How Reputation Does (and Does Not) Drive People to Punish Without Looking

Jillian J. Jordan
Nour S. Kteily
How Reputation Does (and Does Not) Drive People to Punish Without Looking

Jillian J. Jordan
Harvard Business School

Nour S. Kteily
Northwestern University

Working Paper 22-073
How reputation does (and does not) drive people to punish without looking

Jillian J. Jordan¹*, Nour S. Kteily²

¹ Harvard Business School, Soldiers Field, Boston, MA 02163
² Kellogg School of Management, Northwestern University,
   2211 Campus Dr, Evanston, IL 60208

*Corresponding author: Jillian J. Jordan
Email: jijordan@hbs.edu
Phone number: (651) 895-3771
Mailing address: Soldiers Field, Boston, MA 02163

Classification: Social Sciences/Psychological and Cognitive Sciences

Keywords: Opposing perspectives; Outrage culture; Signaling; Ideology; Moralistic punishment

OSF page: https://osf.io/3es2k/?view_only=e272c31aa7304eca841f084d582187fb
Abstract

Punishing wrongdoers can confer reputational benefits, and people sometimes punish without careful consideration. But are these observations related? Does reputation drive people to people to “punish without looking”? And if so, is this because unquestioning punishment looks particularly virtuous? To investigate, we assigned “Actors” to decide whether to sign punitive petitions about politicized issues (“punishment”), after first deciding whether to read articles opposing these petitions (“looking”). To manipulate reputation, we matched Actors with co-partisan “Evaluators”, varying whether Evaluators observed (i) nothing about Actors’ behavior, (ii) whether Actors punished, or (iii) whether Actors punished and whether they looked. Across four studies of Americans (total $n = 10,343$), Evaluators rated Actors more positively, and financially rewarded them, if they chose to (vs. not to) punish. Correspondingly, making punishment observable to Evaluators (i.e., moving from our first to second condition) drove Actors to punish more overall. Furthermore, because some of these individuals did not look, making punishment observable increased rates of punishment without looking. Yet punishers who eschewed opposing perspectives did not appear particularly virtuous. In fact, Evaluators preferred Actors who punished with (vs. without) looking. Correspondingly, making looking observable (i.e., moving from our second to third condition) drove Actors to look more overall—and to punish without looking at comparable or diminished rates. We thus find that reputation can encourage reflexive punishment—but simply as a byproduct of generally encouraging punishment, and not as a specific reputational strategy. Indeed, rather than fueling unquestioning decisions, spotlighting punishers’ decision-making processes may encourage reflection.
**Introduction**

Consider the following narrative: after an accusation of wrongdoing, an “outrage mob” appears. Initially, just a few condemn the accused. But before long, more and more people “pile on,” hastily meting out punishment with little regard for evidence or opposing perspectives. According to some critics of “outrage culture,” such events have become troublingly frequent. A growing number of voices have expressed concern that society has become too quick to enact punishment without due consideration, yielding outcomes that can be disproportionate, unfair, and sometimes downright cruel (1, 2).

Even highly-educated professors can, by their own admission, punish without sufficient reflection. In 2022, 38 Harvard faculty members signed an open letter condemning the University for sanctioning John Comaroff, an Anthropology Professor who was accused of sexual harassment. Their letter expressed “dismay” at the sanctions, endorsing Comaroff’s narrative (3). After more details emerged, however, 35 of the letter’s 38 signatories retracted their signatures, acknowledging that they were “lacking full information about the case”. Following the retraction, critics alleged that “the letter’s uncritical engagement with limited perspectives, voiced in such decisive and unambiguous terms, was hasty and uninformed” (4).

Here, we ask what drives people to punish without careful consideration, which we term “punishment without looking.” There are many ways that somebody might “look” (i.e., gather information to assess the merits of punishment) before deciding whether to punish: one might research basic facts, attempt to verify claims, question relevant sources, or consider additional perspectives—including those that support, or oppose, punishment. We conceptualize “punishment without looking” as the decision to punish *without* engaging in relevant looking behaviors. In particular, we focus on the form of punishment without looking that the 38 Harvard faculty members were accused of: punishment without consideration of opposing perspectives. One might see a trend towards punishment with less consideration of opposing perspectives as negative or positive for society, depending on the context and one’s orientation towards the relevant moral issue. When merited, punishment is essential for discouraging wrongdoing and holding perpetrators accountable (5, 6). Consider how movements like #MeToo and #BlackLivesMatter have punished transgressions, like sexual assault and racism, that many believe have historically been committed with impunity. From this perspective, it is problematic when people are too hesitant to support justified punitive efforts. Regardless of whether one believes that people should push more or less reflexively, however, punishment without looking clearly has meaningful social consequences—raising the question of what drives such behavior.

**Does reputation drive people to punish without looking?**

We consider whether *reputation motives* (7–12) encourage punishment without looking. This hypothesis is partially inspired by social commentary suggesting that “virtue signaling” can inspire punishment that is hasty, underserved, or goes too far (13–15). For example, Sunstein (15) considers “what happens when a group of people, outraged by a real or imagined transgression, responds in a way that is disproportionate.” He suggests that such responding represents a “quick, automatic reaction”, fueled by reputation (“people want to appear at least as appalled as others in their social group.”) Similarly, Haidt and Rose-Stockwell (14) discuss “what happens when people use moral talk to enhance their prestige in a public forum”, arguing that “nuance and truth are casualties in this competition to gain the approval of the audience.” Such perspectives raise the question of whether reputation contributes to reflexive punishment.
Furthermore, academic research highlights the power of reputation to fuel moralistic punishment in general. Punishment can be socially rewarded (16) and signal trustworthiness (17–22) by conveying that the punisher is unlikely to themself transgress. And consequently, people punish more when their decisions are observable to others (23, 24), and when punishment has greater signaling value (17, 18).

Yet while reputation clearly drives punishment in general, and social commentary suggests that virtue signaling may fuel reflexive punishment, the hypothesis that reputation drives people to punish specifically without looking remains untested. We investigate two distinct pathways through which reputation might have this effect.

First, reputation could drive punishment without looking as a byproduct of driving punishment in general. When reputation motivates people to punish, some individuals may choose not to look (e.g., to save time or effort, or avoid engaging with disliked sources). Thus, by generally encouraging punishment, reputation may give rise to some reflexive punishment.

Second, reputation could drive punishment without looking as a specific reputational strategy, if people appear especially virtuous when they punish unquestioningly. People who punish without considering opposing perspectives might benefit from appearing particularly committed to the relevant moral cause—especially in politicized domains, where people are increasingly intolerant of other viewpoints (25). Furthermore, theoretical modeling highlights that cooperating “without looking” (i.e., without attending to the costs of cooperation) can be preferentially rewarded (26). Indeed, cooperative decisions that are faster, more intuitive, and less calculating tend to be evaluated more positively (9, 27–29), and people make less calculating cooperative decisions when others can observe their decision-making process (27). Thus, cooperating without looking can serve as a specific reputational strategy, and punishment without looking might function similarly.

Yet there are also reasons to doubt this proposal. Punishment—unlike cooperation—is morally bad when undeserved, so punishers who eschew opposing perspectives might incur the reputational cost of seeming less fair-minded. Furthermore, deliberative decisions are often extolled as wiser (30–32), so punishment without looking might signal reduced competence.

Across four pre-registered studies of Americans from Mturk and Prolific (total \( n = 10,343 \)), we show that reputation can drive people to punish without looking. Critically, however, we find that reputation fuels such punishment as a byproduct of incentivizing punishment in general, and not as a specific reputation strategy. While punishers were reputationally rewarded in our studies, punishers earned the best reputations by considering opposing perspectives before punishing. Thus, placing reputational scrutiny on punishment in general increased punishment without looking. But spotlighting potential punishers’ decision-making processes actually encouraged engagement with other perspectives—and did not influence or even decreased punishment without looking.

We also consider how the influence of reputation may depend on the reputational audience, and their ideological conviction towards the relevant moral issue. People may frequently face audiences who share their ideological viewpoints (i.e., “homophily”) (33, 34), and in particular hold extreme aligned views (i.e., “acrophily”) (35). Interestingly, in our studies, even ideologically extreme audiences did not preferentially reward punishers who declined (vs. chose) to look. Yet we find that evidence that such audiences may nonetheless preferentially fuel punishment without looking—both by doing do more to encourage punishment in general, and less to encourage consideration of other viewpoints.
Paradigm Overview

To investigate punishment without looking, we designed an incentive-compatible paradigm with meaningful stakes. We invited Actor subjects to sign real punitive petitions about politicized moral issues from the website Change.org (“punishment”), with or without first reading articles expressing perspectives opposing the petitions (“looking”). And we invited Evaluator subjects to respond to Actors’ decisions, while allocating financial resources to them.

We assigned Democrat vs. Republican subjects to engage with distinct petitions, that we expected to resonate with their respective political parties. Democrats read one of two petitions (because the first closed before we finished conducting all studies in this paper), calling for the firing of either (i) University of Central Florida professor Charles Negy, in light of “abhorrent racist comments…on his personal Twitter account,” or (ii) Los Angeles Police Department Chief Michel Moore, following Moore’s statement that the death of George Floyd is on protestors’ and looters’ hands. These petitions align with the #BlackLivesMatter movement, which is preferentially supported by liberal Americans. Republicans read a petition calling for the removal of “Blue Lives Murder” merchandise from Amazon, which the petition characterized as “hatred merchandise”. This petition aligns with the #BlueLivesMatter countermovement, which is preferentially supported by conservative Americans.

Before Actors decided whether to sign the petitions, they could consider opposing perspectives. In particular, Actors could read real news articles with headlines suggesting that punishment might be unmerited (e.g., “UCF professor behind tweets deemed racist says he is subject of ‘witch hunt’” for the Negy petition; “Despite criticism, LAPD Chief Michel Moore maintains support in political circles” for the Moore petition; “Spring woman says ‘Blue Lives Murder’ shirts for sale on Amazon are hate speech. Activist disagrees” for the Amazon petition).

Together, this paradigm allowed us to track whether Actors looked (i.e., considered opposing perspective articles) and punished (i.e., signed their assigned petition), and thus to identify individuals who punished without looking. To incorporate reputation, we used a “Dictator Game”: Evaluators received an endowment of 50¢ and decided how much to share with an Actor of their same political affiliation, on the basis of the Actor’s study decisions. Of note, stakes of this magnitude in online economic game studies tend to produce comparable results to in-person studies with stakes ten times larger (36–38).

In Studies 1-2, we used this paradigm to ask how Evaluators create reputational incentives for Actors, when rewarding their behavior. In these studies, Evaluators decided how much money to share, if given different information (or no information) about an Actor’s behavior (the “strategy method” (39)). They also rated Actors, using 0-100 scales, on overall positivity of impression, fairness, competence, and loyalty to the relevant moral cause. In Studies 3-4, we turned to asking how reputational incentives shape the behavior of Actors. In these studies, we compared Actors’ punishment and looking behavior across conditions in which Evaluators could observe (i) nothing about their decisions (“Nothing Observable”), (ii) just whether they punished (“Punishment Observable”), or (iii) whether they punished and whether they looked (“Both Observable”).

Results

Study 1

Study 1 investigated Evaluators. In Study 1a, we assigned $n = 629$ Democrat Evaluators to Actors who had either considered the Negy ($n = 308$) or Moore ($n = 321$) petition. Both groups showed equivalent results (see SI Section 2.1), so our analyses collapse across petition. In
Study 1b, we assigned \( n = 600 \) Republican Evaluators to Actors who had considered the Amazon petition.

**Did Evaluators create reputational incentives for punishment in general?**

We first seek to confirm: did Evaluators create reputational incentives for punishment *in general*, as must be true for even the “byproduct” hypothesis to hold? We thus compare Evaluators’ reactions to Actors who did vs. did not punish, when given no information about whether the Actor looked (Figure 1A).

In Study 1a, Democrats formed more positive overall impressions of Actors who *did* (vs. did not) punish (\( b = 24.08 \ [21.55, 26.61], t = 18.69, p < .001 \), shared a larger percentage of their endowment with punishers (\( b = 15.80 \ [13.79, 17.81], t = 15.44, p < .001 \), and rated punishers as more loyal supporters of #BlackLivesMatter (\( b = 36.82 \ [34.59, 39.05], t = 32.47, p < .001 \), more fair (\( b = 19.23 \ [17.06, 21.41], t = 17.38, p < .001 \), and more competent (\( b = 17.42 \ [15.37, 19.46], t = 16.71, p < .001 \), \( n = 629 \).

