Why Do God and Humans Punish? Perceived Retributivist Punishment Motives Hinge on Views of the True Self

Young-eun Lee, James P. Dunlea and Larisa Heiphetz

Department of Psychology, Columbia University

1190 Amsterdam Ave., New York, NY 10027, United States

Lee, Y., Dunlea, J. P., & Heiphetz, L. (2024). Why do God and humans punish? Perceived retributivist punishment motives hinge on views of the true self. *Personality and Social Psychology Bulletin*, *50*, 1167-1181. doi: 10.1177/01461672231160027

Author Note
Correspondence concerning this article should be addressed to Young-eun Lee at yl4898@columbia.edu.

Acknowledgments
This project was made possible (in part) through the support of a Visiting Scholar Award from the Russell Sage Foundation, grants #61080 and #61808 from the John Templeton Foundation, NSF CAREER Award #2044360, and Columbia University to LH. Any opinions expressed here are those of the authors alone and should not be construed as representing the opinions of any organizations that provided support for this project. The authors wish to thank members of the Cooperation Lab, the Morality Lab, and the Social and Moral Cognition Lab for providing feedback on this project.

Abstract

Laypeople often believe that God punishes transgressions; however, their inferences about God's punishment motives remain unclear. We addressed this topic by asking laypeople to indicate *why* God punishes. We also examined participants' inferences about why humans punish to contribute to scholarly conversations regarding the extent to which people may anthropomorphize God's mind. In Studies 1A-1C, participants viewed God as less retributive than humans. In Study 2, participants expected God (versus humans) to view humans' true selves more positively; this difference mediated participants' views of God as less retributive than humans. Study 3 manipulated agents' views of humans' true selves and examined how such information influenced each agent's perceived motives. Participants viewed a given agent as less retributive when that agent regarded the true self as good (versus bad). These findings extend scholarship on lay theories of punishment motives and highlight links between religious and moral cognition.

Keywords: moral cognition; pre-registered; punishment; religious cognition; true self

Why Do God and Humans Punish? Perceived Retributivist Punishment Motives Hinge on Views of the True Self

Punishment is a central component of humans' psychological repertoire (Fehr & Fischbacher, 2003). Across cultures and development, people indicate that individuals who break laws, rules, or widely held moral norms should receive some sort of punishment for their actions (e.g., Dunlea & Heiphetz, 2021; Henrich et al., 2010; Marshall et al., 2021). Some scholars suggest that punitive desires are functional: meting out punishment may help sustain cooperation among humans (Balliet et al., 2011; Mathew & Boyd, 2011). Evolutionary accounts of religion suggest that the "costliness" associated with humans enacting punishment, in part, prompted the widespread and enduring cultural belief in supernatural entities (e.g., God) who can enact punishment (for a review, see Laurin et al., 2012). While people endorse the idea that both earthly and divine agents can enact punishment, relatively less work has focused on people's judgments about *why* such agents choose to punish.

The current work began to address this topic by comparing the extent to which people attribute different punishment motives to another human and, separately, God. Philosophical theories of punishment often highlight two theories of justice: *retributivism*, arguing that transgressors should suffer in proportion to their moral offense (Hume, 1739/1888), and *consequentialism*, arguing that punishment is justified only if it confers social benefits (e.g., prevention of future harms; Bentham, 1823/1970). Building on philosophers' normative approach of asking why people should punish, a robust literature within psychology has examined the extent to which punishment satisfies people's retributivist and, separately, consequentialist motives (e.g., Carlsmith, 2008; Darley, 2009; Funk et al., 2014; Marshall et al., 2021). To contribute to scholarship on punishment motives, we also focused on people's

inferences regarding different agents' retributivist motives, with several studies additionally investigating consequentialism.

Comparing perceptions of God's and another human's punishment motives was important for two main reasons. First, religion plays an important role in people's lives. Most people around the globe identify with a religious group (Pew Research Center, 2015), and many of these people believe in supernatural agents (e.g., God) that intervene in human affairs (e.g., Johnson & Bering, 2006). Despite its importance, religion is fairly understudied within psychology. By probing people's inferences about God, we advance the scientific understanding of a central component of human life.

Second, asking about God's punishment motives can inform scientific understanding about why people may consider God to be a moral agent (Ginges et al., 2016; Heiphetz et al., 2018; Pasek et al., 2020). Perceiving punishment motives as retributivist or consequentialist has significant consequences in how people view punishers; for instance, people consider a punisher with consequentialist motives as more moral and trustworthy than a punisher with retributivist motives (Dhaliwal et al., 2022). However, religious texts often provide readers with conflicting information about God's punishment motives. Some religious texts describe God as retributive (e.g., "I am about to pour out my wrath on you and spend my anger against you. I will judge you according to your conduct and repay you for all your detestable practices" from Ezekiel 7:8). In contrast, other religious texts teach that God punishes with consequentialist motives (e.g., "My son, do not despise the Lord's discipline, and do not resent his rebuke, because the Lord disciplines those he loves, as a father the son he delights in" from Proverbs 3:11-12). Therefore, investigating perceptions of God's punishment motives could clarify which aspects of official

religious teachings align more with laypeople's religious cognition and potentially shed light on why people view God as a moral agent.

The present work examined how people perceive God's punishment motives. For comparison, we also assessed how people perceive another human's punishment motives. Doing so allowed us to contribute to ongoing conversations in psychology regarding the extent to which people anthropomorphize—or attribute human-like characteristics to—God's mind (Ginges et al., 2016; Heiphetz et al., 2018; Richert et al., 2016). Past work leads to three possibilities regarding the extent to which people attribute similar punishment motives to the minds of earthly and divine agents.

One possibility is that people may view God and humans as possessing similar retributivist *and* consequentialist motives. We refer to this possibility as the *similarity account*. This possibility stems from scholarship suggesting that, under some conditions, people view God's mind as similar to that of a human (e.g., Knight et al., 2004; Shtulman & Lindeman, 2016). For instance, people sometimes attribute similar morally-relevant mental states (e.g., believing that helping another person is morally acceptable) to other humans and God (Heiphetz et al., 2018). Given that people sometimes ascribe similar morally-relevant beliefs to both God and humans, people may infer that God and humans also possess similar punishment motives.

A second possibility is that people may view God and humans as possessing different punishment motives. We refer to this possibility as the *divergence account*. Theologians and religious studies scholars often argue that God's mind is strikingly different from the minds of humans (Armstrong, 1993). Similarly, when asked to indicate their views of God, adults sometimes provide "theologically correct" answers (Barrett, 1999; Bering & Johnson, 2005). For example, adults living across different cultures, including China and Italy, readily indicate that

God's mind is more powerful than the minds of humans (e.g., compared to humans, God has a stronger ability to reason and plan; Demoulin et al., 2008; Haslam et al., 2008). Furthermore, religious texts often portray God's motives as different from those of humans (e.g., "'For my thoughts are not your thoughts, neither are your ways my ways,' declares the Lord. 'As the heavens are higher than the earth, so are my ways higher than your ways and my thoughts than your thoughts'" from Isaiah 55:8-9). If people in fact distinguish God's mind from human minds, they may attribute different punishment motives to God versus to another person.

A third possibility is that people may report both similarities and differences between humans' and God's punishment motives. For instance, people may infer that God and humans possess similar consequentialist motives, but that God and humans differ in their desires to mete out punishment for retributivist reasons. We refer to this possibility as the *mixed account.* People's reasoning about complex topics (e.g., religion) often reflects bundles of co-existing concepts—those that align with information conveyed via testimony and those that instead align with lay theories of the social world (Legare et al., 2012; Shtulman & Lombrozo, 2016). For example, people tend to perceive God as more moral and benevolent compared to other humans (Ginges et al., 2016; Pasek et al., 2020). Furthermore, people consider those with retributivist motives to be less moral than those with consequentialist motives (Dhaliwal et al., 2022; Herrmann et al., 2008). Because retributivism is often associated with less morality compared to consequentialism, people might perceive God to hold less retributivist motives than humans. The current work tested among these three competing possibilities—the similarity account, divergence account, and mixed account.

In addition to probing how participants perceive God's punishment motives, we also asked *why* participants might hold the perceptions that they do. Specifically, we examined the

role that concepts of the true self might play in participants' responses. People often view their own and others' true selves—who they really are deep, deep down inside—as morally good (Bench et al., 2015; De Freitas et al., 2018; Newman et al., 2014). However, judgments about others' true selves may be heterogenous (i.e., people view some individuals' true selves as better than others' true selves; Heiphetz, 2020; Zhang & Alicke, 2021). Furthermore, viewing others' true selves as immoral can have deleterious consequences: perceived badness typically augments, whereas perceived goodness buffers against, the likelihood of outcomes such as severe punishment and unfavorable attitudes towards individuals who receive punishment (e.g., Carlsmith & Sood, 2009; Dunlea & Heiphetz, 2022; Heiphetz, 2020; Nadler & McDonnell, 2011). For example, when people think the agent views humans' true selves as morally bad (versus good), they might expect the agent to show more severe punishment motives. Because the goal of retributivist punishment is to harm transgressors by making them suffer in proportion to their moral offense (Hume, 1739/1888), people might infer that the agent who believes humans' true selves as bad would show retributivist punishment motives.

