killed, as the victim of that animal's attack increased in perceived significance (the "victim identity" effect). This result held even though, on a priori grounds, it is not plausible to think that an offending animal, like a shark, should be any more dangerous having killed a more sympathetic victim. The victim identity effect appears straightforwardly to corroborate a basic retributive principle of proportionality—the notion that punishment is owed to an offender in proportion to the severity of his wrong (or alternatively, his blameworthiness).

Study 3 replicated this victim identity effect while also showing that participants found the killing of the actual "guilty" animal more deserved than the killing of a nearly identical yet "innocent" animal from the same species (the "targeted punishment" effect). Since retribution focuses on punishing only the guilty (see e.g., Flew, 1954; Mabbott, 1939; Quinton, 1954; Rachels, 1997; but cf. Cottingham, 1979), participants' clear concern for guilt suggests that their judgments were motivated by retribution. By contrast, from an incapacitationist perspective, if two animals are equally dangerous to humans, it should not matter whether one of them has previously attacked and killed a human, or not. All that should matter is their future potential to do so again.[9]

Both of these effects—the victim identity effect and the targeted punishment effect—remained highly reliable having accounted for differential perceptions of the sharks' dangerousness. And the targeted punishment effect was mediated by the perception that killing the correct shark would make better amends (a construct rooted in the notion of justice) for its killing of the victim. We, therefore, interpret both effects as corroborating the existence of retributive motives directed toward animals.

The most compelling evidence that our participants' punishment decisions reflect retribution, rather than solely consequentialist concerns, comes from Studies 4 and 5. Study 4 extended the targeted punishment effect by showing that individuals were more accepting of the infliction of pain and suffering in the responsive killing of a guilty animal (one that attacked a human) than in the responsive killing of an almost identical but innocent animal. This result cannot be explained on incapacitationist grounds, since the death of the relevant animal was assured in each case. Moreover, mediation analysis showed that measures of retributive sentiment and not perceptions of the shark's dangerousness, or the importance of incapacitating it, best explained this effect. In a similar vein, Study 5 extended the victim identity effect by showing that individuals were more accepting of the infliction of pain and suffering on a shark that killed a more sympathetic victim. As

in Study 4, this effect was mediated only by retributive sentiment, and not by perceptions of dangerousness or the the importance of incapacitation. Therefore, it, too, cannot be explained on incapacitationist grounds.

Other explanations for our findings—outside of the realm of incapacitative or deterrence theories—are also unconvincing when compared to a retribution-based account. For example, one alternative is that individuals' apparent differential support for punishing the animal attackers based on the level of sympathy evoked by the victim merely reflects a differential desire to express sympathy in a symbolic way. According to this view, the punishment judgments are simply abstract expressions of perceptions of the value of the lives that were ended (it is known for instance, that 10-year-old lives are seen as especially valuable, Goodwin & Landy, 2014). However, while plausible at face value, this alternative account cannot explain our full pattern of data. For one thing, while it may represent a possible alternative explanation for the victim identity effect, it cannot explain the targeted punishment effect, since the animal's victim was held constant in those cases. Moreover, even for the victim identity effect, the problem that this and other alternative explanations face, is that participants' punitive responses were mediated by explicitly retributive measures (for instance, the importance of avenging, making amends for, and seeking retribution for the shark's attack). This mediation evidence is directly predicted by the retributivist account, but cannot easily be explained by the "symbolic expression" alternative. If anything, this alternative account would predict that the measures tapping the general importance of incapacitating the offending animal should have mediated the victim identity effect, which they did not.[10]

Another alternative is that perhaps participants thought the punishment of animal attackers would deter would-be *human* offenders who might learn of the animal's punishment (we thank an anonymous reviewer for bringing up this concern). However, this account also faces difficulties. First, it seems unlikely that the punishment of an animal (as opposed to a human) would be a matter of public record in such a way that it could effectively deter humans. This is particularly the case for the assignment of pain to the animal—it is possible that a newspaper might report on the capture and killing of a shark, but it is not plausible that it would report on its exact method of killing, including the likely pain it was caused. It is, therefore, hard to see how this information could be thought to deter humans, especially given that it is improbable that many members of the public would even make the connection between a shark attack and their own decisions about whether or not to commit a crime. Second, this deterrence-based account also cannot explain the mediation of our effects via the retributive variables (although we acknowledge it would be interesting in future studies to include deterrence-focused mediators to gauge their effects).