Similarly, Republicans in Study 1b rated punishers more positively overall (\( b = 22.10 \ [18.77, 25.42], t = 13.05, p < .001 \), shared more money with punishers (\( b = 15.32 \ [12.81, 17.82], t = 12.02, p < .001 \), and rated punishers as more loyal supporters of #BlueLivesMatter (\( b = 26.91 \ [23.49, 30.34], t = 15.45, p < .001 \), more fair (\( b = 19.86 \ [17.15, 22.58], t = 14.37, p < .001 \), and more competent (\( b = 18.02 \ [15.45, 20.58], t = 13.80, p < .001 \), \( n = 600 \).

Thus, Evaluators rewarded punishers, creating incentives for Actors to punish in general. We also find that Evaluators who identified as “strong” (vs. “weak”) partisans showed relatively stronger preferences for punishers (vs. non-punishers) (Figure 1A and Table 1), suggesting that more ideological audiences create especially strong incentives for punishment.

**Did Evaluators create reputational incentives to punish specifically without looking?**

Next, we ask: did Evaluators create reputational incentives for punishment *specifically without looking*, as the “specific reputational strategy” hypothesis predicts? We thus compare Evaluators’ reactions to Actors who punished without vs. with looking (Figure 1B).

In Study 1a, Democrats formed less positive impressions of punishers who declined (vs. chose) to look (\( b = -10.46 \ [-12.32, -8.60], t = -11.07, p < .001 \), and shared less money with them (\( b = -4.39 \ [-5.83, -2.94], t = -5.96, p < .001 \). They also rated punishers who declined to look as less fair (\( b = -16.30 \ [-18.35, -14.25], t = -15.61, p < .001 \) and less competent (\( b = -11.10 \ [-12.81, -9.39], t = -12.73, p < .001 \), but as more loyal (\( b = 5.55 \ [3.95, 7.14], t = 6.82, p < .001 \).

Similarly, Republicans in Study 1b rated punishers who declined to look less positively (\( b = -4.55 \ [-6.30, -2.79], t = -5.08, p < .001 \), shared less money with them (\( b = -1.72 \ [-3.14, -0.29], t = -2.36, p = .018 \) and saw them as less fair (\( b = -8.62 \ [-10.38, -6.86], t = -9.62, p < .001 \) and less competent (\( b = -5.20 \ [-6.87, -3.52], t = -6.10, p < .001 \), but also as more loyal (\( b = 6.90 \ [4.91, 8.88], t = 6.83, p < .001 \).

Thus, Evaluators *did* see punishing without looking as an especially strong signal of loyalty. Yet eschewing opposing perspectives also had reputational *costs*: it made punishers seem less fair and less competent. And the reputational costs outweighed the benefits, such that Evaluators rated punishers who declined to look less positively overall and sent them less money. Indeed, non-preregistered mediation analyses reveal (i) positive indirect effects of declining to look on overall positivity via enhanced ratings of loyalty, and (ii) countervailing (and significantly larger) *negative* indirect effects via diminished ratings of fairness and competence (see SI Section 2.3).
Our design also allows for between-subjects analyses of responses to punishers who declined vs. chose to look (because Evaluators encountered these two Actor profiles first, in random order). These analyses were also pre-registered, and produce results that are similar to our within-subject analyses, but, in line with their reduced statistical power, less consistently significant (see SI Section 2.4).

We also again investigate the role of Evaluator ideology (Figure 1B and Table 1). We find that strong (vs. weak) partisans reacted relatively less negatively to punishment without (vs. with) looking. Still, even strong partisans did not prefer punishment without looking. Strong Democrats rated punishers significantly less positively if they declined (vs. chose) to look, but shared comparable amounts of money with both types of punishers; strong Republicans did not significantly differentiate between punishers who declined vs. chose to look on either global evaluation measure.

Finally, Study 1 reveals that Evaluators reacted positively to looking even when the Actor ultimately declined to punish; see SI Section 2.5 for details.

Figure 1. Evaluators react positively to punishment, but do not prefer punishment without looking. Shown are evaluations of Actors who (A) did vs. did not punish, and (B) punished with vs. without looking, among weak partisan Evaluators, strong partisan Evaluators, and all Evaluators. Bars plot mean positivity ratings, and error bars are 95% CIs. For a version of Figure 1 that plots money shared, see SI Section 2.2.
Table 1. Strong partisan Evaluators react more positively to punishment, and less negatively to punishment without looking. In the left columns, for both positivity and money shared, we predict evaluations of Actors who did vs. did not punish. We report simple effects of punishment among both weak and strong partisan Evaluators, and the interaction between partisanship and punishment. In the right columns, we predict evaluations of Actors who punished without vs. with looking; we report simple effects of not looking, and the interaction between partisanship and not looking.

**Study 2**

In Study 1, Evaluators created reputational incentives for punishment in general, highlighting the potential for reputation to drive punishment without looking as a *byproduct*. However, Evaluators in Study 1 did not preferentially reward punishment without (vs. with) looking, casting doubt on the hypothesis that reputation drives punishment without looking as a *specific reputational strategy*. In Study 2, which specifically recruited *Democrat* subjects, we probed the robustness of this conclusion (Figure 2).

Given that punishment without looking can serve as a strong signal of loyalty, we might expect it to pay greater reputational dividends when one’s loyalty is in doubt. In Study 2a, we thus asked: do Evaluators reward punishment without looking when given reason to doubt the Actor’s loyalty? To investigate, we assigned *n* = 1796 Democrat Evaluators (across six conditions) to Actors who had considered the Moore petition. In the “baseline condition”, mirroring Study 1, Evaluators learned that their Actor was a Democrat. In the “Other participant” treatment, by contrast, Evaluators merely learned that their Actor was another participant—leaving the Actor’s values (and thus loyalty towards the petition’s cause) completely ambiguous. Finally, in four different treatments, Evaluators learned that their Actor was a Democrat, but also received *additional* information: they saw a (real) screenshot of their Actor responding to survey question(s). In each of these treatments, we anticipated that the responses in the screenshot would call the Actor’s loyalty into question.

Specifically, in the “Ignored injustice” treatment, the featured Actor described witnessing racial injustice, but doing nothing to respond. In two “Conflict of Interest” treatments, the Actors described reasons that signing the petition—which advocated firing Police Chief Moore—might conflict with their personal self-interest. (In the “Police” version of this treatment, the Actor...
reported having a police officer uncle; in the “Privacy” version, the Actor noted that signing could make their information public, elicitng spam.) Finally, in the “Previously Independent”
treatment, the Actor described a questionable history of partisan commitment (in which they voted in just one of the last four elections, and before becoming a Democrat identified as an
Independent and “praised Trump”).

Thus, across five treatments, we attempted to cast doubt on Actor loyalty, either by
removing or adding information—and, as reported in Table 2A, four of our five treatments
successfully achieved this aim (Table 2A). Yet none of our treatments caused Evaluators to
reward punishment without looking as a specific reputation strategy. In all six conditions of
Study 2a, Evaluators reacted significantly more negatively to Actors who punished without (vs.
with) looking. And none of our five treatments significantly mitigated the reputation costs of not
looking, relative to the baseline condition. In fact, “Ignored injustice” and both “Conflict of
Interest” treatments actually caused Evaluators to share relatively less money with punishers who
declined (vs. chose) to look (Table 2C and Figure 2B). Supplementary analyses of Study 2 (see
SI Section 2.3) suggest that these findings may reflect that, when given active reason to doubt an
Actor’s loyalty, Evaluators cease to see punishment without looking as a positive signal of
loyalty—perhaps because they become more inclined to interpret not looking as laziness, rather
than moral commitment.

Taking the reverse approach—and establishing a positive track record of Actor loyalty—
also did not cause Evaluators to preferentially reward punishment without looking. In Study 2b
(n = 595 Evaluators across two conditions), we compared a baseline condition to a treatment in
which we added information to highlight the Actor’s loyalty. Specifically, we described the
Actor as a Democrat, and provided a screenshot where the Actor described responding to racial
injustice. This “Responded to injustice” treatment increased perceived Actor loyalty (Table 2A).
Yet in both conditions of Study 2b, Evaluators again reacted more negatively to punishment
without (vs. with) looking, and our treatment (vs. baseline condition) did not significantly
mitigate the reputation costs of not looking (although we do observe a marginally significant
effect in this direction for our money shared DV) (Table 2C and Figure 2B). Supplementary
analyses suggest that highlighting Actor loyalty did make punishment without looking seem less
unfair (perhaps because it seemed less hasty or uninformed), but also diminished its value as a
signal of loyalty (perhaps because loyalty was already well-established)—such that the overall
reputation value of punishment without looking did not significantly change.

While our Study 2 treatments did not cause Evaluators to reward punishment without
looking as a specific reputation strategy, some of them did increase the reputation value of
punishment in general (Table 2B and Figure 2A). In all seven conditions of Study 2, Evaluators
reacted more positively to Actors who did (vs. did not) punish—but this preference for
punishment was significantly stronger, relative to the relevant baseline condition, in both the
“Other participant” treatment of Study 2a and the “Responded to justice” treatment of Study 2b.

Together, Study 2 thus reveals that Actors may gain more from punishing in general
when their loyalty is in doubt, or is particularly well-established—but they are still unlikely to be
preferentially rewarded for punishing without looking. These results cast further doubt on the
hypothesis that reputation drives punishment without looking as a specific reputational strategy.

Finally, we note that the baseline conditions of Studies 2a-b replicated all reported Study 1
effects on fairness, competence, and loyalty—including our mediation analyses, which we pre-
registered in Study 2. Study 2 also provides further evidence, both within the baseline conditions
and overall, for the moderating effects of Evaluator ideology (whereby strong partisans react
more positively to punishment in general, and less negatively to punishment without looking). Across all conditions of Study 2, however, even strong partisan Evaluators did not prefer punishment without looking. See SI Sections 3.3-3.5 for more information.

Figure 2. Evaluators do not prefer punishment without looking, even when given reason to doubt or trust the Actor’s loyalty. Shown are evaluations of Actors who did vs. did not punish, and who punished with vs. without looking, across conditions in Study 2. (A) Evaluators in all conditions preferred punishers to non-punishers, and this preference was significantly stronger in the “Other Participant” and “Responded to injustice” treatments. (B) Evaluators in all conditions dispreferred punishers who declined (vs. chose) to look, and no treatments significantly mitigated the reputation costs of not looking. Bars plot mean positivity ratings, and error bars are 95% CIs. For a version of Figure 2 that plots money shared, see SI Section 3.1.
Table 2. (A) Treatments in Study 2 successfully cast doubt on (Study 2a) or highlighted (Study 2b) Actor loyalty. As manipulation checks, we analyze Evaluators’ ratings of Actor loyalty, when given no information about the Actor’s punishment or looking behavior. We report descriptive statistics within each condition (first rows) and, for each treatment, compare ratings in the treatment vs. baseline condition (second rows; significant differences highlighted in blue).

(B) Some treatments increased the reputation value of punishment in general. We analyze evaluations of Actors who did vs. did not punish. For both positivity and sharing, we report simple effects of Actor punishment within each condition (first rows; significant effects highlighted in grey) and, for each treatment, compare the effect of punishment in the treatment vs. baseline condition (i.e., we test for punishment X treatment interactions) (second rows; significant interactions highlighted in blue). (C) No treatments caused Evaluators to prefer punishment without looking. We repeat the above approach, but analyze evaluations of Actors who punished without vs. looking (and report simple effects of not looking, and not looking X treatment interactions). A version of Table 2 with CIs on regression coefficients (not shown here for readability), see SI Section 3.2.

Studies 3 and 4

In Studies 1-2, Evaluators rewarded punishment in general, but not punishment specifically without looking. These studies thus highlight the reputational incentives that Evaluators create for Actors. In Studies 3-4, we transition to investigating how such reputational incentives do (and do not) drive Actors to punish without looking. We thus shift our focus from Evaluators to Actors.

We consider two distinct reputational contexts, that varied with respect to audience ideology. In Studies 3a (n = 1808 Democrat Actors who considered the Moore petition) and 3b (n = 1796 Republican Actors who considered the Amazon petition), Actors were assigned to a co-partisan Evaluator who identified as “a weak [Democrat/Republican], who only leans towards the party”. In Studies 4a (n = 1974 Democrat Actors who considered the Negy petition) and 4b (n = 1145 Republican Actors who considered the Amazon petition), Actors learned that their co-partisan Evaluator identified as “a strong [Democrat/Republican] who strongly supports [Black
Lives Matter/Blue Lives Matter]. Thus, Actors faced more ideological audiences in Study 4 than Study 3.