Overview of the Current Research

We examined laypeople's inferences about different agents' punishment motives. Studies 1A-1C addressed this topic by investigating the extent to which participants attribute retributivist and, separately, consequentialist punishment motives across earthly and divine minds. Study 2 extended the results of Studies 1A-1C by probing the extent that one factor—namely, participants' inferences about how agents view humans' true selves—underlies the main difference in perceived punishment motives found in Studies 1A-1C. Study 3 built on Study 2's correlational approach by manipulating agents' views of humans' true selves and subsequently examining how such information influenced each agent's perceived motives. Together, these

studies shed light on people's lay theories of punishment motives, contribute to scholarly conversations regarding the extent to which people anthropomorphize God's mind, and highlight links between religious and moral cognition.

Study 1A

In an initial pre-registered study (https://aspredicted.org/blind.php?x=sv9q67), we investigated the extent to which people attribute similar punishment motives across earthly and divine minds.

Method

Here and in all subsequent studies, we report all manipulations, measures, participant exclusions, and sample size determinations. Materials, anonymized data, analysis syntax, and codebooks for this and all subsequent studies can be accessed via an online repository (https://osf.io/c84xy/?view_only=a75cc5577199481d9855753d8e781f2d).

Participants. Our final sample included 475 adults between 18 and 72 years old (M_{age}=37.09 years, SD_{age}=11.50 years; 56% female, 44% male, <1% other).[1] A sensitivity analysis established that our sample size was large enough to detect small main and interaction effects (f=0.13) with sufficient power (0.80) at alpha .05. In this and subsequent studies, we used G*Power (Faul et al., 2007) to run a sensitivity analysis.[2]

[1] Our initial sample included 177 participants. However, during the review process, we received a request to increase statistical power. Therefore, to achieve a power of f = 0.125, we recruited 298 more participants from the same participant pool. The pattern of results with the original sample (n=177) is similar to the results reported in the main text (please see p. 11 in SOM for details about the findings with the original sample).

[2] Although the current study used a within-subject design and we analyzed the data as such, in the current and all subsequent studies, we conservatively ran sensitivity analysis based on a between-subject design. Because between-subjects designs require more participants than within-subjects or mixed designs to detect similarly-sized effects, we reasoned that if a between-subjects version of our analysis had sufficient power to detect the effects we observed, our actual (more sensitive) analyses also had sufficient power.

Participants self-identified as White or European-American (76%), Asian or Asian-American (9%), Black or African-American (7%), Native American or Pacific Islander (3%), Multiracial (2%), and other (1%), remainder unspecified. Additionally, 10% of adults self-identified as Hispanic or Latinx. Participants also reported their present religious affiliation as Catholic (29%), Protestant (19%), other Christian (10%), Jewish (3%), Muslim (<1%), other/not listed (3%), and non-religious/atheist/agnostic (37%; our demographic questionnaire grouped these options together in this and subsequent studies).

We recruited participants online via Amazon's Mechanical Turk, which we configured so that only United States residents whose approval rating was at least 95% could participate. Only those who did not participate in a similar study before were eligible for the study. Participants received $0.33 if they properly summarized a passage and correctly answered two attention check items presented throughout the testing session (i.e., items asking them to indicate which agent's perspective they adopted); otherwise, they received $0.05.[3] Consistent with our pre-registration, we excluded data from 346 participants because they failed to meet at least one of these criteria. We did not include excluded participants in the demographic breakdown above in this and all subsequent studies. Unless otherwise noted, the patterns of results reported below remained unchanged when including all respondents.

Procedure. First, participants read a vignette about a man named Michael Scarrow who killed another man (adapted from Shariff et al., 2014). Participants then read definitions of retributivism and consequentialism as motivations for punishment and, subsequently, indicated on two separate Likert scales the extent to which a person named Joan or John, gender-matched to the participant, would agree that retributivism and, separately, consequentialism is important

[3] The studies reported here were conducted during a period of time when norms for paying participants were changing in psychology. As a result, we increased our hourly pay after conducting Study 1A.

in punishing Scarrow (e.g., "I feel that retributivism should be an important motivation in punishing Michael Scarrow"). Hereafter, we refer to the human agent as Joan for brevity. Participants also indicated the extent to which God would agree that retributivism and, separately, consequentialism is important in punishing Scarrow. Values ranged from 1 (*Strongly disagree*) to 7 (*Strongly agree*). Participants answered all items from the perspective of one agent before answering items from the perspective of the other agent.

Participants answered each item from the perspective of another human or God. To make the script between agents identical as much as possible, and to help participants more fully imagine themselves into the mind of God or another person, we designed all items using first-person language to match the target sentence across agents (e.g., after instructing participants to adopt the perspective of God or another person, we asked them how much they agreed or disagreed with the statement, "*I* feel that retributivism should be an important motivation in punishing Michael Scarrow"). We adapted this method from scholarship probing people's lay beliefs about humans' and God's moral judgments (Ginges et al., 2016; Pasek et al., 2020).

After answering the main experimental items, participants indicated which agent's perspective they took when answering the items on the previous page. These questions served as an attention check in this and all subsequent studies. All participants included in the data analysis answered these items correctly. The order in which participants answered items from each agent's perspective and the order of experimental items (i.e., items about perceived retributivist and consequentialist motives) were counterbalanced across participants.

Results

All statistical analyses were conducted with R statistical software (R version 4.1.1; R Core Team, 2021) except for a mediation analysis (Study 2) with MEMORE (Montoya & Hayes, 2017), for which we used Version 28.0 of SPSS (IBM Corporation, 2021).

We first analyzed participants' views of each agent's punishment motives using a 2 (Agent: God vs. Joan) x 2 (Motive: retributivist vs. consequentialist) repeated measures ANOVA.[4] This analysis revealed a significant main effect of Agent, $F(1,474)=13.87$, $p<.001$, $\eta_p^2=.03$, which was qualified by a significant Agent x Motive interaction, $F(1,474)=13.88$, $p<.001$, $\eta_p^2=.03$. The main effect of Motive did not reach significance, $F(1,474)=1.99$, $p=.159$.

To examine the Agent x Motive interaction, we conducted two sets of tests (Figure 1). Consistent with our pre-registration, we first compared perceived retributivist and, separately, consequentialist motives across agents. This analysis resulted in two comparisons; therefore, after applying a Bonferroni correction, *p* values needed to be .025 or lower to remain significant. We applied this correction in this and subsequent studies. In line with the *mixed account*, participants perceived God as less retributive than Joan, $t(474)=-4.93$, $p<.001$, $d=0.23$, 95% CI_{diff}: [-.76, -.33]. However, we did not observe a significant difference in their perceptions of consequentialist motives between God and Joan, $t(474)=1.75$, $p=.081$, 95% CI_{diff}: [-.02, .40].

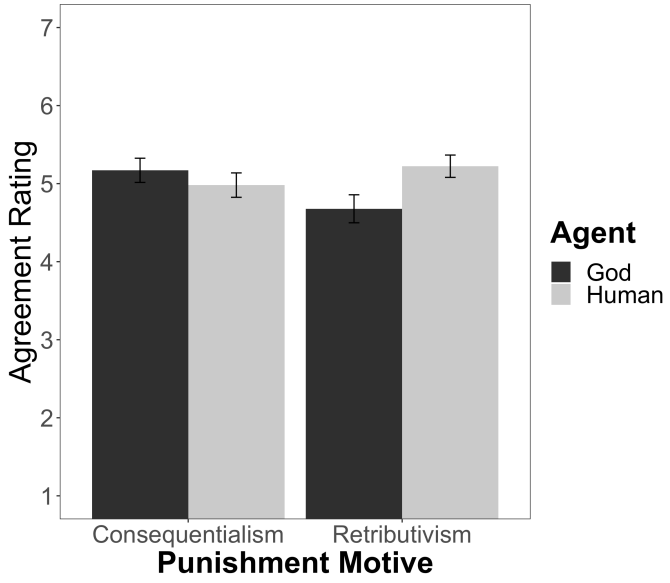
In addition to the pre-registered analyses, we conducted a series of exploratory analyses comparing perceived retributivist, versus perceived consequentialist, motives within each agent. Participants viewed God as more motivated by consequentialist rather than retributivist concerns; $t(474)=3.57$, $p<.001$, $d=0.16$, 95% CI_{diff}: [.22, .76]. In contrast, participants tended to view Joan

[4] Our pre-registration for Study 1 specified particular comparisons of theoretical interest rather than the omnibus ANOVA. However, to have a comprehensive understanding of the data, we first analyzed participants' views of each agent's punishment motives using a 2 (Motive: retributivist vs. consequentialist) x 2 (Agent Perspective: God vs. Joan) repeated measures ANOVA.

as more motivated by retributivism than by consequentialism with marginal significance; $t(474)=-1.90$, $p=.058$, $d=0.09$, 95% CI_{diff}: [-.49, .01].