One further potential concern regarding our findings is that the mean levels of punitive responses were quite low on our scales (typically below the mid-points of our scales), which may call into question the extent to which the data indicate a genuinely retributive motive. However, while these data may indicate something about the *degree* of retributive sentiment that our participants expressed on aggregate, we do not see them as impinging on our central claim that the studies document the existence of retributivism. Our core

argument hinges only on whether or not there are condition differences between various victims (*victim identity* effect) and various agents (*targeted punishment* effect) with respect to our key dependent measures, not on the overall levels of response on these measures. Retributivism, as we are using the term, refers to punishment responses driven by a notion of moral balance or proportionality—punishment in accordance with an offender's just deserts—which does not require an extreme or vicious response. Moreover, there was considerable variation within our samples such that while many, if not most, of our participants did not express active or extreme vengeance toward animals, there were some individuals who expressed more extreme responses. The key point is that, on aggregate, and based on the differences between means across our conditions, the data do indicate that participants expressed some measurable amount of retributive sentiment, exactly as we predicted. As a result, we think the most accurate interpretation of our effects is that they indicate a fairly widespread, but possibly low-level, and probably implicit, retributivism directed toward animals.[11]

7.2. Retributive sentiment can extend to nonhuman actors

The second main theoretical contribution of our work is to show that the retributive motive can extend more broadly than has generally been assumed in prior theories of punishment—that is, to the punishment of nonhuman actors, namely animals. The results raise interesting new questions regarding the requisite mental states and capacities for an agent to be deemed an appropriate target of retribution. In general, people presumably do not think that sharks (and most other animals) have the moral capacity to be able to distinguish right and wrong actions, yet we nonetheless observed responses to shark attacks that indicate retributive motives. These results are undoubtedly puzzling. However, we surmise that they might, in part, be explained by participants' ascribing relevant mental states (and attributes) to animals (e.g., some kind of low-level purposefulness or intentionality). While these mental states may fall short of the traditional "guilty mind" (*mens rea*) standard required for criminal culpability under the law, the ascription of such states may enable participants to view animals as appropriate targets of retribution.

This hypothesis, though untested in our current data, is consistent with a large body of recent findings on anthropomorphization and mental state attribution. Researchers have shown, for instance, that individuals are liable to attribute human-like traits and capacities to non-human agents, including animals, under certain conditions (Epley, Akalis, Waytz, & Cacioppo, 2008). Gray, Gray, and Wegner (2007) found that although animals (such as a chimpanzee and a dog) were generally perceived to be low on a dimension of mind they termed "agency," which involves such capacities as self-control, morality, memory, and planning, these animals were not seen as entirely lacking in such attributes (see also Piazza, Landy, & Goodwin, 2014). Individuals appear to be especially prone to find agency following moral events—that is, instances in which harm has occurred or a benefit has accrued (Gray & Wegner, 2010). This suggests that one reason why our participants were retributive toward animals is that, following a bad event, they attributed mental

states relevant to blameworthiness (e.g., some degree of intentionality or purposefulness) to animal offenders.

In a related vein, Rosset (2008) found that people's interpretations of actions are guided by an "intentionality bias," such that the default interpretation of most actions is that they are intentional. This appears to be particularly true with respect to negative events (Knobe, 2005; Morewedge, 2009). Thus, this research also suggests that when individuals make sense of an animal's attack, they may be prompted to view the animal's actions as reflecting a human-like mental state of deliberateness or intentionality.

7.3. Relation to other relevant existing findings

The victim identity effect observed in our studies is somewhat similar to the effects observed in the literature on outcome bias, severity effects, and moral luck (see e.g., Alicke, Weigold, & Rogers, 1990; Baron & Hershey, 1988; Burger, 1981; Royzman & Kumar, 2004; Rucker, Polifroni, Tetlock, & Scott, 2004; Walster, 1966; Williams & Nagel, 1976). In one recent study, for instance, it was shown that participants are liable to punish an individual on the basis of an outcome that he only had partial control over —and that individuals will punish an individual based on the negative outcomes he has caused, even when his intentions were benevolent (Cushman, Dreber, Wang, & Costa, 2009). The present results could be described in a broadly similar way—as illustrating a kind of outcome bias with respect to the punishment of animals.