Of note, our Study 4a data was collected in two batches; however, all key results (i) held significantly within the first batch, and (ii) survive corrections for “peeking” between batches (see SI Section 4.1).

**Does making punishment observable drive punishment without looking?**

We begin investigating whether reputation drives Actors to punish without looking as a byproduct of incentivizing punishment in general. We thus ask: does making punishment observable to Evaluators—thereby placing reputational scrutiny on whether Actors punish in general—encourage punishment without looking? To answer, we compare the “Nothing Observable” and “Punishment Observable” conditions, which differed only in whether punishment was observable (Figure 3A).

**Study 3.** Before analyzing punishment without looking, we ask whether making punishment observable increased punishment overall. We find that, in Study 3, Democrats were marginally significantly more likely to punish in Punishment Observable (23%) than Nothing Observable (19%), \(b = .04 [-.01, .09], t = 1.69, p = .091, n = 1222\). Similarly, Republicans were marginally significantly more likely to punish in Punishment Observable (27%) than Nothing Observable (22%), \(b = .05 [-.002, .09], t = 1.89, p = .060, n = 1214\).

Next, we directly test the byproduct hypothesis by asking: did making punishment observable increase rates of punishment without looking? Among Democrats, rates of punishment without looking were comparable across the Punishment Observable (13%) and Nothing Observable (12%) conditions, \(b = .01 [-.02, .05], t = 0.68, p = .497, n = 1222\). However, Republicans were significantly more likely to punish without looking in Punishment Observable (19%) than Nothing Observable (14%), \(b = .05 [.01, .10, t = 2.57, p = .010, n = 1214\).

Thus, in Study 3—which featured less ideological Evaluators—making punishment observable had marginally significant effects on punishment overall, and increased punishment without looking among Republicans but not Democrats.

**Study 4.** In Study 4, making punishment observable significantly increased punishment overall. Democrats were more likely to punish in Punishment Observable (30%) than Nothing Observable (19%), \(b = .12 [.07, .16], t = 4.98, p < .001, n = 1319\). Similarly, Republicans were more likely to punish in Punishment Observable (25%) than Nothing Observable (15%), \(b = .10 [.05, .16], t = 3.59, p < .001, n = 763\).

Making punishment observable also significantly increased punishment without looking. Among Democrats, rates of punishment without looking were higher in Punishment Observable (16%) than Nothing Observable (10%), \(b = .06 [.02, .09], t = 3.13, p = .002, n = 1319\). Similarly, Republicans were more likely to punish without looking in Punishment Observable (20%) than Nothing Observable (12%), \(b = .08 [.03, .14], t = 3.14, p = .002, n = 763\).

Thus, in Study 4—which featured more ideological Evaluators—making punishment observable increased punishment overall, and punishment without looking specifically.

Together, Studies 3-4 suggest that reputation can fuel punishment without looking by incentivizing punishment in general. Furthermore, consistent with our finding from Studies 1-2 that more ideological Evaluators create stronger reputational incentives for punishment, we found stronger evidence for this process in Study 4 than Study 3.
Does making looking observable drive punishment without looking?

Next, we investigate whether reputation drives punishment without looking as a specific reputational strategy. We thus ask: does additionally making looking observable—thereby placing reputational scrutiny on how Actors come to their punishment decisions—further encourage punishment without looking? To answer, we compare the “Punishment Observable” and “Both Observable” conditions, which differed only in whether looking was observable (Figure 3B).

**Study 3.** Before analyzing rates of punishment without looking, we ask whether making looking observable influenced looking overall. In Study 3, Democrats were significantly more likely to look in Both Observable (52%) than Punishment Observable (35%), \( b = .17 \) [.11, .22], \( t = 5.88, p < .001, n = 1206 \). Similarly, Republicans were more likely to look in Both Observable (38%) than Punishment Observable (24%), \( b = .14 \) [.08, .19], \( t = 5.13, p < .001, n = 1197 \). These results suggest that Actors expected Evaluators to reward them for considering opposing perspectives—and thus that placing reputational scrutiny on looking is unlikely to encourage punishment without looking.

Next, we directly test how making looking observable influenced punishment without looking. Among Democrats, we find significantly lower rates of punishment without looking in Both Observable (9%) than Punishment Observable (13%), \( b = -.04 \) [-.08, -.01], \( t = -2.35, p = .019, n = 1206 \). And among Republicans, we find no significant difference between rates of punishment without looking in Both Observable (16%) and Punishment Observable (19%), \( b = -.03 \) [-.08, -.01], \( t = -1.53, p = .127, n = 1197 \).

Thus, Study 3—which featured less ideological Evaluators—provides no evidence that making looking observable encircles punishment without looking. Rather, making looking observable increased looking overall, and, among Democrats, significantly decreased punishment without looking.

**Study 4.** In Study 4, making looking observable again increased looking. Democrats were significantly more likely to look in Both Observable (49%) than Punishment Observable (41%), \( b = .08 \) [.02, .13], \( t = 2.72, p = .007, n = 1284 \). Similarly, Republicans were more likely to look in Both Observable (28%) than Punishment Observable (15%), \( b = .13 \) [.07, .19], \( t = 4.52, p < .001, n = 779 \).

Furthermore, making looking observable still did not encourage punishment without looking. Among Democrats, we find no significant difference between rates of punishment without looking in Both Observable (13%) and Punishment Observable (16%), \( b = -.03 \) [-.07, .01], \( t = -1.41, p = .158, n = 1284 \). Similarly, Republicans show no significant difference between Both Observable (19%) and Punishment Observable (20%), \( b = -.01 \) [-.07, .04], \( t = -.45, p = .651, n = 779 \). Thus, in Study 4—which featured more ideological Evaluators—making looking observable again increased looking, and did not encourage punishment without looking.

Studies 3-4 thus provide no evidence that reputation can drive punishment without looking as a specific reputational strategy, even when the reputational audience is strongly ideological. In both studies, showcasing looking decisions encouraged Actors to consider opposing perspectives, and diminished or did not influence rates of punishment without looking. Interestingly, however, making looking observable had directionally smaller positive effects on looking, and negative effects on punishment without looking, in Study 4 than in Study 3. This pattern is consistent with our finding from Studies 1-2 that more ideological Evaluators react less negatively to punishment without (vs. with) looking.
In sum, Studies 3-4 suggest that reputation can drive punishment without looking as a byproduct of driving punishment in general, but not as a specific reputational strategy. Moreover, they suggest that ideological audiences may facilitate punishment without looking by doing more to encourage punishment in general, and less to encourage looking and discourage punishment without looking.

Finally, we note that we report how making punishment observable influenced looking in SI Section 3.2, and how making looking observable influenced punishment in Section 3.3. These analyses were also pre-registered, but are not reported in the main text because they are less relevant to our key theoretical questions.

Figure 3. Making punishment observable drives punishment without looking, while making looking observable does not. (A) Placing reputational scrutiny on punishment (by making punishment observable) drove Actors to punish more overall, and to punish specifically without looking. We plot the proportion of subjects who chose to (i) punish and (ii) punish without looking, across the Nothing Observable vs. Punishment Observable conditions. (B) Placing reputational scrutiny on looking (by making looking observable) increased rates of looking overall, and did not drive punishment without looking. We plot the proportion of subjects who chose to (i) look and (ii) punish without looking, across the Punishment Observable vs. Both Observable conditions. Error bars are 95% CIs.
Discussion

Across four studies (total $n = 10,343$), we have shown that reputation fuels punishment without looking—not as a specific reputational strategy, but as a byproduct of encouraging punishment in general. Highlighting the general power of reputation to encourage punishment, Evaluators rewarded punishers, and Actors punished more when punishment was observable. And because some Actors who were reputationally-induced to punish chose not to look, making punishment observable increased not just punishment overall, but also punishment without looking specifically. Thus, reputation drove punishment without looking as a byproduct.

Yet reputation did not drive punishment without looking as a specific reputational strategy. Evaluators saw punishing without looking as an especially strong signal of loyalty, suggesting that eschewing opposing perspectives can enhance a punisher’s perceived commitment to a moral cause. However, declining to look also made punishers appear less fair and competent—perhaps because people see hastily-enacted punishment as potentially unjustified, and associate deliberation with competence. On net, then, Evaluators dispreferred punishers who declined to look. And correspondingly, making looking observable did not encourage Actors to punish without looking, and actually increased looking overall.

Our results also highlight the robustness of Evaluators’ preferences for deliberative punishment. Despite our finding that punishment without looking can be a strong signal of loyalty, casting doubt on an Actor’s loyalty did not make Evaluators more likely to reward such punishment. In fact, when Evaluators had active basis to question an Actor’s loyalty, they did not interpret punishment without looking as a signal of loyalty in the first place. Furthermore, giving Evaluators reason to trust an Actor’s loyalty did not cause them to preferentially reward punishment without looking, either.

Thus, we find no support for the hypothesis that reputation drives people to punish specifically without looking, in order to appear especially virtuous. Rather, our results suggest that punishers who ignore conflicting viewpoints simply prefer not to look. This preference could reflect that evaluating opposing perspectives takes time and effort, and can require engaging with dissimilar (and potentially disliked) others. Moreover, some people may avoid looking because they are motivated to see punishment as merited (e.g., for ideological reasons, or because punishing has reputational benefits) and do not wish to have their minds changed. Discriminating between these possibilities is an interesting direction for future research.

Our findings have implications for how society might encourage more deliberative decision-making. According to the current zeitgeist, the deep human sensitivity to reputation is a problem to be managed, rather than a tool to be leveraged, when it comes to encouraging reflective punishment. Our work, however, challenges this perspective. While we do find that reputation inspires punishment that is sometimes unreflective, we also find that people correctly believe that engaging with opposing perspectives will be perceived positively by others. Thus, by giving people opportunities to broadcast their engagement with other viewpoints, it may actually be possible to leverage reputation to motivate careful deliberation.

Still, this approach might not be a silver bullet. When reputation inspires “looking”, the result may sometimes be more performative than truly reflective. Our specific methodology suggests that reputation can genuinely increase intentions to consider opposing perspectives: when looking was observable, Actors clicked more links to opposing perspective articles, despite not knowing that their link-clicking was tracked. Yet we do not know how frequently or deeply Actors actually read the linked articles. When individuals face reputational incentives both to consider opposing perspectives and to ultimately punish, they may sometimes engage with
countervailing perspectives superficially or in motivated ways (that bias them towards punishment) — especially if their incentives to punish are stronger than their incentives to look.

Our results also highlight the potential importance of audience ideology — both for generally driving punishment, and specifically driving reflexive punishment. We found that more ideological Evaluators reacted more positively to punishment in general, and less negatively to punishment without (vs. with) looking. Moreover, comparing Studies 3 and 4 suggest that, in the presence of more ideological audiences, (i) making punishment observable may do more to encourage punishment overall and punishment without looking, and (ii) making looking observable may do less to encourage looking and discourage punishment without looking. Still, the comparison between Studies 3 and 4 remains merely suggestive: we did not randomly assign Actors to Study 3 vs. 4, and the studies featured different Democrat petitions. Future work should continue exploring the influence of audience ideology on punishment behavior.

Importantly, even strongly partisan Evaluators did not significantly prefer punishment without (vs. with) looking in our studies. Yet our ideology findings, and our finding that punishment without looking does strongly signal loyalty, suggest a framework for thinking about when punishment without looking might be evaluated most positively. For example, punishment without looking might conceivably be rewarded as a specific reputation strategy by audiences that are extremely ideological, or that prize loyalty over fairness and competence (e.g., extremist groups) — or in situations that encourage even more moderate audiences to value loyalty more strongly.

Future research should evaluate the generalizability of our results across populations (e.g., representative samples of Americans, or samples from other cultures), moral domains (given that our studies focused on just a few petitions), and social environments (including environments beyond opt-in online participant pools; an investigation of punishment on social media platforms would be of particular interest). It would also be valuable to consider different operationalizations of punishment and of looking. For example, the reputation value of looking might increase for severe punishments (e.g., physical aggression, firing somebody) and lessen for milder punishments (e.g., condemnatory gossip in a private, dyadic conversation).

Similarly, there are many ways that somebody might look (i.e., gather relevant information) before deciding whether to punish. Future studies could investigate other looking behaviors, including gathering or verifying facts, or evaluating perspectives that are not countervailing (i.e., seeking confirmatory evidence that punishment is merited). Of note, relative to these other looking behaviors, considering opposing perspectives might appear particularly disloyal. It is perhaps especially striking, then, that looking was still rewarded in our studies. However, the amount of looking might matter too: if an individual considers many opposing perspectives, or demands a huge amount of evidence in favor of punishment, their looking might become a net reputational negative.