Figure 1

Average Agreement with Different Punishment Motives by Agent Type in Study 1A



Note. Higher values reflect greater agreement that a given agent possesses a specific punishment motive. Error bars represent 95% confidence intervals.

Discussion

Study 1A investigated which punishment motives people attribute to God versus to another human. Two main findings emerged. First, participants viewed God as less retributive than Joan; however, a significant difference in perceived consequentialist motives between agents did not emerge. This pattern of results supports the *mixed account*, which is consistent with the interpretation that people may anthropomorphize some—but not all—aspects of God's mind when reasoning about punishment motives. This result is also consistent with previous findings that people view God as more moral than humans (Ginges et al., 2016; Pasek et al.,

2020) and that retributivist punishment motives signal less moral traits (Dhaliwal et al., 2022; Herrmann et al., 2008). In conjunction with past work, the current findings may be relevant to *why* people may view God as a particularly moral agent—i.e., people's views of God's less retributivist punishment motives, as compared to a human, may shape their views of God's morality.

Second, participants viewed God, but not Joan, as more motivated by consequentialist rather than retributivist concerns. This finding dovetails with scholarship suggesting that people's estimates of God's mental states are more egocentric than estimates of other people's beliefs (Epley et al., 2009). When asked to indicate their own punishment motives, laypeople generally report being motivated to punish for consequentialist as opposed to retributivist reasons (Carlsmith, 2008). Because this pattern emerged only in participants' judgments of God's punishment motives, the current finding, in conjunction with prior work documenting people's stated punishment motives (e.g., Carlsmith, 2008), is in line with the notion that laypeople's estimates of God's mental states may be more egocentric than estimates of other people's beliefs.

Study 1B

Study 1A found that, collapsing across individuals from different religious backgrounds, participants on average viewed God as less retributivist than another person. The main goal of Study 1B (see https://aspredicted.org/RY2_KPR for pre-registration) was to determine whether this result would generalize across participants from two specific religious traditions: Catholicism and Judaism. Catholics might be more likely than Jews to perceive God as retributivist—for instance, by reporting belief in a hell where God punishes people for their wrongdoings (Pew Research Center, 2014). Thus, testing members of these two religious groups allows us to determine the extent to which socialization into a particular view of a retributive

versus benevolent God shapes participants' attributions of retributivist punishment motives to God. For consistency with Study 1A, we also asked participants to attribute punishment motives to Joan.

Method

Participants. Our final sample included 388 adults between 18 and 79 years old (M_{age}=39.31 years, SD_{age}=15.80 years; 63% female, 34% male, 2% non-binary). A sensitivity analysis established that our sample size was large enough to detect small- to medium-sized main and interaction effects (f=0.14) with sufficient power (0.80) at alpha .05.

Half of our participants were Catholics (n=193), while the other half were Jews (n=195). Participants self-identified as White or European-American (89%), Multiracial (4%), Asian or Asian-American (4%), Black or African-American (2%), and other (1%). Additionally, 8% of participants self-identified as Hispanic or Latinx.

We recruited participants online via Prolific, an organization that allowed us to screen potential participants for their religious identity prior to allowing them to begin the study. We also configured our study so that only United States residents who passed two tasks designed to screen out bots could participate. However, we did not screen out participants based on approval rate. Participants received $1.34 if they correctly answered attention check items presented throughout the testing session (i.e., items asking them to indicate which agent's perspective they adopted); otherwise, they received no payment. Based on pre-registered criteria, we excluded data from 70 participants who answered these attention check items incorrectly or who reported prior participation in a similar study.

Procedure. The procedure in Study 1B was identical to Study 1A except for two main changes. First, we did not present questions about consequentialism because this item did not

differ significantly between God and Joan in Study 1A. Instead, we focused on replicating the difference in retributivism between God and Joan with members of two distinct religious groups.

Second, at the end of the study, participants completed the Views of God Scale (Shariff & Norenzayan, 2011) to confirm whether Jewish and Catholic participants reported, on average, different God concepts. Using a scale from 1 (*Strongly disagree*) to 7 (*Strongly agree*), participants rated their agreement with 7 items portraying God as benevolent (e.g., "God is loving") and 7 items portraying God as punitive (e.g., "God is punishing"). We then averaged the positive items (α=0.97) and negative items (α=0.89) to create a "Benevolent God" and "Punitive God" measure, respectively. In line with the original work by Shariff and Norenzayan (2011), we then subtracted the Benevolent God average from the Punitive God average to create an overall God negativity score, with larger numbers indicating more negative views of God.

Results

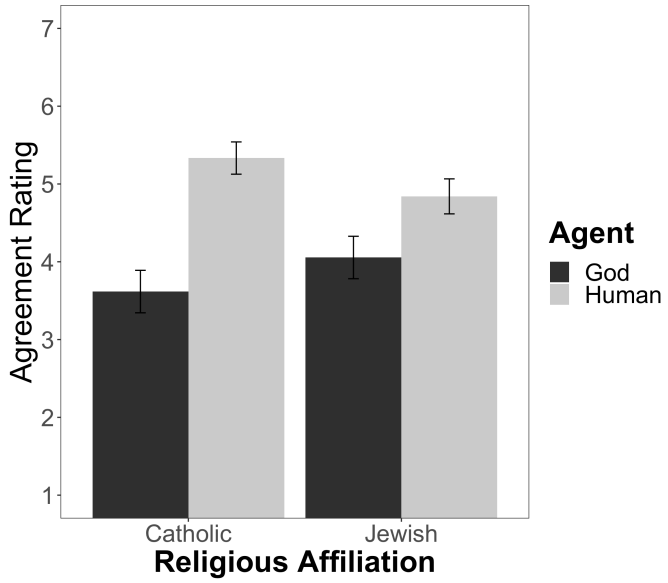
As pre-registered, we analyzed participants' perceptions of retributivist motives using a 2 (Agent: God vs. Joan) x 2 (Participant Religion: Catholic vs. Jewish) mixed ANOVA. We observed a main effect of Agent, $F(1,386)=139.63$, $p<.001$, $\eta_p^2=.27$, which was qualified by an Agent x Participant Religion interaction, $F(1,386)=22.92$, $p<.001$, $\eta_p^2=.06$ (Figure 2). The main effect of Participant Religion did not reach significance, $F(1,386)<1$, $p=.860$.

The interaction occurred because Catholic participants showed a stronger differentiation between God's and Joan's retributivist motive compared to Jewish participants, while both groups were more likely to attribute retributivist motives to Joan than to God; $t(192)=-11.14$, $p<.001$, $d=0.80$, 95% CI_{diff}: [-2.15, -1.51] for Catholic participants, and $t(194)=-5.27$, $p<.001$, $d=0.38$, 95% CI_{diff}: [-1.06, -.48] for Jewish participants.

Additionally, the result from the Views of God Scale confirmed that Catholic and Jewish participants hold significantly different views about God. Specifically, one-sample *t*-tests indicated that participants from both religious groups exhibited positive perceptions of God when compared against 0 ($t(192)=-16.82$, $p<.001$, $d=1.21$, 95% CI: [-3.24, -2.56] in Catholic participants and $t(194)=-12.08$, $p<.001$, $d=0.86$, 95% CI: [-2.15, -1.54] in Jewish participants). However, a paired-samples *t*-test revealed that although both religious groups showed positive God perceptions, Catholic participants ($M=-2.90$, $SD=2.39$) were more likely to perceive God positively than Jewish participants ($M=-1.85$, $SD=2.13$), $t(380.10)=-4.57$, $p<.001$, $d=0.46$, 95% CI_{diff}: [-1.51, -.60], indicating that Catholics would have relatively more positive God perceptions than Jews. Together with participants' attributions of punishment motives, this result shows that even members of two religious groups who hold significantly different views of God are both less likely to attribute retributivism to God than to a person.

Figure 2

Average Agreement with Retributivism by Agent Type and Religious Affiliation in Study 1B



Note. Higher values reflect greater agreement that a given agent possesses a retributivist punishment motive. Error bars represent 95% confidence intervals.

Discussion

Study 1B replicated the finding that participants were less likely to attribute retributivism to God than to Joan with two distinct religious samples. Both Catholics and Jews reported that God was less likely than Joan to punish for retributivist motives. Although Catholics showed a stronger distinction between God and Joan than did Jews, members of both religious groups perceived Joan as more retributive than God.

Study 1C

Building on Study 1B, which probed the extent to which Study 1A's findings would generalize across members of different religious groups, Study 1C (see https://aspredicted.org/blind.php?x=CNT_T64 for pre-registration) examined the degree to which findings would generalize across different types of transgressions. Specifically, whether a person commits religious versus secular transgression could affect perceptions of God's and Joan's retributivist motives. One possibility is that participants might perceive Joan to be more retributive than God regardless of whether a person violates religious rules or secular rules. Such a result would show that attributing greater retributivism to Joan than to God might reflect a general reasoning process that is not limited to specific transgression domains. Another possibility is that participants might perceive God to be more retributive than Joan when a person violates religious rules, but not when a person violates secular rules. This is plausible because people may perceive a stronger affront against God's desired behavior in a religious versus secular context.