However, our studies extend beyond what is known about outcome biases in three ways. First, our work is the first that we know of to document such effects for animal perpetrators. Second, in previous work, experimenters have usually focused on manipulating causation (rather than concerning themselves with the nature of an action's harmful consequences). For example, in one prior study an individual intended to poison a competitor and, through pure chance, the competitor was (a) killed by the poison, (b) not killed by the poison, or (c) killed as a result of some action by a third party (see e.g., Cushman, 2008, Experiment 4). In contrast, the causal element remained constant with respect to our victim identity effect: the animal always attacked and caused harm. What differed was the perceived significance of the animal's victim, if any. Thus, our focus is on better isolating how altering the *significance of a harm* (rather than *whether or how it occurs*) affects the potency of the retributive motive. Third, and relatedly, such outcome biases have rarely been demonstrated with respect to the particular type of *victim* an individual happens to kill. One exception is a study by Alicke and Davis (1989), which showed that people want to punish someone who killed a dangerous criminal (in self-defense) less severely than they want to punish someone who killed an innocent person (also in self-defense). The "victim effect" in this study is somewhat similar to ours, although unlike in the present study, it is possible that participants in Alicke and Davis's (1989) study were motivated by deterrence considerations—that is, by the belief that punishment outcomes ought to send the signal to the community that it is worse for *people* to kill innocent victims (note that it was not Alicke and Davis's aim to rule out such an interpretation). No such deterrence considerations can explain our results.

In documenting the victim identity effect, the present studies offer strong experimental support for previous theories advanced by economists to explain patterns in actual sentencing data. Using records from the Bureau of Justice Statistics, Glaeser and Sacerdote (2003) noted a correlation between victim characteristics and the severity of sentences in homicide cases. The authors suggested that this might have to do with a visceral vengeance response focused on the nature of the victim, but the data were unable to provide a clear causal mechanism because of the complexity of real life criminal prosecutions (e.g., the killer of a little girl might be given a longer sentence than the killer of an older man because prosecutors work harder and are, thus, more effective advocates when the victim is a little girl, not because the nature of the victim naturally leads judges or jurors to be more retributive). By controlling for alternative explanations, our experimental work helps resolve the issue and directly aligns with recorded trends in actual United States punishment practices.

7.4. Concluding remarks

Studies of the punishment of human actors are notoriously difficult to interpret because of the relevance of deterrence as a motive for punishing offenders. By studying the punishment of animals, and, consequently, by taking deterrence off the table, we have provided more decisive evidence for the existence of retributive punishment, while also showing that the existing understanding of the scope of retribution has been too narrow. Retributive motives can, and do, extend to nonhuman targets.

Acknowledgments

This research is an output from an NSF grant, award no. SES-1228231, awarded to both authors. The authors thank Dena Gromet, Justin Landy, and Jonathan Phillips for their valuable comments on an earlier draft of this manuscript, Sara Ghebremariam, Sarah Patrick, Elizabeth Smarrelli, and Jeffrey Stanton for their research assistance, and Alex Geisinger, Jon Hanson, Joe Simmons, and Uri Simonsohn for their thoughtful comments on this research. The authors also extend their gratitude to participants at the Harvard Law School SALMS Lecture Series, the Conference on Empirical Legal Studies (CELS), the Moral Research Lab (MoRL), the Drexel American Constitution Society Lecture Series, the University of Tulsa College of Law Faculty Workshop Series, the Pace University School of Law Faculty Colloquium Series, the Brooklyn Law School Faculty Workshop Series, and the Vanderbilt Law School Criminal Justice Roundtable for their valuable insights.

Notes

1. The use of a forced choice methodology presents a further, related problem, which is that participants may have felt obliged to balance out their answers across the

different response categories. Given that detection rate and publicity are clearly more relevant to deterrence than to retribution, a desire for balanced responding may have led participants to place harm severity and perpetrator motivation in the retribution category more than they might have done if presented with a different option set.

2. Although Carlsmith et al. (2002) did not include crime frequency in their study, we included it because we suspected that it would be seen as particularly relevant to general deterrence (which, indeed, it was).
3. While investigating people's sensitivity to crime severity may yield ambiguous evidence with regard to the retributive motive in the case of human offenders, it provides a much clearer test for animal offenders, given that general deterrence is removed from consideration (see earlier discussion).
4. A shark was one of the attackers in the present study.
5. To construct these four assignments, we made an initial random assignment of the four attackers to the four consequences. This initial assignment comprised one of the four versions of the study. To create the remaining three versions, while preserving the order of both attacker and consequences lists, we fixed the list of attackers and simply shifted the list of consequences through three different permutations until the initial assignment was reached again.
6. In the interests of clarity, across all of the studies, we refer to this as the "victim identity" manipulation despite the fact that in one of the four cases in Study 2—property damage—the perpetrator did not kill a victim.
7. For all of the analyses in this study that test for trend effects on the victim identity variable, we ran nonparametric Kendall's tau analyses because there is no assurance that the levels of the victim identity variable are equally spaced (making parametric linear trend tests questionable). Kendall's tau produces equivalent results to Jonckheere's nonparametric trend test, but we prefer it to the Jonckheere test because it also produces a measure of effect size.
8. The high punishment for the human who caused property damage likely reflects the fact that the man was essentially a terrorist who intended to seriously harm or kill innocent people. The high punishment for the pitbull that caused property damage may reflect strong negative background beliefs about pit bulls held by some participants.
9. Of course, a prior attack could influence perceptions of future dangerousness, which is why we measured and accounted for perceptions of dangerousness in our studies.
10. A possible concern with our mediation evidence is that the relevant mediators were all measured prior to the main dependent variables. This ordering reflects the purported causal sequence of mental events, and adheres to an established practice of mediation analysis (see e.g., Preacher & Hayes, 2008b, p. 36). While this method raises the possibility that the mediators may have influenced the dependent variable in some way, with respect to the present data, this general concern does not explain why only the retributive mediators were found to explain punitive judgments in