In conclusion, people sometimes punish alleged wrongdoers without first considering opposing perspectives. We find that reputation can fuel such “punishment without looking” by incentivizing punishment in general. Yet we find no evidence that people use unquestioning punishment as a specific reputational strategy, and show that engaging with opposing viewpoints can actually be socially rewarded.

Materials and Methods

For all studies, full procedures and stimuli are documented in SI Sections 1 and 6, and data, scripts, materials, and pre-registrations are online.
All studies were approved by an IRB, and began with subjects providing informed consent. Subjects also completed attention checks (collected prior to random assignment) and comprehension questions about study tasks. Our main text analyses restrict to subjects who passed the attention checks; see SI Sections 2.6 and 4.4 for analyses that additionally restrict to subjects who passed all comprehension questions (and produce very similar results).

For our Evaluator studies (i.e., Studies 1-2), we recruited a small number of Actors who were not subjects in any of our studies (although they did provide informed consent), but were recruited so that we could describe real Actors to Evaluators, and pay study bonuses. We matched each featured Actor with multiple Evaluators, although Evaluators did not learn this. Likewise, for our Actor studies (i.e., Studies 3-4), we recruited a small number of Evaluators (who provided informed consent) to match with Actors. See SI Section 1.4 for more details about the matching procedures for each study.

Studies 1a-b
In May-June 2021, we recruited a target of \( n = 600 \) Evaluators from Prolific for each of Studies 1a (final \( n = 629 \) Democrats, 46% male, average age = 34 years) and 1b (final \( n = 600 \) Republicans, 45% male, average age = 37 years). In each study, we (i) introduced the Dictator Game, (ii) described the Actor’s punishment and looking opportunities, (iii) measured evaluations of Actors, and (iv) presented demographic and other questions (including a binary strength of partisanship measure, used in our moderation analyses).

In stage (iii), subjects first evaluated Actors who did vs. did not look (described as spending an above vs. below average amount of time considering opposing perspectives), before punishing. Next, in corresponding order, subjects evaluated Actors who did vs. did not look, before not punishing. Subjects then evaluated Actors who did vs. did not punish (without looking information). Finally, subjects evaluated an Actor about whom they had no information. For each Actor, subjects (i) decided what to share if matched with the Actor (between 0 and 50¢ in 5-cent increments; we analyze the percentage of 50¢ shared), (ii) rated the Actor on overall positivity, and then (iii) in random order, rated the Actor on loyalty, competence, and fairness.

For all studies, we adhered closely to our pre-registered analysis plans with some minor deviations. Of particular note, in Study 1 we pre-registered money shared as our primary DV, with analyses of overall positivity ratings as secondary. After conducting Study 1, however, we concluded, based on these variables’ distributions, that positivity ratings were a more informative global evaluation measure. Whereas many Evaluators shared none or exactly half of their endowment, and almost no Evaluators shared more than half, positivity ratings were more continuously distributed. We thus pre-registered positivity as our primary DV in Study 2. For both Studies 1-2, we report analyses of both global evaluation DVs, but plot positivity in our main text figures; we reproduce these figures for sharing in the SI. For a complete discussion of all pre-registration deviations, see SI Section 5.

Studies 2a-b
In October 2022, we recruited a target of (i) \( n = 1800 \) Evaluators for Study 2a (final \( n = 1796 \) Democrats, 41% male, average age = 37 years), and (ii) \( n = 600 \) Evaluators for Study 2b (final \( n = 595 \) Democrat, 49% male, average age = 38 years), from Prolific. Study 2, which was conducted chronologically last, did not include Republicans; we did not believe that enough Republicans on Prolific remained naïve to our paradigm to afford sufficient statistical power.
The method for Study 2 closely mirrored Study 1a, except that we gave subjects different background information about the Actor they were matched with (which varied across conditions, as described in the main text).

**Studies 3a-b**

In January-March 2021, we recruited a target of \( n = 1800 \) Democrat Actors from MTurk for Study 3a (final \( n = 1808, 39\% \) male, average age = 38 years). For Study 3b, we recruited a target \( n = 1800 \) Republican Actors, first from Mturk and then from Prolific, once it became clear that an insufficient number of Republicans on Mturk remained naïve to our paradigm (final \( n = 1796, 52\% \) from MTurk and 48% from Prolific, 46% male, average age = 40 years). The flow of Study 3 mirrored Studies 1-2, except that subjects *themselves* had the chance to punish and look, and we manipulated what they were told that the Evaluator would learn about their behavior.

To measure looking, we presented a “looking screen” with headlines and links for two opposing perspective articles. We tracked subjects’ link-clicking behavior and define “looking” as clicking at least one link to an opposing perspective article. Subjects in Both Observable were told that the Evaluator would learn how long they spent on the looking screen, but not that we would track their link-clicking behavior. See SI Section 4.5 for pre-registered secondary analyses that define looking as either (i) time spent on the looking screen or (ii) the continuous number of links clicked (and produce very similar results).

To measure punishment, we presented a link to the petition, and tracked whether subjects clicked it (again, without informing subjects of this tracking). We also asked subjects whether they signed the petition. To discourage false reporting, we asked subjects, if they signed, to report information about the screen they saw after signing. We define “punishing” as clicking the link to the petition and self-reporting signing.

**Studies 4a-b**

In October-November 2020, we conducted Studies 4a-b. For each study, we pre-registered initial targets of \( n = 1200 \) Actors from MTurk (less than we targeted in Studies 3a-b because, in Study 3 but not Study 4, we hypothesized that making looking observable might *decrease* punishment without looking—but expected this effect to be small, and thus require a larger sample to detect).

For Study 4a (which recruited Democrats), upon reaching our initial target of \( n = 1200 \), our key results were all significant. However, while finishing data collection for Study 4b (which recruited Republicans, and took longer to complete), we decided to direct Democrats to Study 4a, increasing its sample size. We thus amended our Study 4a preregistration, increasing our target to \( n = 2000 \) subjects, and registering a plan to correct, in our analyses, for “peeking” at the first batch of data before collecting the second batch. All significant results survive these corrections (see SI Section 4.1). Our final sample sizes were \( n = 1974 \) for Study 4a (47% male, average age = 40 years) and \( n = 1145 \) for Study 3b (44% male, average age = 43 years).

The Study 4 method was otherwise identical to the Study 3 method, except that (i) subjects were matched with more ideological Evaluators and (ii) Study 4a featured the Negy petition (whereas Study 3a featured the Moore petition).
References


Supplementary Information
for
How reputation does (and does not) drive people to punish without looking

1. Full study procedures ........................................................................................................... 3
   1.1 Studies 1a-b ...................................................................................................................... 3
   1.2 Studies 2a-b .................................................................................................................... 4
   1.3 Studies 3-4 ..................................................................................................................... 4
   1.4 Matching Actors and Evaluators ................................................................................... 5

2. Supplemental analyses of Studies 1a-b ............................................................................... 6
   2.1 Analyses of results for the Negy vs. Moore petitions in Study 1a .................................... 6
   2.2 Reproducing main text Figure 1 with sharing ................................................................. 7
   2.3 Mediation analyses ......................................................................................................... 7
   2.4 Between-subjects analyses ............................................................................................. 8
   2.5 Analyses of evaluations of non-punishers who declined vs. chose to look .................... 9
   2.6 Analyses of perfect comprehenders .............................................................................. 9
       2.6.1 Evaluations of punishers vs. non punishers ............................................................ 10
       2.6.2 Evaluations of punishers who declined vs. chose to look ....................................... 10

3. Supplemental analyses of Studies 2a-b .............................................................................. 10
   3.1 Reproducing main text Figure 2 with sharing ................................................................. 10
   3.2 Reproducing main text Table 2 with 95% CIs ............................................................... 11
   3.3 Study 2 analyses for all DVs .......................................................................................... 12
       3.3.1 Ratings of Actors, when given no information about punishment or looking .......... 12
       3.3.2 Ratings of Actors who did vs. did not punish ....................................................... 13
       3.3.3 Ratings of Actors who punished without vs. with looking .................................... 13
       3.3.4 Ratings of Actors who did not punish without vs. with looking ......................... 14
       3.3.5 Interpretation of analyses ....................................................................................... 14
   3.4 Mediation analyses ......................................................................................................... 16
   3.5 Moderation analyses ...................................................................................................... 17

4. Supplemental Analyses of Studies 3-4 ............................................................................. 20
   4.1 Analyses of Study 4a by batch, and corrections for “peeking” ....................................... 20
   4.2 Analyses of how making punishment observable influences looking ......................... 20
   4.3 Analyses of how making looking observable influences punishment ........................... 21
   4.4 Analyses of perfect comprehenders ............................................................................. 22
4.4.1 Effects of making punishment observable .......................................................... 22
4.4.2 Effects of making looking observable ............................................................... 23
4.5 Analyses of alternative specifications of looking .................................................. 23
   4.5.1 Looking time (natural-log transformed seconds) ........................................... 23
   4.5.2 Number of articles ......................................................................................... 24
5. Discussion of preregistration deviations .................................................................. 25
   5.1 Studies 1-2 ......................................................................................................... 25
   5.3 Studies 3-4 ......................................................................................................... 26
6. Appendix: Experimental stimuli ................................................................................ 26
   6.1 Studies 1a-b ....................................................................................................... 26
   6.2 Studies 2a-b ....................................................................................................... 40
   6.3 Studies 3-4 ......................................................................................................... 46
1. Full study procedures

Below, we provide a narrative description of the procedure for each study. In Section 6 of this SI, we show the complete stimuli from each study; this section thus provides the specific wording we used to explain all study tasks and measure all dependent variables.

1.1 Studies 1a-b

As in all studies in this paper, subjects in Studies 1a-b began by reporting their gender and political party, and then completing two attention checks. Next, we introduced subjects to the Dictator Game involving the Actor. We told subjects that they (i.e., the Evaluator) would be “Player 1” and another subject (i.e., the Actor), of their same political affiliation, would be “Player 2”. Furthermore, subjects learned that they would receive 50 cents, and decide how much to share with Player 2. Subjects then answered three comprehension questions about the Dictator Game.

Next, we introduced subjects to the Actor’s punishment and looking decisions. We told subjects that Player 2 had the chance to sign the relevant punitive petition, and provided subjects with a brief description, and screenshot, of the petition. Next, we told subjects that Player 2 had the chance to consider opposing perspectives before deciding whether to sign. In particular, we told subjects that Player 2 received links to articles providing opposing perspectives (and provided a screenshot of an example headline) and that Player 2 could search the Internet for other opposing perspectives. We also told subjects that Player 2s varied in the time they spent considering opposing perspectives.

Next, we collected subjects’ evaluations of different Actors. First, we asked subjects to consider a Player 2 who did sign the petition, after considering opposing perspectives for an above- or below-average amount of time (manipulated between-subjects). Thus, we described a punisher who did, or did not, look. We told subjects that they might be matched with this Player 2, and then measured the amount of money they wished to share with Player 2 in this event (between 0 and 50 cents in 5-cent increments). Afterwards, subjects rated the Player 2 on overall positivity, and then, in random order, loyalty towards Black Lives Matter or Blue Lives Matter, competence, and fairness (on 0-to-100 scales).

Next, we told subjects that they would make sharing decisions about other Player 2s they might be matched with. Subjects then evaluated six other Player 2s, with six other behavioral profiles. The second Player 2 was a punisher who did or did not look ( whichever the subject did not evaluate first). Next, subjects evaluated non-punishers who did, or did not, look (in an order corresponding to the order in which they evaluated punishers). Then, Subjects evaluated non-punishers and punishers, without receiving information about their looking (in random order). Finally, subjects evaluated a Player 2 without receiving any information about their decisions.

Afterwards, we presented subjects with a post-experimental survey in which they evaluated the petition, described how they made their decisions, shared their impressions of the study, rated their previous participation in related studies and beliefs about whether the petition and Player 2 were real, and answered demographic and ideology questions. In this survey, subjects answered a binary strength of partisanship measure, used in our moderation analyses (in which they were asked to identify as either “a strong Democrat/Republican, who strongly supports the party” or “a weak Democrat/Republican, who only leans towards the party”).
1.2 Studies 2a-b

The method for Study 2 closely mirrored Study 1a, with a few differences. In Study 1, subjects evaluated multiple distinct Player 2s, who they learned that they might be matched with, and who behaved in different ways. In contrast, subjects in Study 2 evaluated just one Player 2, who they learned that they were matched with, in the event that this Player 2 behaved in different ways. This allowed us to, in our Study 2 treatments that added information, provide subjects with specific background information about the Player 2 they were paired with. Furthermore, to encourage subjects to reflect on this background information, we asked subjects across all conditions of Study 2 to write a paragraph about their initial impression of Player 2, before they evaluated different potential Player 2 behaviors.