Study 1C addressed these possibilities by probing the extent to which people attribute retributivist motives to God and Joan when reading about transgressions that violated religious versus secular rules. Because Study 1B revealed that participants from two religions with different concepts of God viewed God as more retributive than Joan, Study 1C and all subsequent studies allowed any interested respondent to participate and collapsed across participants from different religious backgrounds.

Method

Participants. Our final sample included 337 adults between 18 and 73 years old (M_{age}=38.34 years, SD_{age}=11.84 years; 54% female, 45% male, <1% other). A sensitivity analysis established that our sample size was large enough to detect small- to medium-sized main and interaction effects (f=0.15) with sufficient power (0.80) at alpha .05.

Participants self-identified as White or European-American (80%), Asian or Asian-American (7%), Black or African-American (5%), Native American or Pacific Islander (4%), Multiracial (3%), and other (1%). Additionally, 12% of participants self-identified as Hispanic or Latinx. Participants reported their present religious affiliation as Catholic (34%), Protestant (20%), other Christian (8%), Jewish (1%), Muslim (<1%), other/not listed (6%), and non-religious/atheist/agnostic (30%).

We recruited participants online via Amazon's Mechanical Turk, which we configured so that only United States residents whose approval rating was at least 98% could participate. Respondents were only eligible to participate if they passed two tasks screening out bots and if they had not participated in any similar prior studies. Participants received $1.33 if they correctly answered attention check items and $0.15 if they did not answer these items correctly. Based on

pre-registered criteria, we excluded data from 76 participants who did not provide correct responses to the attention check items.

Procedure. As in Study 1B, we measured participants' perceptions of God's and Joan's retributivist punishment motives; also as in Study 1B, we did not measure perceptions of consequentialist motives. We randomly assigned participants to complete either the religious transgression condition (n=165) or the secular transgression condition (n=172). In both conditions, participants read three vignettes—one about lying, one about theft, and one about murder—in randomized order. While the vignettes about lying and theft were new, the vignette about murder was identical to that in Studies 1A-1B. Importantly, we portrayed the transgression as violating religious rules (in the religious transgression condition) or secular rules (in the secular transgression condition). For example, in the lying vignette of the religious transgression condition, participants read the following:

> Across the world, religious rules often forbid people from lying in serious circumstances. For instance, the Bible states, "You shall not give false witness against your neighbor" (Exodus 20:16). This text means that it wrong to lie under some circumstances, and people who engage in certain types of dishonesty are deliberately breaking a religious rule.

After reading this script, participants read a story about a person who lied. Then, as in Study 1B, participants indicated the extent to which God and Joan would agree that retributivism is important in punishing the person who lied.

The secular transgression condition was identical to the religious transgression condition except that the religious text was replaced with legal texts and thus lying was framed as violating legal rules, as follows:

> Across the world, legal rules often forbid people from lying in serious circumstances. For instance, Title 18 in the U.S. Code states, "Whoever having taken an oath […] subscribes any material matter which he does not believe to be true […] is guilty of perjury." This

> text means that it wrong to lie under some circumstances, and people who engage in certain types of dishonesty are deliberately breaking a legal rule.

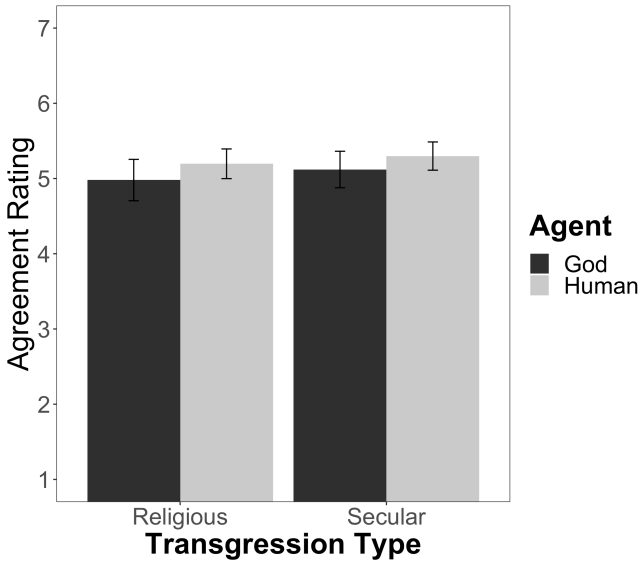
In each condition, participants completed two blocks in counterbalanced order: one block asking their perception of God's retributivist motive about lying, theft, and murder, respectively, and another block asking their perception of Joan's retributivist motive about the same transgressions. Participants answered all items from the perspective of one agent before answering items from the perspective of the other agent, and we randomized item order (i.e., order in which participants answered about lying versus theft versus murder) within blocks.

Results

As pre-registered, we analyzed participants' perceptions of retributivist motives using a 2 (Agent: God vs. Joan) x 2 (Transgression Type: Religious vs. Secular) mixed ANOVA. This analysis revealed a significant main effect of Agent, $F(1,335)=4.02$, $p=.046$, $\eta_p^2=.012$ (Figure 3). Overall, participants were more likely to attribute retributivism to Joan ($M=5.25$, $SD=1.26$) than to God ($M=5.05$, $SD=1.70$), which replicated findings from Studies 1A and 1B. No other main effects or interactions reached significance, $ps>.351$.

Figure 3

Average Agreement with Retributivism by Agent Type and Transgression Type in Study 1C



Note. Higher values reflect greater agreement that a given agent possesses a retributivist punishment motive. Error bars represent 95% confidence intervals.

Discussion

Regardless of whether we portrayed transgressions as violating religious or secular rules, participants were less likely to attribute retributivist motive to God than to Joan. The result suggests that the tendency to perceive less retributivism in God (versus Joan) generalizes across religiously- and secularly-framed transgressions.

Study 2

Study 1 established that participants attributed less retributivist punishment motives to God than to another person (Study 1A), that this effect occurred among both Jewish and Catholic participants (Study 1B), and that this effect generalized across transgressions framed as violating both religious and secular rules (Study 1C). The goal of Study 2 (see https://aspredicted.org/blind.php?x=zx7mt4 for pre-registration) was to examine a possible

mechanism that might underlie this effect. Specifically, we wondered whether judgments about moral character may underlie participants' propensity to view God, versus Joan, as less influenced by retributivist punishment motives. Because perceived badness is often associated with outcomes such as severe punishment and unfavorable attitudes towards individuals who receive punishment (e.g., Carlsmith & Sood, 2009; Dunlea & Heiphetz, 2022; Heiphetz, 2020; Nadler & McDonnell, 2011), it is possible that a belief in the good true self is associated with a decrease in perceived retributivism. Specifically, participants may expect God, versus Joan, to be more optimistic about people's true selves, and this difference may lead participants to view God as less retributive than Joan. Such a result would extend work on the "good true self" by demonstrating that participants expect God to view humans' true selves more positively than Joan. Study 2 tested this possibility.

Method

Participants. Our final sample included 494 adults between 19 and 87 years old (M_{age}=36.72 years, SD_{age}=11.71 years; 53% female, 46% male, <1% other, remainder unspecified).[5] A sensitivity analysis indicated that our sample size was large enough to detect small main and interaction effects (f=0.13) with sufficient power (0.80) at alpha .05. Additionally, prior recommendations for powering mediation models (Fritz & MacKinnon, 2007) indicated that our sample size was large enough to detect a small mediation effect with sufficient power (0.80) at alpha .05.

Participants self-identified as White or European-American (80%), Asian or Asian-American (8%), Black or African-American (7%), Native American or Pacific Islander (2%),

[5] Our initial sample included 265 participants. However, during the review process, we received a request to increase statistical power. Therefore, to achieve a power of f = 0.125, we recruited 229 more participants from the same participant pool. The pattern of results with the original sample (n=265) is similar to the results reported in the main text (please see p. 12 in SOM for details about the findings with the original sample)

Multiracial (2%), remainder unspecified. Additionally, 8% of adults self-identified as Hispanic or Latinx. Participants reported their present religious affiliation as Catholic (40%), Protestant (23%), other Christian (6%), Jewish (2%), Muslim (<1%), other/not listed (3%), and non-religious/atheist/agnostic (25%), remainder unspecified.

We recruited participants online via Amazon's Mechanical Turk, which we configured so that only United States residents whose approval rating was at least 95% could participate. Participants received $2.00 if they properly summarized a passage and correctly answered two attention check items asking them to recall questions they had answered earlier in the study; otherwise, they received $0.20. Consistent with our pre-registration, we excluded data from 475 participants because they failed to meet at least one of these criteria.