both Studies 4 and 5. Nevertheless, it may be useful in future studies to counterbalance whether these mediators are measured before or after the main dependent variables.

11. It also bears mention that there are several reasons why participants might be unlikely to express high responses on many of our dependent variables. First, retribution is but one of a number of motivations that likely influenced participants' assessments (i.e., incapacitation concerns likely drove some component of the responses as well, despite not fully explaining the condition differences). Second, the killing of animals for nonfood purposes is typically frowned upon in modern Western societies, which may have led some participants to be reluctant to express any such strong desire. Third, participants were likely aware, at some level, that indicating that a particular animal strongly deserved to be killed might appear irrational, given that most people presumably do not really believe that animals are capable of distinguishing right from wrong in the first place.

References

Aharoni, E., & Fridlund, A. J. (2012). Punishment without reason: Isolating retribution in lay punishment of criminal offenders. *Psychology, Public Policy, and Law*, *18*, 599–625.

Alicke, M. D., & Davis, T. L. (1989). The role of a posteriori victim information in judgments of blame and sanction. *Journal of Experimental Social Psychology*, *25*, 362–377.

Alicke, M. D., Weigold, M. F., & Rogers, S. L. (1990). Inferring intentions and responsibility from motives and outcomes: Evidential and extra-evidential judgments. *Social Cognition*, *8*, 286–305.

Baron, J. (2009). Belief-overkill in political judgments. *Informal Logic*, *29*, 368–378.

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, *54*, 569–579.

Baron, J., & Ritov, I. (2009). The role of probability of detection in judgments of punishment. *Journal of Legal Analysis*, *1*, 553–590.

Beccaria, C. (1963). *On crimes and punishments* (H. Paolucci, trans.). Englewood Cliffs, NJ: Prentice Hall. (Original work published in 1764).

Bentham, J. (1962). Principles of penal law. In J. Bowring (Ed.), *The works of Jeremy Bentham* (pp. 365–580). Edinburgh: W. Tait. (Original work published 1843).

Berman, P. S. (1994). Rats, pigs, and statues on trial: The creation of cultural narratives in the prosecution of animals and inanimate objects. *New York University Law Review*, *69*, 288–326.

Bondeson, J. (1999). *The feejee mermaid and other essays in natural and unnatural history*. Ithaca, NY: Cornell University Press.

Burger, J. M. (1981). Motivational biases in the attribution of responsibility for an accident: A meta-analysis of the defensive-attribution hypothesis. *Psychological Bulletin*, *90*, 496–512.

Carlsmith, K. M. (2006). The roles of retribution and utility in determining punishment. *Journal of Experimental Social Psychology*, *42*, 437–451.

Carlsmith, K. M. (2008). On justifying punishment: The discrepancy between words and actions. *Social Justice Research*, *21*, 119–137.

Carlsmith, K. M., & Darley, J. M. (2008). Psychological aspects of retributive justice. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 40, pp. 193–236). San Diego, CA: Elsevier.

Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology*, *83*, 284–299.