In the “baseline” conditions of Studies 2a-b, the method was otherwise identical to the Study 1a method. In the “other person” treatment of Study 2a, when Player 2 was introduced, Player 2 was characterized as “another participant” and no information was provided about Player 2’s political affiliation. In all other treatments of Studies 2a-b, Player 2 was initially described as a co-partisan (like in Study 1a and the baseline conditions). Then, after subjects learned about Player 2’s punishment and looking opportunities, they were shown a screenshot of some questions that Player 2 was asked, and Player 2’s responses. These screenshots varied across treatments, and were shown on the page where subjects described their initial impression of Player 2. In the main text, we overview the content of these screenshots in each treatment; in Section 6.2 of this SI, we reproduce each screenshot.

1.3 Studies 3-4

After reporting their gender and political party, and completing two attention checks, subjects in Studies 3-4 learned about the Dictator Game. We told subjects that they (i.e., the Actor) would be Player 2 and another subject (i.e., the Evaluator) of their same political affiliation would be Player 1, receive 50 cents, and decide how much to share with them. We described the Player 1 as “a weak [Democrat/Republican], who only leans towards the party” in Study 3, and “a strong [Democrat/Republican] who strongly supports [Black Lives Matter/Blue Lives Matter]” in Study 4. Subjects then answered three comprehension questions about the Dictator Game (and were required to correct any incorrect answers before preceding).

Next, we introduced subjects to their punishment and looking decisions. We presented subjects with a brief description, and screenshot, of the relevant petition, explaining that they would later have the chance to sign it. Subjects in Nothing Observable learned that Player 1 would not find out whether they signed; subjects in Punishment Observable and Both Observable learned that Player 1 would find this out.

Afterwards, subjects learned that they could consider opposing perspectives before deciding whether to sign. We told subjects that we would provide links to opposing perspective articles (and provided a screenshot of an example headline) and that they could also search the Internet for other opposing perspectives. Subjects in Nothing Observable and Punishment Observable learned that Player 1 would not find out how long they spent considering opposing perspectives; subjects in Both Observable learned that Player 1 would find this out. Subjects then answered three comprehension questions about what Player 1 would learn about their behavior (and again were required to correct any incorrect answers).

We then told subjects they would next get to consider opposing perspectives, and reminded them about whether their time spent doing so would be observable. Then, subjects advanced to a screen in which we provided headlines for and links to two opposing perspective
articles, and reminded subjects than they could search for other articles. We told subjects to advance the screen once they were ready to decide whether to sign the petition. On this “looking” screen, we tracked subjects’ link-clicking behavior and, per our preregistration, defined “looking” as clicking at least one link to an opposing perspective article. We did not tell subjects that their link-clicking would be tracked.

We then told subjects that they would next decide whether to sign the petition, and that if they reported signing, we would ask them to show us that they signed by reporting information about the screen that Change.org displayed after they signed. (We also assured subjects that we would not collect their identifying information if they signed). Next, we gave subjects a link to the petition and asked them whether they signed (and prompted them to report information about the completion screen if so). We also asked subjects about their level of commitment to supporting the petition; per our preregistrations, we do not analyze this DV.

On this “petition” screen, we tracked subjects’ link-clicking behavior and, per our preregistration, defined “signing” (i.e., punishing) as clicking the link to the petition and self-reporting signing. We did not tell subjects that their link-clicking would be tracked. We did not incorporate the reported completion screen information into our pre-registered definition of “signing”, because (i) pilot testing revealed that almost everyone who met the above criteria did report this information correctly, but (ii) there is some room for ambiguity in categorizing reports as “correct” (given that some responses are close to but not completely accurate).

Afterwards, we presented subjects with a post-experimental survey in which they evaluated the petition, described how they spent their time on the “looking” screen and made their decisions, rated their previous familiarity with the petition, the extent to which they were motivated by reputation, and their beliefs about whether the petition, articles, and Player 1 were real, and answered demographic and ideology questions.

1.4 Matching Actors and Evaluators

In Studies 1-2, focal subjects were Evaluators. Each Evaluator was matched with a real Actor. To pay each Evaluator’s bonus, we implemented one of their sharing decisions, corresponding to how their Actor actually behaved.

The Actors that Evaluators in Studies 1-2 were matched with were not themselves subjects in any of our studies (i.e., we do not report analyses of their behavior, although they did provide informed consent to participate). Rather, we recruited a small number of Actors, with the specific aim of matching them with Evaluators (so that we could describe real Actors to Evaluators, and pay study bonuses). We matched each featured Actor with multiple Evaluators, although Evaluators did not learn this.

When recruiting Actors to match with Evaluators from Study 2, we asked a set of survey questions that we believed might elicit responses that would cast doubt on, or highlight, Actors’ loyalty. These Actors provided informed consent for their responses to potentially be shown to other online participants (who they learned might view their responses, evaluate them, and decide how much money to share with them).

In Study 1, we simply matched Evaluators with co-partisan Actors. And in the “baseline” and “other participant” conditions of Study 2, we likewise matched Evaluators (who were all Democrats) with Democrat Actors (although Evaluators in the “other participant” condition did not learn that their Actor was a Democrat). In the remaining conditions of Study 2, we matched Evaluators with specific Democrat Actors who we chose to feature, on the basis of their responses to the above-described questions (selecting responses that cast doubt on the Actor’s
loyalty in the relevant conditions of Study 2a, and selecting a response that highlighted the Actor’s loyalty in the relevant condition of Study 2b). Each of these conditions featured a different Actor, and we showed Evaluators in these conditions a screenshot of their Actor’s survey responses.

In Studies 3–4, focal subjects were Actors. Each Actor was matched with a real Evaluator. To pay each Actor’s bonus, we implemented one of their Evaluator’s sharing decisions, corresponding to how the Actor’s observability condition and how the Actor actually behaved (e.g., if Actor was in the “Punishment Only” condition and chose to punish, we paid the Actor the amount that their Evaluator chose to share with a punisher, given no information about looking).

Mirroring Studies 1–2, the Evaluators that Actors in Studies 3–4 were matched with were not themselves subjects in any of our studies (although they did provide informed consent to participate). Rather, we recruited a small number of Evaluators specifically to match with Actors (and matched each featured Evaluator with multiple Actors, without Actors knowing this). In Study 3, we matched Actors with co-partisan Evaluators who reported identifying as weak partisans, and in Study 4, we matched Actors with co-partisan Evaluators who reported identifying as strong partisans, and also reported strongly supporting Black/Blue Lives Matter.

2. Supplemental analyses of Studies 1a–b

2.1 Analyses of results for the Negy vs. Moore petitions in Study 1a

In Study 1a, subjects were randomly paired with Actors who either had the chance to sign the Negy (n = 308) or Moore (n = 321) petition. As reported in the main text, the Negy and Moore petitions produced identical patterns of results; thus, our main text results collapsed over petition. However, per our pre-registration, in Table S1 below, we report our key Study 1 analyses separately for each petition, and also report interaction results that compare results across the two petitions. We find that each of our key results hold significantly within each of the two petitions. However, for some results, the effect size differs significantly across petitions, as reflected by a significant interaction.

<table>
<thead>
<tr>
<th>Evaluations of punishers vs. non-punishers [(Positive coefficients reflect preferences for punishment)]</th>
<th>Evaluations of punishment without vs. with looking [(Positive coefficients reflect preferences for punishment without looking)]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positivity</strong></td>
<td><strong>Positivity</strong></td>
</tr>
<tr>
<td>Negy Petition: b = 26.81 [23.17, 30.45], t = 14.51, p &lt; .001</td>
<td>Negy Petition: b = -8.56 [-11.25, -5.87], t = -6.26, p &lt; .001</td>
</tr>
<tr>
<td>Interaction: b = -5.36 [-10.41, -0.31], t = -2.08, p = .038</td>
<td>Interaction: b = -3.73 [-7.44, -0.03], t = -1.98, p = .048</td>
</tr>
<tr>
<td><strong>Sharing</strong></td>
<td><strong>Sharing</strong></td>
</tr>
<tr>
<td>Moore Petition: b = 13.71 [11.05, 16.36], t = 10.17, p &lt; .001</td>
<td>Moore Petition: b = -5.05 [-6.85, -3.24], t = -5.49, p &lt; .001</td>
</tr>
<tr>
<td>Interaction: b = -4.28 [-8.30, -0.26], t = -2.09, p = .037</td>
<td>Interaction: b = -1.35 [-4.25, 1.56], t = -0.91, p = .363</td>
</tr>
<tr>
<td><strong>Fair</strong></td>
<td><strong>Fair</strong></td>
</tr>
<tr>
<td>Negy Petition: b = 21.60 [18.39, 24.82], t = 13.22, p &lt; .001</td>
<td>Negy Petition: b = -15.73 [-18.50, -12.97], t = -11.19, p &lt; .001</td>
</tr>
<tr>
<td>Moore Petition: b = 16.96 [14.02, 19.89], t = 11.38, p &lt; .001</td>
<td>Moore Petition: b = -16.84 [-19.87, -13.81], t = -10.94, p &lt; .001</td>
</tr>
<tr>
<td>Interaction: b = -4.65 [-8.99, -0.31], t = -2.10, p = .048</td>
<td>Interaction: b = -1.11 [-5.20, 2.99], t = -0.53, p = .594</td>
</tr>
<tr>
<td><strong>Competent</strong></td>
<td><strong>Competent</strong></td>
</tr>
<tr>
<td>Negy Petition: b = 19.62 [16.64, 22.60], t = 12.95, p &lt; .001</td>
<td>Negy Petition: b = -10.18 [-12.74, -7.62], t = -7.82, p &lt; .001</td>
</tr>
<tr>
<td>Moore Petition: b = 15.30 [12.49, 18.11], t = 10.71, p &lt; .001</td>
<td>Moore Petition: b = -11.99 [-14.28, -9.69], t = -10.27, p &lt; .001</td>
</tr>
<tr>
<td>Interaction: b = -4.31 [-8.40, -0.22], t = -2.07, p = .039</td>
<td>Interaction: b = -1.81 [-5.24, 1.62], t = -1.04, p = .300</td>
</tr>
<tr>
<td><strong>Loyal</strong></td>
<td><strong>Loyal</strong></td>
</tr>
<tr>
<td>Negy Petition: b = 36.39 [33.11, 39.66], t = 21.85, p &lt; .001</td>
<td>Negy Petition: b = 4.69 [2.37, 7.00], t = 3.99, p &lt; .001</td>
</tr>
<tr>
<td>Moore Petition: b = 37.24 [34.19, 40.28], t = 24.05, p &lt; .001</td>
<td>Moore Petition: b = 6.37 [4.16, 8.59], t = 5.66, p &lt; .001</td>
</tr>
<tr>
<td>Interaction: b = 0.85 [-3.81, 5.51], t = 0.37, p = .708</td>
<td>Interaction: b = 1.69 [-1.51, 4.88], t = 1.04, p = .300</td>
</tr>
</tbody>
</table>

Table S1. Results of Study 1a by petition.
2.2 Reproducing main text Figure 1 with sharing

Below, in Figure S1, we reproduce main text Figure 1 (which plotted positivity ratings) with our sharing DV. We find qualitatively identical patterns, with one exception. For positivity, strong partisan Democrats showed a significant preference for punishment with (vs. without) looking; for sharing, strong partisan Democrats show no significant difference between punishment without vs. with looking.

**Figure S1. Reproducing main text Figure 1, but with our sharing DV.**

2.3 Mediation analyses

As described in the main text, Evaluators in Studies 1a-b rated punishers who declined (vs. chose) to consider opposing perspectives less positively overall, and also rated them as less fair and less competent. Interestingly, however, they rated punishers who declined to look as *more* loyal. This pattern of results suggests that (i) declining to look has reputational costs and advantages, and (ii) the advantages are outweighed by the costs, such that declining to look is a net reputational negative. Here, we support this proposal with mediation analyses.

Before describing our mediation analyses, we note that two of our mediators, fairness and competence, were highly correlated with each other (among Democrats, $r = .86, p < .001$; among Republicans, $r = .87, p < .001$). In contrast, loyalty was less strongly correlated with the other mediators (correlations with fairness: Democrats, $r = .63, p < .001$; Republicans, $r = .51, p < .001$; correlations with competence: Democrats, $r = .59, p < .001$; Republicans, $r = .49, p < .001$).