Procedure. In Block I, participants rated their agreement with 12 statements regarding a person's view of humans' true selves (e.g., "[John/Joan] believes that, at their core, people are morally good") and an analogous set of 12 statements regarding God's views (e.g., "God believes that, at their core, people are morally good"). We developed these statements by drawing on language used in prior scholarship on lay perceptions of the true self (De Freitas & Cikara, 2018; Newman et al., 2015). Participants indicated their responses on Likert scales with values ranging from 1 (*Not at all*) to 9 (*Very much so*) and answered all items about each agent before answering items about the other agent. In Block II, participants read the vignette about murder from Study 1A and rated the extent to which they expected Joan and God to endorse retributive and consequentialist punishment motives, respectively.

The order in which participants answered questions from each agent's perspective and the order of experimental items were counterbalanced across participants. To conform to recommendations for mediation designs (Baron & Kenny, 1986), all participants completed

Block I before Block II. After completing Blocks I and II, participants answered items for an unrelated study.

Results

Perceptions of the "true self." We first examined the extent to which participants expected Joan and God to have different views about the true self of humans. After determining that each set of items probing views of the true self of humans had acceptable reliability, we collapsed across items measuring Joan's views (α=.91) and, separately, God's views (α=.93).

One-sample *t*-tests indicated that participants expected that both Joan, $t(493)=12.56$, $p<.001$, $d=0.56$, 95% CI: [5.51, 5.70], and God, $t(493)=11.88$, $p<.001$, $d=0.53$, 95% CI: [5.75, 6.05], would view humans' true selves as significantly more positive when compared against 5 (the scale's midpoint; see Figure 4A). However, a paired-samples *t*-test revealed that participants expected God ($M=5.90$, $SD=1.69$) to view humans' true selves more positively than Joan ($M=5.60$, $SD=1.07$); $t(493)=3.81$, $p<.001$, $d=0.17$, 95% CI_{diff}: [.15, .46].

Perceptions of agents' punishment motives. We analyzed participants' perceptions of punishment motives using a 2 (Agent: God vs. Joan) x 2 (Motive: retributivist vs. consequentialist) repeated measures ANOVA.[6] This analysis revealed main effects of Agent, $F(1,493)=22.29$, $p<.001$, $\eta_p^2=.04$, and Motive, $F(1,493)=19.28$, $p<.001$, $\eta_p^2=.04$. The Agent x Motive interaction did not reach significance, $F(1,493)=2.29$, $p=.131$.

Although the interaction was not significant, we conducted a follow-up test in accordance with the analytic plan outlined in Study 2's pre-registration. Specifically, we compared the degree to which participants attributed retributivist and consequentialist motives to God and to

[6] As in Study 1A, our pre-registration for Study 2 specified particular comparisons of theoretical interest rather than the omnibus ANOVA. However, to have a comprehensive understanding of the data, we first tested participants' views of each agent's punishment motives using a 2 (Motive: retributivist vs. consequentialist) x 2 (Agent Perspective: God vs. Joan) repeated measures ANOVA.

Joan (see Figure 4B). As in Study 1A, the pattern of results for Study 2 supported the *mixed account*: participants perceived God as less retributive than Joan, $t(493)=-3.39$, $p=.001$, $d=0.15$, 95% CI_{diff}: [-.56, -.15], but we did not observe a significant difference in their perceptions of consequentialist motives between God and Joan, $t(493)=-0.77$, $p=.442$, 95% CI_{diff}: [-.27, .12].

Additionally, as in Study 1A, we compared perceived retributivist versus consequentialist motives within each agent. Participants viewed both God, $t(493)=3.73$, $p<.001$, $d=0.17$, 95% CI_{diff}: [.25, .79] and Joan, $t(493)=2.16$, $p=.031$, $d=0.10$, 95% CI_{diff}: [.02, .46], as more motivated by consequentialist rather than retributivist concerns.

Figure 4

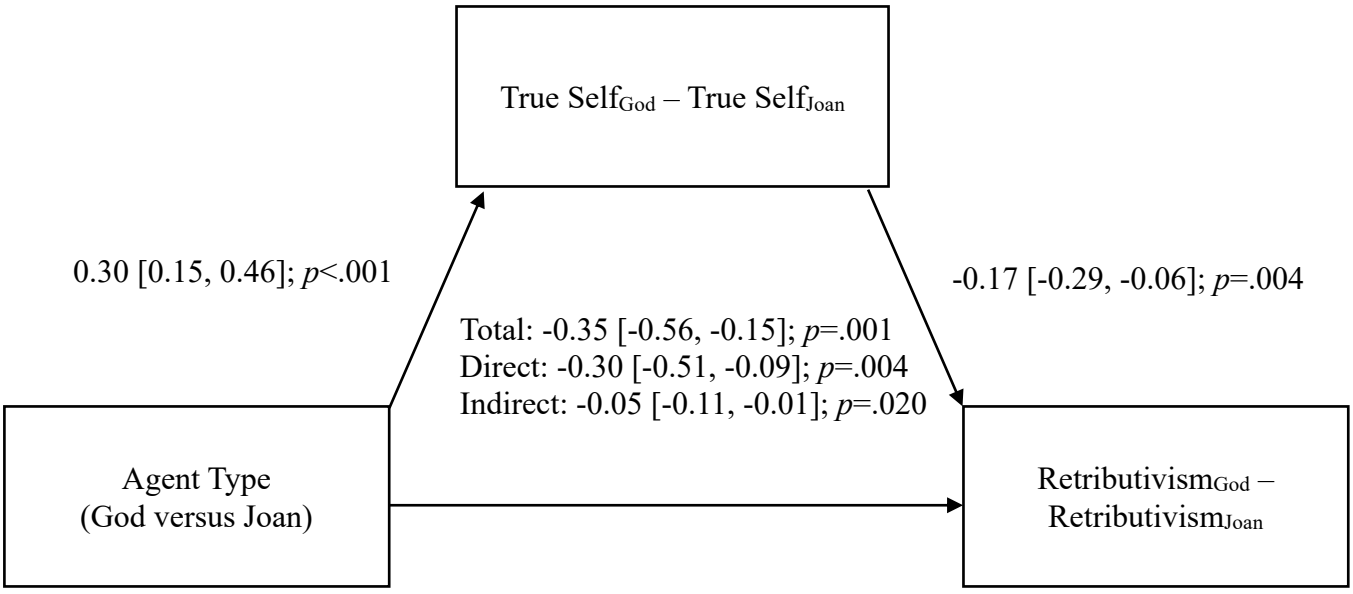
(A) Average Agreement with True Self Perceptions by Agent Type (B) Average Agreement with Different Punishment Motives by Agent Type in Study 2

Note. (A) Higher values indicate greater agreement that a given agent views humans' true selves as good. (B) Higher values indicate greater agreement that a given agent possesses a specific punishment motive. Error bars represent 95% confidence intervals.

Potential mechanism. Finally, we examined whether perceptions of agents' views of the true self mediated the difference in perceived retributivist motives across agents. As specified in our pre-registration, we used MEMORE (Montoya & Hayes, 2017) with 20,000 bootstrapped samples. We entered agent type (God versus Joan) as the independent variable, a difference in ratings between each agent's view of humans' true selves as the mediator, and a difference in ratings between each agent's perceived retributivism as the dependent variable. Because the perceived consequentialist motives between agents did not differ significantly, we did not run a mediation analysis on perceived consequentialism.

Figure 5

Estimates for the Relationship Between Agent Type and Retributivism as Mediated by Agent's View About Humans' True Selves



Note. Numbers in brackets indicate 95% confidence intervals. The mediator represents participants' agreement ratings about God's views of humans' true selves minus those of Joan's views about humans' true selves. The dependent variable represents participants' ratings about God's retributivist motives minus those of Joan's retributivist motives.

Perceptions of how each agent would perceive humans' true selves mediated the relation between agent type and perceived retributivist motives. The more participants viewed God as optimistic about humans' true selves as compared to Joan, the less they viewed God as retributive as compared to Joan (indirect effect: -0.05, p=.020, 95% CI [-.11, -.01]; Figure 5).

Discussion

Study 2 examined a potential mechanism underlying the different perceptions of retributivism between God and Joan. We found that participants' inferences about how God or Joan would view humans' true selves was associated with their differential perceptions of retributivism between God and Joan. The more participants inferred that God views humans' true selves positively compared to Joan, the less they perceived God to hold retributivist punishment motives compared to Joan. This finding dovetails with prior work suggesting that judgments of another individual's inherent goodness may be linked with reasoning about punishment (e.g., Heiphetz, 2020).

Study 3

One benefit of Study 2's design was that it allowed us to measure naturally-occurring variation in participants' attribution of true self beliefs to God and to another human. However, one drawback of Study 2's correlational approach was that it did not allow to determine how attributions regarding true self beliefs might causally impact perceived punishment motives. Therefore, in a final pre-registered study (https://aspredicted.org/blind.php?x=jz7f75), we manipulated agents' views of humans' true selves. Specifically, we manipulated Joan's views of humans' true selves and examined the extent to which such information impacted participants' judgments of Joan's retributivist motives. Separately, we manipulated God's views of humans' true selves and examined the extent to which such information impacted participants' judgments

of God's retributivist motives. This approach allowed us to draw causal inferences about the impact of true self beliefs on each agent's perceived punishment motives.