Cottingham, J. (1979). Varieties of retributivism. *The Philosophical Quarterly*, *29*, 238–246.
Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*, 353–380.
Cushman, F., Dreber, A., Wang, Y., & Costa, J. (2009). Accidental outcomes guide punishment in a "trembling hand" game. *PLoS ONE*, *4*, e6699.
Darley, J. M., & Pittman, T. (2003). The psychology of compensatory and retributive justice. *Personality and Social Psychology Bulletin*, *7*, 324–336.
Epley, N., Akalis, S., Waytz, A., & Cacioppo, J. T. (2008). Creating social connection through inferential reproduction. *Psychological Science*, *19*, 114–120.
Evans, E. P. (1906). *The criminal prosecution and capital punishment of animals*. London: William Heinemann.
Finkelstein, J. J. (1981). The ox that gored. *Transactions of the American Philosophical Society*, *71*, 1–89.
Flew, A. (1954). The justification of punishment. *Philosophy*, *29*, 291–307.
Girgen, J. (2003). The historical and contemporary prosecution and punishment of animals. *Animal Law*, *9*, 97–133.
Glaeser, E. L., & Sacerdote, B. (2003). Sentencing in homicide cases and the role of vengeance. *The Journal of Legal Studies*, *32*, 363–382.
Goodwin, G. P., & Landy, J. F. (2014). Valuing different human lives. *Journal of Experimental Psychology: General*, *143*, 778–803.
Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*, 619.
Gray, K., & Wegner, D. M. (2010). Blaming god for our pain: Human suffering and the divine mind. *Personality and Social Psychology Review*, *14*(1), 7–16.
Humphrey, N. (2003). *The mind made flesh: Essays from the frontiers of psychology and evolution*. New York: Oxford University Press.
Hyde, W. W. (1916). The prosecution and punishment of animals and lifeless things in the middle ages and modern times. *University of Pennsylvania Law Review*, *64*, 696–731.
Jervis, R. (1976). *Perception and misperception in international politics*. Princeton, NJ: Princeton University Press.
Judd, C. M., Kenny, D. A., & McClelland, G. H. (2001). Estimating and testing mediation and moderation in within-subject designs. *Psychological Methods*, *2*, 115–134.
Kant, I. (1952). The science of right (W. Hastie, trans.). In R. Hutchins (Ed.), *Great books of the western world* (pp. 397–446). Edinburgh: T. & T. Clark (Original work published 1790).
Kant, I. (1991). The metaphysics of morals (H. B. Nisbet, trans.). In H. Reiss (Ed.), *Kant: Political writings* (pp. 131–175). Cambridge, England: Cambridge University Press (Original work published 1797).
Knobe, J. (2005). Theory of mind and moral cognition: Exploring the connections. *Trends in Cognitive Sciences*, *9*, 357–359.
Mabbott, J. D. (1939). Punishment. *Mind*, *48*, 152–167.
Mill, J. S. (1998). Utilitarianism (R. Crisp, Ed.) Oxford, England: Oxford University Press. (Original work published in 1871).
Morewedge, C. K. (2009). Negativity bias in attribution of external agency. *Journal of Experimental Psychology: General*, *138*, 135–145.
Piazza, J., Landy, J. F., & Goodwin, G. P. (2014). Cruel nature: Harmfulness as an important, overlooked dimension in judgments of moral standing. *Cognition*, *131*, 108–124.
Preacher, K. J., & Hayes, A. F. (2008a). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879–891.
Preacher, K. J., & Hayes, A. F. (2008b). Contemporary approaches to assessing mediation in communication research. In A. F. Hayes, M. D. Slater, & L. B. Snyder (Eds.), *The Sage sourcebook of advanced data analysis methods for communication research* (pp. 13–54). Thousand Oaks, CA: Sage Publications.
Quinton, A. M. (1954). On punishment. *Analysis*, *14*, 133–142.

Rachels, J. (1997). Punishment and desert. In H. LaFollette (Ed.), *Ethics in practice* (pp. 470–479). Oxford, England: Blackwell.

Ristroph, A. (2005). Proportionality as a principle of limited government. *Duke Law Journal*, *55*, 263–331.

Rosset, E. (2008). It's no accident: Our bias for intentional explanations. *Cognition*, *108*, 771–780.

Royzman, E., & Kumar, R. (2004). Is consequential luck morally inconsequential? Empirical psychology and the reassessment of moral luck. *Ratio*, *17*, 329–344.

Rucker, D. D., Polifroni, M., Tetlock, P. E., & Scott, A. L. (2004). On the assignment of punishment: The impact of general-societal threat and the moderating role of severity. *Personality and Social Psychology Bulletin*, *30*, 673–684.

Walster, E. (1966). Assignment of responsibility for an accident. *Journal of Personality and Social Psychology*, *3*, 73–79.

Williams, B. A. O., & Nagel, T. (1976). Moral luck. *Proceedings of the Aristotelian Society*, Supplementary Volumes, *50*, 115–151.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1. Exact wordings of scenarios and questions, and additional data analyses.