In light of this correlation structure, we consider two distinct multiple mediation models: one that takes fairness and loyalty as mediators, and one that takes competence and loyalty as mediators. This approach avoids placing fairness and competence together in a single model, which would be uninformative in light of their high collinearity. In each model, we take overall
positivity as our dependent variable, the contrast between punishers who declined vs. chose to look as our independent variable, and the relevant pair of mediators as our mediating variables.

In Table S2, for each of Studies 1a-b, we report the total effect of not looking on overall positivity, as well as analyses from each of our two mediation models. In particular, for each mediator within each model, we report (i) the A path (i.e., the effect of not looking on the mediator), (ii) the B path (i.e., the effect of the mediator on overall positivity, controlling for looking and the other mediator), and (iii) the indirect effect of not looking on overall positivity via the mediator. Additionally, for each model, we report the direct effect of not looking on overall positivity.

For each of Studies 1a-b, we consistently observe significant negative indirect effects for fairness and competence, and significant positive indirect effects for loyalty. Furthermore, the positive indirect effects of loyalty are consistently significantly smaller in magnitude than the negative indirect effects of fairness and competence (as revealed by the 95% CIs reported in Table S1), consistent with the negative total effect of not looking on overall positivity.

<table>
<thead>
<tr>
<th>Model with Fair &amp; Loyal as mediators</th>
<th>Democrats (Study 1a)</th>
<th>Republicans (Study 1b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effects for Fair</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Path B: b = .75 [70, 81], z = 27.59, p &lt; .001</td>
<td>Path B: b = .78 [74, 82], z = 36.77, p &lt; .001</td>
<td></td>
</tr>
<tr>
<td>Indirect effect: b = -12.27 [-14.05, -10.50]</td>
<td>Indirect effect: b = -6.74 [-8.16, -5.32]</td>
<td></td>
</tr>
<tr>
<td>Effects for Loyal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Path A: b = 5.55 [3.95, 7.14], z = 6.83, p &lt; .001</td>
<td>Path A: b = 6.90 [4.92, 8.88], z = 6.83, p &lt; .001</td>
<td></td>
</tr>
<tr>
<td>Path B: b = .12 [0.6, 0.17], z = 4.36, p &lt; .001</td>
<td>Path B: b = .08 [0.04, 0.12], z = 4.35, p &lt; .001</td>
<td></td>
</tr>
<tr>
<td>Indirect effect: b = .65 [30, 1.00]</td>
<td>Indirect effect: b = .56 [26, 86]</td>
<td></td>
</tr>
<tr>
<td>Direct effect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b = 1.16 [-35, 2.67], z = 5.15, p = .13</td>
<td>b = 1.63 [28, 2.98], z = 3.27, p = .018</td>
<td></td>
</tr>
</tbody>
</table>

| Model with Competent & Loyal as mediators |
|------------------------------------------|----------|
| Effects for Competent                    |          |
| Path A: b = -11.10 [-12.81, -9.39], z = -12.74, p < .001 | Path A: b = -8.20 [-6.86, -3.53], z = -6.07, p < .001 |
| Path B: b = .78 [73, 83], z = 30.40, p < .001 | Path B: b = .78 [73, 84], z = 28.86, p < .001 |
| Indirect effect: b = -8.63 [-10.07, -7.19] | Indirect effect: b = -4.07 [-5.41, -2.73] |
| Effects for Loyal                        |          |
| Path A: b = 5.55 [3.95, 7.14], z = 6.83, p < .001 | Path A: b = 6.90 [4.92, 8.88], z = 6.83, p < .001 |
| Path B: b = .13 [0.07, 0.18], z = 4.43, p < .001 | Path B: b = .06 [0.02, 0.10], z = 3.18, p < .001 |
| Indirect effect: b = .70 [33, 1.07] | Indirect effect: b = .41 [13, 70] |
| Direct effect                            |          |
| b = -2.53 [-3.91, -1.15], z = -3.58, p < .001 | b = -0.89 [-2.19, 0.41], z = -1.34, p = .180 |

Table S2. Mediation analyses of Studies 1a-b.

2.4 Between-subjects analyses

As described in the main text, the designs of Studies 1a-b allow us to examine Evaluators’ responses to punishers who declined vs. chose to look in a between-subjects analysis (because Evaluators encountered these two Actor profiles first, and we randomized their order). These between-subject analyses were also pre-registered, and produce results that are similar in
direction and magnitude to the within-subject analyses we report in our main text. However, in line with the reduced statistical power they afford, they are less consistently significant. Below, we report the between-subject results for each of our dependent variables.

Among Democrats \((n = 629)\), in our between-subjects analyses, Evaluators formed more negative overall impressions of punishers who declined (vs. chose) to look \((b = -9.92 [-13.35, -6.49], t = -5.68, p < .001)\). They also shared directionally less money with punishers who declined to look, although this difference was not significant \((b = -3.44 [-7.65, .77], t = -1.60, p = .109)\). And they rated punishers who declined to look as significantly less fair \((b = -15.83 [-19.38, -12.28], t = -8.76, p < .001)\) and less competent \((b = -10.32 [-13.63, -7.01], t = -6.12, p < .001)\). In contrast, however, they rated punishers who declined to look as more loyal supporters of Black Lives Matter \((b = 7.44 [4.32, 10.56], t = 4.68, p < .001)\).

Among Republicans \((n = 600)\), our between-subjects analyses reveal no significant differences between evaluations of punishers who declined (vs. chose) to look with respect to overall positivity ratings \((b = -2.08 [-5.84, 1.67], t = -1.09, p = .276)\), money shared \((b = -2.49 [-6.80, 1.82], t = -1.14, p = .257)\) or ratings of competence \((b = -1.76 [-5.33, 1.82], t = -0.96, p = .335)\). Punishers who declined to look were, however, rated as significantly less fair \((b = -4.32 [-7.94, -7.0], t = -2.34, p = .020)\) and significantly more loyal supporters of Blue Lives Matter \((b = 10.83 [5.99, 15.66], t = 4.40, p < .001)\).

### 2.5 Analyses of evaluations of non-punishers who declined vs. chose to look

Our main text focuses on how Evaluators in Studies 1a-b evaluated punishers who declined vs. chose to look. However, as described in the main text, the designs of Studies 1a-b also allow us to investigate evaluations of non-punishers who declined vs. chose to look. Here, we report analyses of these evaluations.

Among Democrats \((n = 629)\), we find that Evaluators formed less positive overall impressions of non-punishers who declined (vs. chose) to look \((b = -13.31 [-14.97, -11.65], t = -15.78, p < .001)\), shared less money with them \((b = -7.74 [-9.12, -6.36], t = -11.02, p < .001)\), and rated them as significantly less fair \((b = -17.66 [-19.53, -15.78], t = -18.52, p < .001)\), less competent \((b = -14.38 [-16.14, -12.63], t = -16.13, p < .001)\), and less loyal \((b = -7.19 [-8.63, -5.75], t = -9.81, p < .001)\). Similarly, Republicans \((n = 600)\) rated non-punishers who declined to look as less positive overall \((b = -9.22 [-11.14, -7.30], t = -9.43, p < .001)\), shared less money with them \((b = -4.82 [-6.29, -3.34], t = -6.40, p < .001)\), and rated them as well as less fair \((b = -12.59 [-14.69, -10.48], t = -11.74, p < .001)\), less competent \((b = -10.14 [-12.05, -8.23], t = -10.44, p < .001)\), and marginally significantly less loyal \((b = -1.95 [-4.04, .14], t = -1.83, p = .068)\).

These analyses demonstrate that Evaluators created reputational incentives for Actors to consider opposing perspectives, even if they did not ultimately choose to punish. In this way, they serve to bolster our conclusion that looking can have positive reputational consequences.

### 2.6 Analyses of perfect comprehenders

As reported in the main text, our main text analyses include all subjects who passed two attention checks, regardless of performance on comprehension questions. However, per our pre-registration, here we report secondary analyses that restrict to subjects who correctly answered all comprehension questions (and produce very similar results).
2.6.1 Evaluations of punishers vs. non punishers

We begin by reporting analyses of how Evaluators who showed perfect comprehension reacted to Actors who did vs. did not punish, when given no information about whether the Actor chose to look.

Among Democrats (n = 533), Evaluators formed more positive overall impressions of punishers than non-punishers (b = 23.97 [21.25, 26.69], t = 17.31, p < .001) and also shared significantly more money with punishers (b = 15.42 [13.25, 17.59], t = 13.95, p < .001).

Furthermore, subjects rated punishers as more loyal (b = 37.80 [35.40, 40.19], t = 30.96, p < .001), fair (b = 18.92 [16.55, 21.29], t = 15.69, p < .001), and competent (b = 17.32 [15.10, 19.54], t = 15.34, p < .001). Similarly, Republicans (n = 494) evaluated punishers more positively overall (b = 22.38 [18.65, 26.12], t = 11.79, p < .001), shared more money with punishers (b = 14.29 [11.61, 16.98], t = 10.46, p < .001), and rated punishers as more loyal (b = 28.05 [24.24, 31.87], t = 14.44, p < .001), fair (b = 19.87 [16.86, 22.88], t = 12.97, p < .001), and competent (b = 18.09 [15.23, 20.95], t = 12.45, p < .001).

2.6.2 Evaluations of punishers who declined vs. chose to look

Next, we report analyses of how Evaluators who showed perfect comprehension reacted to Actors who chose to punish, after declining vs. choosing to consider opposing perspectives.

Among Democrats (n = 533), we find that Evaluators formed more negative overall impressions of punishers who declined (vs. chose) to look (b = -10.75 [-12.75, -8.75], t = -10.55, p < .001). They also shared less money with punishers who declined to look (b = -4.60 [-6.09, -3.11], t = -6.06, p < .001), and rated them as less fair (b = -16.71 [-18.95, -14.48], t = -14.69, p < .001) and less competent (b = -11.59 [-13.46, -9.72], t = -12.16, p < .001), but more loyal (b = 6.15 [4.42, 7.82], t = 6.98, p < .001). Similarly, Republicans (n = 494) rated punishers who declined to look less positively overall (b = -4.34 [-6.32, -2.34], t = -4.30, p < .001), shared less money with them (b = -1.21 [-2.82, .39], t = -1.48, p = .138) and rated them as less fair (b = -9.04 [-10.99, -7.08], t = -9.10, p < .001) and less competent (b = -4.93 [-6.75, -3.12], t = -5.33, p < .001), but more loyal (b = 7.47 [5.21, 9.72], t = 6.51, p < .001).

3. Supplemental analyses of Studies 2a-b

In this section, we report supplemental analyses of Studies 2a-b. We note that for these two studies, given the large number of conditions—and the resulting large number of potential analyses—we chose not to pre-register secondary analyses that restricted to subjects who correctly answered all comprehension questions. Therefore, we do not include “perfect comprehender” analyses in this section. However, analyses that restrict to perfect comprehenders show similar patterns, reinforcing our conclusions.

3.1 Reproducing main text Figure 2 with sharing

Below, in Figure S2, we reproduce main text Figure 2 (which plotted positivity ratings) with our sharing DV; we find qualitatively identical patterns.
Figure S2. Reproducing main text Figure 2, but with our sharing DV.

3.2 Reproducing main text Table 2 with 95% CIs

In main text Table 2, we reported analyses, for each Study 2 condition, of Evaluators’ (i) ratings of Actor loyalty, when given no information about punishment or looking, (ii) global evaluations of Actors who did vs. did not punish, and (iii) global evaluations of Actors who punished without vs. with looking. Below, in Table S3, we reproduce this table, but with 95% CIs on the regression coefficients.
Table S3. Reproducing main text Table 2, but with 95% CIs on regression coefficients.

3.3 Study 2 analyses for all DVs

Next, we provide a more complete set of Study 2 analyses for all DVs. Across a series of tables below, we report, for each Study 2 condition, analyses of Evaluators’ (i) ratings of Actors on all DVs, when given no information about punishment or looking (Table S4a), (ii) ratings of Actors who did vs. did not punish on all DVs (Table S4b), (iii) ratings of Actors who punished without vs. with looking on all DVs (Table S4c), and (iv) ratings of Actors who did not punish without vs. with looking on all DVs (Table S4d).