Method

Participants. Our final sample included 1,033 adults between 18 and 80 years old (M_{age}=39.01 years, SD_{age}=12.46 years; 67% female, 33% male, <1% other). A sensitivity analysis indicated that our sample size was large enough to detect small main and interaction effects (f=0.09) with sufficient power (0.80) at alpha .05. Originally, we recruited 920 participants to yield a sample consistent with the pre-registered sample size (n=714). However, an electronic error in randomization during data collection resulted in all participants being assigned to one of three conditions, with zero participants assigned to the fourth condition. After realizing this error, we recruited 350 more participants to the fourth condition. We analyzed Study 3's data after recruiting participants for all conditions.

Participants self-identified as White or European-American (79%), Black or African-American (9%), Asian or Asian-American (6%), Multiracial (4%), Native American or Pacific Islander (1%), other (1%). Additionally, 8% of adults self-identified as Hispanic or Latinx. Participants reported their present religious affiliation as Protestant (25%), Catholic (16%), other Christian (14%), Jewish (2%), Muslim (<1%), other/not listed (4%), and non-religious/atheist/agnostic (39%).

We recruited participants online via Amazon's Mechanical Turk, which we configured so that only United States residents whose approval rating was at least 95% could participate. Only those who did not participate in a similar study before were eligible for the study. Participants received $0.67 if they properly summarized a passage and correctly answered two attention check items asking them to recall questions they had answered earlier in the study; otherwise,

they received $0.05. In accordance with our pre-registration, we excluded data from 235 participants because they failed to meet at least one of these criteria.

Procedure. In Block I, participants read the vignette about one person murdering another from Study 1A. In Block II, participants completed one of four between-participants conditions in a 2 (God's Perception of True Self: good vs. bad) x 2 (Joan's Perception of True Self: good vs. bad) design. Participants read one of four possible paragraphs describing God's and, separately, Joan's views of the transgressor's true self. In one condition, participants learned that God viewed the transgressor's true self as morally bad, but that Joan viewed the transgressor's true self as morally good. In this condition, participants learned that God viewed the transgressor as "a fundamentally evil person" with "no compassion for other people and no concern at all about their well-being" but that Joan held the opposite view, perceiving the transgressor as "a fundamentally good person" and, despite his actions, as having "a profound compassion for other people and a genuine concern about their well-being" (descriptions adapted from Newman et al., 2015). The other three conditions proceeded analogously. After reading one of these paragraphs, participants completed an attention check asking them to indicate how each agent viewed the transgressor. In Block III, participants rated the extent to which they expected Joan and God to endorse retributive and consequentialist punishment motives.

Results

Across Studies 1A and 2, participants reliably attributed different retributivist—but not consequentialist—motives across agents. For brevity, we only report results pertinent to perceived retributivist motives below; see Supplementary Materials for results pertinent to perceived consequentialist motives.

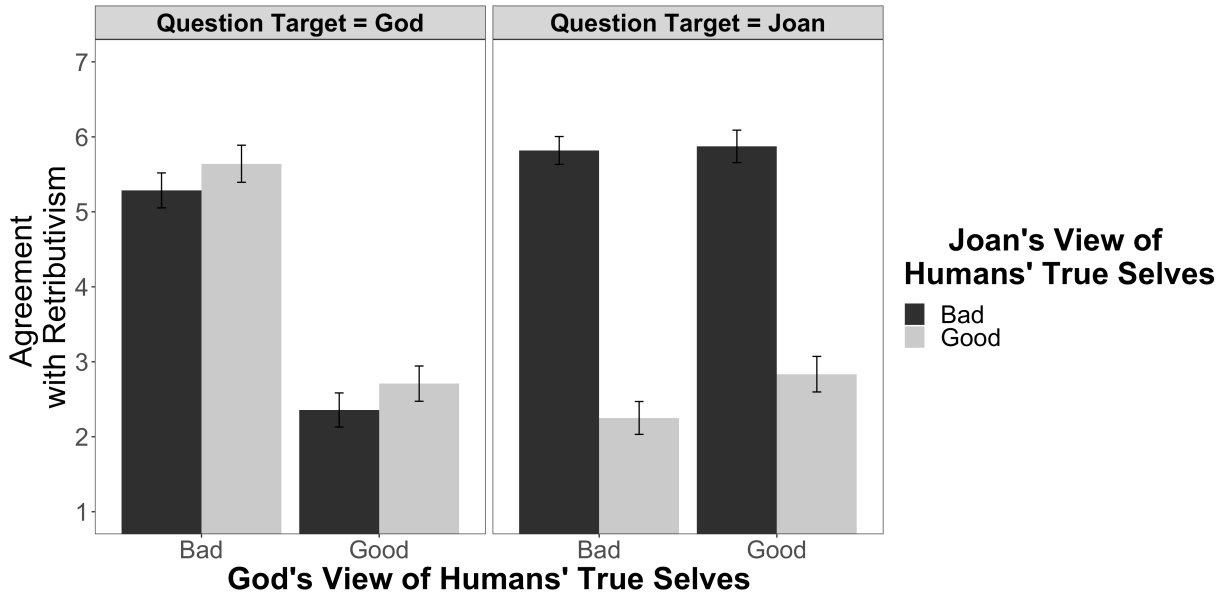
We examined the extent to which manipulating agents' views of the true self influenced perceptions of each agent's retributivist motives using a 2 (God's Perception of True Self: good vs. bad) x 2 (Joan's Perception of True Self: good vs. bad) x 2 (Target of Question: God vs. Joan) mixed ANOVA (Figure 6).[7] We observed main effects of God's Perception of True Self, $F(1,1029)=219.54$, $p<.001$, $\eta_p^2=.18$, Joan's Perception of True Self, $F(1,1029)=279.98$, $p<.001$, $\eta_p^2=.21$, and Target of Question, $F(1,1029)=7.18$, $p=.008$, $\eta_p^2=.007$. These main effects were qualified by a God's Perception of True Self x Target of Question interaction, $F(1,1029)=493.25$, $p<.001$, $\eta_p^2=.32$, and a Joan's Perception of True Self x Target of Question interaction, $F(1,1029)=624.40$, $p<.001$, $\eta_p^2=.38$. No other interactions reached significance, $ps>.068$.

To unpack the two-way interactions, we ran pairwise comparisons. First, when participants learned that God viewed humans' true selves as good (versus bad), they attributed lower retributivist motives to God, $p<.001$, $d=1.51$. However, attribution of retributivism to Joan did not differ significantly depending on how God viewed humans' true selves, $p=.128$, $d=0.09$. Second, when participants learned that Joan viewed humans' true selves as good (versus bad), they attributed lower retributivist motives to Joan, $p<.001$, $d=1.87$. In contrast, attribution of retributivism to God did not differ significantly depending on how Joan viewed humans' true selves, $p=.065$, $d=0.11$.

[7] Although we had pre-registered a 2 (God's Perception of True Self: good vs. bad) x 2 (Joan's Perception of True Self: good vs. bad) between-participants ANOVA, reviewers pointed out that this analytic approach had the issue of running multiple analyses. To minimize multiple testing, we therefore ran a 2 (God's Perception of True Self: good vs. bad) x 2 (Joan's Perception of True Self: good vs. bad) x 2 (Target of Question: God vs. Joan) mixed ANOVA with repeated measures on the last factor.

Figure 6

Average Agreement with Retributivism by Agent's View of Humans' True Selves and Target of Punishment Motive Question in Study 3



Note. "Question Target = God" indicates participants' responses when they attributed retributivist motives to God, while "Question Target = Joan" indicates responses when participants attributed retributivist motives to Joan. Higher values reflect greater agreement that a given agent endorsed retributivist motives. Error bars represent 95% confidence intervals.

Discussion

Study 3 manipulated agents' views of humans' true selves and examined how such information influenced perceived retributivism of each agent. Participants viewed a given agent as less retributive when that agent regarded the true self as good (versus bad). For example, participants expected Joan's motives to be less retributive when Joan believed that the true self was good rather than bad. Importantly, the manipulation of one agent's view of true self changed the attribution of retributivism to the agent specifically, but not that of the other agent. The results of Study 3 move beyond the correlational evidence from Study 2 by showing that beliefs

about the "good true self" causally lower perceptions of an agent's retributivist punishment motives.

General Discussion

We examined laypeople's inferences regarding the extent to which different agents (Joan, God) punish for retributivist versus consequentialist reasons. Several main findings emerged. In Study 1A, participants viewed God as less retributive than Joan. Studies 1B-1C demonstrated that this effect generalizes across distinct religious samples and different transgression domains. Study 2 offered insight regarding a potential underlying mechanism. Namely, participants expected God to view humans' true selves more positively than Joan; this difference predicted participants' perceptions of God as less retributive than Joan. Study 3 demonstrated that manipulating views of humans' true selves impacted participants' judgments of agents' retributivist motives: participants expected agents who were pessimistic (versus optimistic) about humans' true selves to be especially retributive.