3.3.1 Ratings of Actors, when given no information about punishment or looking

<table>
<thead>
<tr>
<th>Democrat</th>
<th>Other participant</th>
<th>Democrat + Ignored injustice</th>
<th>Democrat + Conflict of Interest (Policy)</th>
<th>Democrat + Conflict of Interest (Psych)</th>
<th>Democrat + Previously Independent</th>
<th>Democrat</th>
<th>Democrat + Responded to Injustice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Mean = 57.39, SD = 12.67</td>
<td>Mean = 52.62, SD = 8.87</td>
<td>Mean = 57.30, SD = 13.93</td>
<td>Mean = 55.77, SD = 14.93</td>
<td>Mean = 54.88, SD = 13.72</td>
<td>Mean = 51.93, SD = 13.72</td>
<td>Mean = 58.13, SD = 13.11</td>
</tr>
<tr>
<td>t = 4.73 [5.86, 3.89]</td>
<td>t = 5.05, p &lt; .001</td>
<td>t = 0.06 [-2.22, 2.16]</td>
<td>t = 0.05, p = 0.85</td>
<td>t = 1.53 [0.66, 1.13]</td>
<td>t = 2.33 [0.41, 0.39]</td>
<td>t = 2.43 [0.34, 0.31]</td>
<td>t = 4.61, p &lt; .001</td>
</tr>
<tr>
<td>Sharing</td>
<td>Mean = 32.58, SD = 24.17</td>
<td>Mean = 28.73, SD = 24.49</td>
<td>Mean = 34.05, SD = 23.64</td>
<td>Mean = 34.74, SD = 23.89</td>
<td>Mean = 32.58, SD = 24.25</td>
<td>Mean = 30.39, SD = 24.49</td>
<td>Mean = 38.20, SD = 23.26</td>
</tr>
<tr>
<td>t = 3.79 [7.72, 2.14]</td>
<td>t = 2.37, p &lt; .001</td>
<td>t = 0.46, p = 0.65</td>
<td>t = 2.16 [1.66, 5.98]</td>
<td>t = 0.43, p = 0.65</td>
<td>t = 2.18 [0.32, 0.17]</td>
<td>t = 2.18 [0.32, 0.17]</td>
<td>t = 4.61, p &lt; .001</td>
</tr>
<tr>
<td>Loyal</td>
<td>Mean = 55.10, SD = 12.88</td>
<td>Mean = 50.12, SD = 12.90</td>
<td>Mean = 54.57, SD = 16.78</td>
<td>Mean = 50.80, SD = 14.20</td>
<td>Mean = 52.93, SD = 14.20</td>
<td>Mean = 48.23, SD = 19.01</td>
<td>Mean = 54.53, SD = 12.99</td>
</tr>
<tr>
<td>t = 4.98 [7.07, 2.90]</td>
<td>t = 4.57, p &lt; .001</td>
<td>t = 0.54 [-2.66, 1.78]</td>
<td>t = 0.43 [-2.66, 2.01]</td>
<td>t = 0.19 [0.32, 0.02]</td>
<td>t = 0.87 [0.14, 0.26]</td>
<td>t = 4.61, p &lt; .001</td>
<td>t = 4.61, p &lt; .001</td>
</tr>
<tr>
<td>Competent</td>
<td>Mean = 60.55, SD = 15.99</td>
<td>Mean = 54.26, SD = 16.42</td>
<td>Mean = 57.99, SD = 14.60</td>
<td>Mean = 55.28, SD = 14.17</td>
<td>Mean = 56.02, SD = 14.17</td>
<td>Mean = 52.85, SD = 16.93</td>
<td>Mean = 58.08, SD = 14.93</td>
</tr>
<tr>
<td>t = 4.83 [8.39, 3.46]</td>
<td>t = 4.42, p &lt; .001</td>
<td>t = 0.48, p = 0.63</td>
<td>t = 0.89 [0.31, 0.00]</td>
<td>t = 0.40, p = 0.65</td>
<td>t = 1.99, p = 0.047</td>
<td>t = 0.84, p &lt; .001</td>
<td>t = 1.75, p = 0.081</td>
</tr>
</tbody>
</table>

Table S4a. “Baseline” ratings across conditions of Study 2. We analyze “baseline” ratings of Actors on each of our DVs, when Evaluators are given no information about the Actor’s punishment or looking behavior. We report descriptive statistics within each condition (first rows) and, for each treatment, compare ratings within the relevant treatment to ratings in the baseline condition; second row; significant differences are highlighted in blue.
3.3.2 Ratings of Actors who did vs. did not punish

<table>
<thead>
<tr>
<th>Study 2a</th>
<th>Study 2b</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Democrat</strong></td>
<td><strong>Democrat</strong></td>
</tr>
<tr>
<td>Positive</td>
<td>Other participant</td>
</tr>
<tr>
<td>Positivity</td>
<td>$= -30.58$</td>
</tr>
<tr>
<td>Sharing</td>
<td>$= 24.95$</td>
</tr>
<tr>
<td>Loyal</td>
<td>$= 30.68$</td>
</tr>
<tr>
<td>Fair</td>
<td>$= 22.80$</td>
</tr>
<tr>
<td>Competent</td>
<td>$= 21.59$</td>
</tr>
</tbody>
</table>

Table S4b. Evaluations of punishment across conditions of Study 2. We analyze ratings of Actors who did vs. did not punishers vs. non-punish on each of our DVs, when Evaluators are given no information about looking. We report the effect of punishment on the relevant DV within each condition (first rows: significant effects are highlighted in grey) and, for each treatment, compare the effect of punishment within the relevant treatment vs. the baseline condition (i.e., we test for punishment X treatment interactions) (second rows; significant interactions are highlighted in blue).

3.3.3 Ratings of Actors who punished without vs. with looking

Table S4c. Evaluations of punishment without vs. with looking across conditions of Study 2. We repeat our approach from Table S4b, but analyze ratings of Actors who punished without vs. with looking. We thus report effects of not looking, conditional on punishing (and report both simple effects of not looking, and not looking X treatment interactions).
3.3.4 Ratings of Actors who did not punish without vs. with looking

<table>
<thead>
<tr>
<th>Study 2a</th>
<th>Study 2b</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positively</strong></td>
<td><strong>Positively</strong></td>
</tr>
<tr>
<td>Democrat</td>
<td>-1.15 [-1.48, -0.82]</td>
</tr>
<tr>
<td>Other participant</td>
<td>-1.24 [-1.54, -0.95]</td>
</tr>
<tr>
<td>Democrat + Ignored injustice</td>
<td>-1.10 [-1.40, -0.80]</td>
</tr>
<tr>
<td>Democrat + Conflict of Interest (Police)</td>
<td>-1.06 [-1.36, -0.77]</td>
</tr>
<tr>
<td>Democrat + Conflict of Interest (Privacy)</td>
<td>-1.31 [-1.64, -0.98]</td>
</tr>
<tr>
<td>Democrat + Previously Independent</td>
<td>-1.07 [-1.39, -0.74]</td>
</tr>
<tr>
<td>Democrat</td>
<td>-1.32 [-1.63, -1.01]</td>
</tr>
<tr>
<td>Democrat + Responded to injustice</td>
<td>-1.15 [-1.46, -0.84]</td>
</tr>
<tr>
<td><strong>Sharing</strong></td>
<td><strong>Sharing</strong></td>
</tr>
<tr>
<td>Democrat</td>
<td>-1.43 [-1.78, -1.08]</td>
</tr>
<tr>
<td>Other participant</td>
<td>-1.39 [-1.72, -1.06]</td>
</tr>
<tr>
<td>Democrat + Ignored injustice</td>
<td>-1.16 [-1.50, -0.83]</td>
</tr>
<tr>
<td>Democrat + Conflict of Interest (Police)</td>
<td>-1.03 [-1.40, -0.66]</td>
</tr>
<tr>
<td>Democrat + Conflict of Interest (Privacy)</td>
<td>-1.26 [-1.61, -0.92]</td>
</tr>
<tr>
<td>Democrat + Previously Independent</td>
<td>-1.01 [-1.39, -0.63]</td>
</tr>
<tr>
<td>Democrat</td>
<td>-1.20 [-1.53, -0.87]</td>
</tr>
<tr>
<td>Democrat + Responded to injustice</td>
<td>-1.08 [-1.46, -0.70]</td>
</tr>
<tr>
<td><strong>Loyal</strong></td>
<td><strong>Loyal</strong></td>
</tr>
<tr>
<td>Democrat</td>
<td>-1.02 [-1.34, -0.69]</td>
</tr>
<tr>
<td>Other participant</td>
<td>-1.10 [-1.42, -0.79]</td>
</tr>
<tr>
<td>Democrat + Ignored injustice</td>
<td>-0.97 [-1.31, -0.64]</td>
</tr>
<tr>
<td>Democrat + Conflict of Interest (Police)</td>
<td>-0.80 [-1.20, -0.40]</td>
</tr>
<tr>
<td>Democrat + Conflict of Interest (Privacy)</td>
<td>-1.09 [-1.43, -0.76]</td>
</tr>
<tr>
<td>Democrat + Previously Independent</td>
<td>-0.77 [-1.15, -0.40]</td>
</tr>
<tr>
<td>Democrat</td>
<td>-1.04 [-1.35, -0.74]</td>
</tr>
<tr>
<td>Democrat + Responded to injustice</td>
<td>-0.89 [-1.25, -0.53]</td>
</tr>
<tr>
<td><strong>Fair</strong></td>
<td><strong>Fair</strong></td>
</tr>
<tr>
<td>Democrat</td>
<td>-0.86 [-1.18, -0.54]</td>
</tr>
<tr>
<td>Other participant</td>
<td>-0.88 [-1.19, -0.56]</td>
</tr>
<tr>
<td>Democrat + Ignored injustice</td>
<td>-0.81 [-1.13, -0.50]</td>
</tr>
<tr>
<td>Democrat + Conflict of Interest (Police)</td>
<td>-0.66 [-0.95, -0.37]</td>
</tr>
<tr>
<td>Democrat + Conflict of Interest (Privacy)</td>
<td>-0.94 [-1.23, -0.65]</td>
</tr>
<tr>
<td>Democrat + Previously Independent</td>
<td>-0.76 [-1.05, -0.47]</td>
</tr>
<tr>
<td>Democrat</td>
<td>-0.89 [-1.20, -0.58]</td>
</tr>
<tr>
<td>Democrat + Responded to injustice</td>
<td>-0.77 [-1.06, -0.48]</td>
</tr>
<tr>
<td><strong>Competent</strong></td>
<td><strong>Competent</strong></td>
</tr>
<tr>
<td>Democrat</td>
<td>-1.44 [-1.76, -1.12]</td>
</tr>
<tr>
<td>Other participant</td>
<td>-1.38 [-1.70, -1.07]</td>
</tr>
<tr>
<td>Democrat + Ignored injustice</td>
<td>-1.13 [-1.47, -0.79]</td>
</tr>
<tr>
<td>Democrat + Conflict of Interest (Police)</td>
<td>-1.00 [-1.33, -0.68]</td>
</tr>
<tr>
<td>Democrat + Conflict of Interest (Privacy)</td>
<td>-1.24 [-1.56, -0.93]</td>
</tr>
<tr>
<td>Democrat + Previously Independent</td>
<td>-0.95 [-1.28, -0.62]</td>
</tr>
<tr>
<td>Democrat</td>
<td>-1.18 [-1.50, -0.87]</td>
</tr>
<tr>
<td>Democrat + Responded to injustice</td>
<td>-1.03 [-1.34, -0.72]</td>
</tr>
</tbody>
</table>

Table S4d. Evaluations of non-punishment vs. with looking across conditions of Study 2. We repeat our approach from Table S4c, but analyze ratings of Actors who did not punish without vs. with looking. We thus report effects of not looking, conditional on not punishing.

3.3.5 Interpretation of analyses

Together, the above tables reveal that both baseline conditions of Study 2 replicated all reported Study 1 effects on all DVs. As in Study 1, within these conditions, we find that (i) punishers are evaluated more positively than non-punishers on all DVs, (ii) punishers who decline (vs. choose) to look are evaluated less positively on all DVs except loyalty; for loyalty, punishment without looking is rated more positively, and (iii) non-punishers who decline (vs. choose) to look are evaluated less positively on all DVs.

Furthermore, the above tables provide insight into why our treatments had their effects. To highlight these insights, in the below sections we interpret the results for our treatments that sought to (i) cast doubt on loyalty by removing information, (ii) cast doubt on loyalty by adding information, and (iii) establish loyalty by adding information.

**Casting doubt on loyalty by removing information**

As reported in the main text, relative to the baseline condition, casting doubt on Actor loyalty by removing information (as we did in our “Other participant” treatment) successfully decreased “baseline” loyalty ratings (when Evaluators had no information about whether the Actor punished or looked), and increased the reputation value of punishment in general on both of our global evaluation DVs. However, it did not significantly influence the reputation value of punishment without vs. with looking on either global evaluation DV.