Contributions of the Current Research

The current work extended the scientific understanding of moral reasoning in three main ways. First, the present studies extended scholarship on people's judgments concerning punishment. Prior work suggests that people consistently report that individuals who transgress should receive punishment (e.g., Dunlea & Heiphetz, 2021; Henrich et al., 2010; Marshall et al., 2021). In addition to the belief that other *humans* should punish those who act antisocially, people readily endorse the idea that *God* can punish transgressions (e.g., Laurin et al., 2012). While converging lines of evidence suggest that people endorse the idea that both earthly and divine agents can enact punishment, less work has focused on people's judgments about *why*

such agents punish. By addressing this question, the current work extended scholarship on punishment by charting lay theories of others' punishment motives.

Second, the current work increased scientific understanding of how people view the true self. In Study 2, participants expected God (versus Joan) to be more optimistic about humans' true selves. This finding extended past work showing that people themselves think the true self is good (e.g., De Freitas et al., 2018; Newman et al., 2014) by demonstrating the beliefs people hold about other agents' views of the true self. This finding may also have implications for reducing retribution and other forms of interpersonal conflict. Whereas conflict-reduction interventions often require elaborate methods (e.g., Paluck, 2009), our findings hint that inducing positive perspectives about others' true selves might potentially reduce retribution. This knowledge might be particularly useful in a religious intergroup context. For instance, a belief that God views humans as morally good deep inside might mitigate retributivism against people from another religious out-groups.

Finally, the current work illuminated the degree to which people attribute similar morally-relevant mental states across earthly and divine minds. This comparison was possible because we measured how participants perceive the punishment motives of Joan—another human they do not know—rather than asking participants' own motives directly. If we had assessed participants' own punishment motives, it would have been difficult to disentangle the extent to which their responses are due to anthropomorphism—attribution of similar characteristics to God and humans—versus egocentric thinking (e.g., self-positivity bias). By measuring participants' perceptions of Joan's motives, the current work contributed to broader conversations regarding the extent to which people view God's mind as broadly humanlike (Ginges et al., 2016; Heiphetz et al., 2018; Richert et al., 2016).

Past work led to three possibilities regarding the extent to which participants would view God's mind and a human mind as similar to or different from each other. One possibility—the *similarity account*—was that people may view God and humans as possessing similar retributivist and consequentialist motives. This possibility stems from scholarship suggesting that, under some conditions, people view God's mind as humanlike (e.g., Heiphetz et al., 2018; Knight et al., 2004; Shtulman & Lindeman, 2016). A second possibility—the *divergence account*—was that people may view God and humans as possessing different motives. This possibility stemmed from prior work suggesting that adults sometimes provide "theologically correct" answers when asked to reason about God's mind (i.e., that God's mind is strikingly different from humans' minds; Barrett, 1999). Finally, a third possibility—the *mixed account*—was that people may report both similarities and differences between humans' and God's punishment motives. People's reasoning about complex topics (e.g., religion) often reflects co-existing concepts (Legare et al., 2012; Shtulman & Lombrozo, 2016). In line with many "theologically correct" perspectives conveyed via testimony (i.e., those suggesting that human and divine minds are dissimilar; Armstrong, 1993), it was possible that people would report that some punishment motives differ across agents. Yet, in line with prior work detailing people's lay theories of God's mind (e.g., Ginges et al., 2016; Pasek et al., 2020), it was possible that people could concurrently report that some punishment motives are similar across agents' minds.

The current work supported the *mixed account*. Across Studies 1A-1C and Study 2, participants indicated that, compared to another human, God was less motivated by retribution. However, in Studies 1A and 2, a difference in participants' attributions of consequentialist motives between agents did not emerge. One potential reason for this difference in attributions of retributivist versus consequentialist motives relates to the idea that people's inferences about

God's attributes may not strongly overlap with their inferences about the attributes of retributive individuals. People often view God as possessing more pro-social tendencies than other humans (e.g., Ginges et al., 2016; Pasek et al., 2020). Additionally, people sometimes perceive retributivist—but not consequentialist—motives as signaling anti-social desires (Dhaliwal et al., 2022; Herrmann et al., 2008). It is possible that people have been more reticent to attribute retributivist motives to God versus other humans because people are likely to view God as possessing *pro*-social tendencies (e.g., Pasek et al., 2020) and retributivist motives as signaling *anti*-social tendencies (e.g., Dhaliwal et al., 2022), although causality among these factors remains unclear.

Directions for Future Research

The current work examined people's inferences about why different agents (another human, God) are motivated to punish. In doing so, the present studies made important theoretical contributions to literatures spanning moral and religious cognition. However, as in all programs of research, additional lines of inquiry remain open for future investigation.

One potentially fruitful direction for future work includes examining the generalizability of the current findings. The present work probed whether perceptions of punitive motives generalize across different religious groups and different transgression domains. For instance, Study 1B showed that Catholic and Jewish participants with different views of God nevertheless were less likely to attribute retributivist motive to God than to Joan. However, the finding from their views of God indicates that both religious groups had a tendency to view God positively, even though Catholics' views were more positive than Jews' views. Hence, future work can investigate whether our findings generalize to religions with more punitive concepts of God, such as Evangelical Protestantism (e.g., Pew Research Center, 2014).

Also, to examine God's perceived punishment motives, in this study we needed to tell participants that punishment occurred. However, not everyone might endorse the concept of God's punishment. For example, although prior work has shown that religious believers and non-believers report similar theological concepts in some circumstances (Heiphetz et al., 2018; Kelemen, 2004), non-believers may differ from participants who affiliate with a religious group in their views of God's punishment motives. Additionally, we collected the data for all studies reported here in the United States, which is a predominantly Christian culture. Therefore, Christian notions of God might have affected the current results. It would be important for future research to investigate how members of different cultural and religious groups reason about God's punishment motives.

Future research can also investigate how inferences about an agent's punishment motives shapes responses to transgression. Because retributivism focuses on causing the transgressor suffering, retributivist punishment spotlights transgressors while leaving victims "in the dark." Following this conceptualization, people who perceive a punitive agent as primarily motived by retribution may respond to transgression in ways that center transgressors as opposed to restoring the victim. For instance, people may respond to transgression in more self-centered ways when they are primed to think that another human (versus God) knows about their transgression. Future work can explore these possibilities.

Conclusion

We examined laypeople's inferences regarding why different agents (another human, God) punish. Three central findings emerged. First, participants perceived God as less retributive than another human—a finding that emerged across members of different religious groups and generalized across transgressions framed in both religious and secular ways. Second, participants

expected God to view humans' true selves more positively than another human would; this difference predicted the difference in perceived retributivist motives. Third, participants expected agents to endorse retributivism less when viewing the true self as good versus bad. Taken together, these results suggest that perceived retributivist punishment motives hinge on views of humans' true selves. Moreover, these findings extend prior theorizing by charting people's lay theories about others' punishment motives and highlighting links between religious and moral cognition.

References

Armstrong, K. (1993). *A history of God: The 4,000-year quest of Judaism, Christianity, and Islam*. Random House Publishers.

Balliet, D., Mulder, L. B., & Van Lange, P. A. M. (2011). Reward, punishment, and cooperation: A meta-analysis. *Psychological Bulletin*, *137*, 594–615. https://doi.org/10.1037/a0023489

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182. doi: 10.1037//0022-3514.51.6.1173

Barrett, J. L. (1999). Theological correctness: Cognitive constraint and the study of religion. *Method & Theory in the Study of Religion*, *11*, 325–339. doi: 10.1163/157006899X00078

Bench, S. W., Schlegel, R. J., Davis, W. E., & Vess, M. (2015). Thinking about change in the self and others: The role of self-discovery metaphors and the true self. *Social Cognition*, *33*, 169-185. doi: 10.1521/soco.2015.33.3.2

Bentham, J. (1823/1970). An introduction to the principles of morals and legislation. Althone Press.

Bering, J., & Johnson, D. (2005). "O Lord… You Perceive my Thoughts from Afar": Recursiveness and the Evolution of Supernatural Agency. *Journal of Cognition and Culture*, *5*, 118-142. doi: 10.1163/1568537054068679

Carlsmith, K. M. (2008). On justifying punishment: The discrepancy between words and actions. *Social Justice Research*, *21*, 119-137. doi: 10.1007/s11211-008-0068-x

Carlsmith, K. M., & Sood, A. M. (2009). The fine line between interrogation and retribution. *Journal of Experimental Social Psychology*, *45*, 191-196. doi: 10.1016/j.jesp.2008.08.025

Darley, J. M. (2009). Morality in the law: The psychological foundations of citizens' desires to punish transgressions. *Annual Review of Law and Social Science*, *5*, 1-23. doi: 10.1146/annurev.lawsocsci.4.110707.172335

De Freitas, J., & Cikara, M. (2018). Deep down my enemy is good: Thinking about the true self reduces intergroup bias. *Journal of Experimental Social Psychology*, *74*, 307-316. doi: 10.1016/j.jesp.2017.10.006

De Freitas, J., Sarkissian, H., Newman, G. E., Grossmann, I., De Brigard, F., Luco, A., & Knobe, J. (2018). Consistent belief in a good true self in misanthropes and three interdependent cultures. *Cognitive Science*, *42*, 134-160. doi: 10.1111/cogs.12505

Demoulin, S., Saroglou, V., & Van Pachterbeke, M. (2008). Infra–humanizing others, supra–humanizing gods: The emotional hierarchy. *Social Cognition*, *26*, 235-247. doi: 10.1521/soco.2008.26.2.235

Dhaliwal, N. A., Skarlicki, D. P., Hoegg, J., & Daniels, M. A. (2022). Consequentialist motives for punishment signal trustworthiness. *Journal of Business Ethics*, 176, 451-466. doi: 10.1007/s10551-020-04664-5

Dunlea, J. P. & Heiphetz, L. (2021). Moral psychology as a necessary bridge between social cognition and law. *Social Cognition*, *39*, 183-199. doi: 10.1521/soco.2021.39.1.183

Dunlea, J. P., & Heiphetz, L. (2022). Language shapes children's attitudes: Consequences of internal, behavioral, and societal information in punitive and non-punitive contexts. *Journal of Experimental Psychology: General*.

Epley, N., Converse, B. A., Delbosc, A., Monteleone, G. A., & Cacioppo, J. T. (2009). Believers' estimates of God's beliefs are more egocentric than estimates of other people's beliefs. *Proceedings of the National Academy of Sciences*, *106*, 21533-21538. doi: 10.1073/pnas.0908374106

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191. doi: 10.3758/BF03193146

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*, 785-791. doi: 10.1038/nature02043

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, *18*, 233-239. doi: 10.1111/j.1467-9280.2007.01882.x

Funk, F., McGeer, V., & Gollwitzer, M. (2014). Get the message: Punishment is satisfying if the transgressor responds to its communicative intent. *Personality and Social Psychology Bulletin*, *40*, 986-997. doi: 10.1177/0146167214533130

Ginges, J., Sheikh, H., Atran, S., & Argo, N. (2016). Thinking from God's perspective decreases biased valuation of the life of a nonbeliever. *Proceedings of the National Academy of Sciences*, *113*, 316-319. doi: 10.1073/pnas.1512120113

Haslam, N., Kashima, Y., Loughnan, S., Shi, J., & Suitner, C. (2008). Subhuman, inhuman, and superhuman: Contrasting humans with nonhumans in three cultures. *Social Cognition*, *26*, 248-258. doi: 10.1521/soco.2008.26.2.248

Heiphetz, L. (2020). The development and consequences of moral essentialism. *Advances in Child Development and Behavior*, *59*, 165-194. doi: 10.1016/bs.acdb.2020.05.006

Heiphetz, L., Lane, J. D., Waytz, A., & Young, L. L. (2018). My mind, your mind, and God's mind: How children and adults conceive of different agents' moral beliefs. *British Journal of Developmental Psychology*, *36*, 467-481. doi: 10.1111/bjdp.12231

Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., ... & Ziker, J. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science*, *327*, 1480-1484. doi: 10.1126/science.1182238

Hester, N., & Gray, K. (2020). The moral psychology of raceless, genderless strangers. *Perspectives on Psychological Science*, *15*, 216-230. doi: 10.1177/1745691619885840

Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science*, *319*, 1362-1367. doi: 10.1126/science.1153808

Hume, D. (1739/1888). *A treatise of human nature*. Oxford, UK: Oxford University Press.

IBM Corporation (2021). IBM SPSS Statistics for macOS, Version 28.0.

Johnson, D., & Bering, J. (2006). Hand of God, mind of man: Punishment and cognition in the evolution of cooperation. *Evolutionary Psychology*, *4*, 219-233. doi: 10.1177/147470490600400119

Kelemen D. (2004). Are children "intuitive theists"? Reasoning about purpose and design in nature. *Psychological science*, *15*(5), 295–301. https://doi.org/10.1111/j.0956-7976.2004.00672.x

Knight, N., Sousa, P., Barrett, J. L., & Atran, S. (2004). Children's attributions of beliefs to humans and God: Cross-cultural evidence. *Cognitive Science*, *28*, 117-126. doi: 10.1207/s15516709cog2801_6

Laurin, K., Shariff, A. F., Henrich, J., & Kay, A. C. (2012). Outsourcing punishment to God: Beliefs in divine control reduce earthly punishment. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 3272-3281. doi: 10.1098/rspb.2012.0615

Legare, C. H., Evans, E. M., Rosengren, K. S., & Harris, P. L. (2012). The coexistence of natural and supernatural explanations across cultures and development. *Child Development*, *83*, 779-793. doi: 10.1111/j.1467-8624.2012.01743.x

Marshall, J., Yudkin, D. A., & Crockett, M. J. (2021). Children punish third parties to satisfy both consequentialist and retributivist motives. *Nature Human Behaviour*, *5*, 361-368. doi: 10.1038/s41562-020-00975-9

Marshall, J., Gollwitzer, A., & Bloom, P. (2022). Why do children and adults think other people punish? *Developmental Psychology*. https://doi.org/10.1037/dev0001378

Mathew, S., & Boyd, R. (2011). Punishment sustains large-scale cooperation in prestate warfare. *Proceedings of the National Academy of Sciences*, *108*, 11375-11380. doi: 10.1073/pnas.1105604108

Montoya, A. K., & Hayes, A. F. (2017). Two-condition within-participant statistical mediation analysis: A path-analytic framework. *Psychological Methods, 22*, 6-27. doi:10.1037/met0000086

Nadler, J., & McDonnell, M. H. (2011). Moral character, motive, and the psychology of blame. *Cornell Law Review*, *97*, 255-304.

Newman, G. E., Bloom, P., & Knobe, J. (2014). Value judgments and the true self. *Personality and Social Psychology Bulletin*, *40*, 203-216. doi: 10.1177/0146167213508791

Newman, G. E., De Freitas, J., & Knobe, J. (2015). Beliefs about the true self explain asymmetries based on moral judgment. *Cognitive Science*, *39*, 96-125. doi: 10.1111/cogs.12134

Nussbaum, M. C. (2016). *Anger and forgiveness: Resentment, generosity, justice*. Oxford University Press.

Paluck, E. L. (2009). Reducing intergroup prejudice and conflict using the media: A field experiment in Rwanda. *Journal of Personality and Social Psychology*, *96*, 574–587. doi: 10.1037/a0011989.

Pasek, M. H., Shackelford, C., Smith., J., Vishkin, A., Lehner, A, & Ginges, J. (2020). God values the lives of my outgroup more than I do: Evidence from Fiji and Israel. *Social Psychological and Personality Science*. doi: 10.1177/1948550620904516.

Pew Research Center. (2014). Belief in hell. Retrieved on June 30th, 2022, from https://www.pewresearch.org/religion/religious-landscape-study/belief-in-hell/

Pew Research Center. (2015). America's changing religious landscape. Retrieved on September 9th, 2020, from https://www.pewforum.org/2015/05/12/americas-changing-religious-landscape/

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Richert, R.A., Shaman, N.J., Saide, A.R., & Lesage, K.A. (2016). Folding your hands helps God hear you: Prayer and anthropomorphism in parents and children. *Research in the Social Scientific Study of Religion*, *27*, 140-157. doi: 10.1163/9789004322035_010

Shariff, A. F., & Norenzayan, A. (2011). Mean gods make good people: Different views of god predict cheating behavior. *International Journal for the Psychology of Religion*, *21*(2), 85–96. https://doi.org/10.1080/10508619.2011.556990

Shariff, A. F., Greene, J. D., Karremans, J. C., Luguri, J. B., Clark, C. J., Schooler, J. W., Baumeister, F., & Vohs, K. D. (2014). Free will and punishment: A mechanistic view of human nature reduces retribution. *Psychological Science*, *25*, 1563-1570. doi: 10.1177/0956797614534693

Shariff, A. F., Willard, A. K., Andersen, T., & Norenzayan, A. (2016). Religious priming: A meta-analysis with a focus on prosociality. *Personality and Social Psychology Review*, *20*(1), 27–48. https://doi.org/10.1177/1088868314568811

Shtulman, A., & Lindeman, M. (2016). Attributes of God: Conceptual foundations of a foundational belief. *Cognitive Science*, *40*, 635-670. doi: 10.1111/cogs.12253

Shtulman, A., & Lombrozo, M. (2016). Bundles of contradiction: A coexistence view of conceptual change. In D. Barner & A. Baron (Eds.). Core knowledge and conceptual change (pp. 49-67). Oxford University Press.

Waytz, A., Young, L. L., & Ginges, J. (2014). Motive attribution asymmetry for love vs. hate drives intractable conflict. *Proceedings of the National Academy of Sciences*, *111*, 15687-15692. doi: 10.1073/pnas.1414146111

Zhang, Y., & Alicke, M. (2021). My true self is better than yours: Comparative bias in true self judgments. *Personality and Social Psychology Bulletin*, *47*(2), 216-231.