The above tables add further insight into these patterns. They reveal that, for all DVs (i.e., including loyalty, fairness and competence), the “Other participant” treatment significantly (i) decreased baseline ratings, (ii) increased the reputation value of punishment in general, and (iii) did not influence the reputation value of punishment without vs. with looking, or non-punishment without vs. with looking—with just one exception (the negative effect of “Other participant” on baseline sharing, which was only marginally significant). Thus, describing the Actor simply as “another participant” made Evaluators generally less positive about the Actor at baseline, and more sensitive to the question of whether or not the Actor decided to punish. However, once Evaluators knew whether the Actor punished, a lack of information about the Actor’s party affiliation did not influence Evaluators’ sensitivity to the Actor’s looking decision.
This pattern of results suggests that Evaluators may have seen the choice to punish as very informative, such that punishment information served as a substitute for information about partisanship. Under this interpretation, removing information about Actor partisanship did not change Evaluators’ relative assessments of punishment without vs. with looking, because Evaluators had punishment information (to compensate for a lack of partisanship information) when making these assessments. Thus, the informativeness of punishment may help to explain why we did not find support for the hypothesis that casting doubt on Actor loyalty by removing information can enhance the reputation value of punishment without looking.

**Casting doubt on loyalty by adding information**

As reported in the main text, relative to the baseline condition, casting doubt on loyalty by adding information (as we did in “Democrat + Ignored injustice”, both “Democrat + Conflict of Interest” treatments, and “Democrat + Previously Independent”) was largely successful at reducing baseline loyalty ratings (although “Democrat + Ignored injustice” did not have this effect). However, none of these treatments significantly (i) influenced the reputation value of punishment in general, or (ii) increased the reputation value of punishment specifically without looking, on either of our global evaluation DVs. In fact, “Democrat + Ignored injustice” and both “Democrat + Conflict of Interest” treatments caused Evaluators to share relatively less money with punishers who declined (vs. chose) to look.

The above tables add further insight into these patterns. First, they reveal that, beyond reducing baseline loyalty ratings, some of the aforementioned treatments also reduced baseline ratings of other DVs (in particular, positivity, fairness, and competence). Thus, the information that we added to cast doubt on Actors’ loyalty caused Evaluators to evaluate Actors less positively at baseline in several ways.

Second, we find that none of the aforementioned treatments enhanced the reputation value of punishment in general on any DVs. In fact, all four treatments decreased the loyalty benefit of punishment (and one treatment also decreased the competence benefit of punishment). Thus, the information that we added to cast doubt on Actors’ loyalty caused Evaluators to interpret punishment as a less meaningful signal of loyalty.

Third, we find that none of the aforementioned treatments enhanced the reputation value of punishment without (vs. with) looking on any DVs, with the one exception that “Democrat + Previously Independent” decreased the fairness cost of declining to look. Furthermore, in all four treatments, we did not observe the significant loyalty benefit of declining to look that we found in Study 1 (and replicated in the baseline conditions of Studies 2a-b). And for two treatments, this translated to a significant interaction effect, whereby the “not looking” effect size was significantly more negative in the relevant treatment than the baseline condition. Thus, we find evidence that the information that we added to cast doubt on Actors’ loyalty caused Evaluators to interpret punishment without (vs. with) looking as a less meaningful signal of loyalty.

Finally, we find that none of the aforementioned treatments influenced the reputation value of non-punishment without vs. with looking on any DVs, with the one exception that “Democrat + Previously Independent” decreased the fairness cost of declining to look.

These patterns help illuminate why casting doubt on Actor loyalty by adding information did not cause the loyalty-signaling property of punishment without looking to pay larger reputational dividends, as one might have expected it to. We instead found that, when given active reason to doubt an Actor’s loyalty, Evaluators formed more negative baseline impressions of Actors, and saw the choice to punish in general as a less meaningful indicator of loyalty.
Furthermore, Evaluators also saw the choice to punish specifically without looking as a less meaningful indicator of loyalty, perhaps because eschewing opposing perspectives seemed less like moral commitment, and more like laziness. And consequentially, casting doubt on Actor loyalty by adding information did not increase the reputation value of punishment in general, or punishment specifically without looking.

**Establishing loyalty by adding information**

As reported in the main text, relative to the baseline condition, establishing loyalty by adding information (as we did in “Democrat + Responded to injustice”) successfully increased baseline loyalty ratings, and increased the reputation value of punishment in general on both of our global evaluation DVs. However, it did not significantly influence the reputation value of punishment without vs. with looking on either global evaluation DV.

The above tables add further insight into these patterns. First, they reveal that “Democrat + Responded to injustice” increased baseline loyalty ratings, but did not increase baseline ratings of other DVs.

Second, we find that “Democrat + Responded to injustice” increased the reputation value of punishment in general on all DVs. This effect can be interpreted as a “moral consistency” effect, whereby having displayed loyalty previously (by responding to injustice) increases the importance of displaying loyalty in the present study (by punishing).

Third, we find that “Democrat + Responded to injustice” decreased the fairness cost, but also decreased the loyalty benefit, of punishing without (vs. with) looking—and did not influence relative evaluations of punishment without looking on any other DVs. Thus, the information that we added to establish loyalty caused Evaluators to interpret punishment without looking as relatively less unfair, perhaps because it seemed less hasty or uninformed coming from an individual with an established record of engagement with the relevant moral cause. But establishing loyalty also caused Evaluators to interpret punishment without looking as a relatively less meaningful signal of loyalty, perhaps because loyalty was very well-established for both types of punishers. Together, these countervailing effects seem to have netted out to create no change in the overall reputation value of punishment without vs. with looking (as measured by our global evaluation DVs).

Finally, we find that “Democrat + Responded to injustice” did not influence the reputation value of non-punishment without vs. with looking on any DVs.

Together, these patterns help illuminate why establishing Actor loyalty by adding information served to increase the reputation value of punishment in general, but not punishment specifically without looking.

**3.4 Mediation analyses**

As noted in the main text, the baseline conditions of Studies 2a-b replicate the results of our exploratory mediation analyses from Studies 1a-b (which were pre-registered analyses for the baseline conditions of Study 2). In Table S5, we reproduce Table S2 but with data from the baseline conditions of Study 2; we find qualitatively identical patterns.
Table S5. Mediation analyses of Studies 2a-b. We reproduce Table S2, but with the “baseline” conditions of Studies 2a-b.

We also explored mediation within the treatment conditions of Studies 2a-b. In particular, within each treatment condition, we ran the same two mediation models that we report results for above. The results of these analyses, which we do not report here for brevity, were qualitatively identical to the above-reported results, with one key exception. As illustrated in Table S4a, the four Study 2a treatments that were designed to cast doubt on loyalty by adding information (i.e., “Democrat + Ignored injustice”, both “Democrat + Conflict of Interest” treatments, and “Democrat + Previously Independent”) did not produce significant positive effects of punishment without (vs. with) looking on loyalty. Therefore, within each of these conditions, in both models, the A path and indirect effects for loyalty are not significant.

3.5 Moderation analyses

As noted in the main text, Study 2 provides further evidence, both within the baseline conditions and overall, for the moderating effects of Evaluator ideology (whereby strong partisans show relatively more positive evaluations of punishment in general, and less negative evaluations of punishment without looking).

To support this claim, in Table S6, we reproduce main text Table 2, but with data from the baseline conditions of Studies 2a-b. This table reveals significant partisanship moderation in the baseline condition of Study 2a for overall positivity (both for evaluations of punishment, and evaluations of punishment without vs. with looking) and sharing (for evaluations of punishment,
but not evaluations of punishment without vs. with looking). In the baseline condition of Study 2b, however, we do not find significant moderation.

<table>
<thead>
<tr>
<th>Study 2a (Democrats)</th>
<th>Evaluations of punishers vs. non-punishers (Positive coefficients reflect preferences for punishment)</th>
<th>Evaluations of punishment without vs. with looking (Positive coefficients reflect preferences for punishment without looking)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positivity</strong></td>
<td>Weak Partisans: (b = 22.55 \ [16.72, 28.38], t = 7.67, p &lt; .001)</td>
<td>Weak Partisans: (b = -15.66 \ [-19.63, -11.69], t = -7.82, p &lt; .001)</td>
</tr>
<tr>
<td></td>
<td>Strong Partisans: (b = 35.41 \ [30.45, 40.36], t = 14.10, p &lt; .001)</td>
<td>Strong Partisans: (b = -7.34 \ [-10.99, -3.68], t = -3.96, p &lt; .001)</td>
</tr>
<tr>
<td></td>
<td>Interaction: (b = 12.86 \ [5.26, 20.46], t = 3.33, p = .001)</td>
<td>Interaction: (b = 8.32 \ [2.96, 13.68], t = 3.06, p = .002)</td>
</tr>
<tr>
<td><strong>Sharing</strong></td>
<td>Weak Partisans: (b = 14.99 \ [9.63, 19.56], t = 5.82, p &lt; .001)</td>
<td>Weak Partisans: (b = -5.32 \ [-8.91, -1.72], t = -2.93, p = .004)</td>
</tr>
<tr>
<td></td>
<td>Strong Partisans: (b = 21.36 \ [17.49, 25.23], t = 10.88, p &lt; .001)</td>
<td>Strong Partisans: (b = -2.41 \ [-6.53, 0.71], t = -1.52, p = .129)</td>
</tr>
<tr>
<td></td>
<td>Interaction: (b = 6.77 \ [0.51, 13.02], t = 2.13, p = .034)</td>
<td>Interaction: (b = 2.91 \ [-1.82, 7.63], t = 1.21, p = .227)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study 2b (Democrats)</th>
<th>Evaluations of punishers vs. non-punishers (Positive coefficients reflect preferences for punishment)</th>
<th>Evaluations of punishment without vs. with looking (Positive coefficients reflect preferences for punishment without looking)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positivity</strong></td>
<td>Weak Partisans: (b = 21.90 \ [17.30, 26.51], t = 9.39, p &lt; .001)</td>
<td>Weak Partisans: (b = -15.47 \ [-19.37, -11.58], t = -7.85, p &lt; .001)</td>
</tr>
<tr>
<td></td>
<td>Strong Partisans: (b = 24.86 \ [19.18, 30.55], t = 8.64, p &lt; .001)</td>
<td>Strong Partisans: (b = -10.89 \ [-15.48, -6.30], t = -4.69, p &lt; .001)</td>
</tr>
<tr>
<td></td>
<td>Interaction: (b = 2.96 \ [4.32, 10.24], t = 0.80, p = .424)</td>
<td>Interaction: (b = 4.58 \ [-1.41, 10.57], t = 1.51, p = .133)</td>
</tr>
<tr>
<td><strong>Sharing</strong></td>
<td>Weak Partisans: (b = 12.63 \ [8.79, 16.47], t = 6.50, p &lt; .001)</td>
<td>Weak Partisans: (b = -8.22 \ [-11.24, -5.20], t = -5.38, p &lt; .001)</td>
</tr>
<tr>
<td></td>
<td>Strong Partisans: (b = 14.86 \ [10.17, 19.55], t = 6.26, p &lt; .001)</td>
<td>Strong Partisans: (b = -9.07 \ [-13.01, -5.14], t = -4.56, p &lt; .001)</td>
</tr>
<tr>
<td></td>
<td>Interaction: (b = 2.23 \ [-3.80, 8.25], t = 0.73, p = .468)</td>
<td>Interaction: (b = 0.85 \ [-5.78, 4.09], t = -0.34, p = .735)</td>
</tr>
</tbody>
</table>

Table S6. The moderating role of ideology in Studies 2a-b. We reproduce Table 2 from the main text, but with Studies 2a-b.

Then, in Figures S3a-b, we reproduce Figure 2 from the main text, but separately show results among strong vs. weak partisans—both for positivity (Figure S3a) and sharing (Figure S3b). These figures provide insight into moderation patterns across conditions. For both DVs, across conditions, we see that strong partisans generally show more positive evaluations of punishment in general, and less negative evaluations of punishment without looking. Notably, however, across all conditions, even strong partisans consistently react more negatively to punishers who decline than choose to look, on both DVs.
Figure S3a. The moderating role of ideology on positivity ratings in Studies 2a-b. We reproduce Figure 2 from the main text, but separately show results among strong vs. weak partisans.

Figure S3b. The moderating role of ideology on sharing in Studies 2a-b. We reproduce Figure 2 from the main text, but separately show results among strong vs. weak partisans, and use our sharing DV.
4. Supplemental Analyses of Studies 3-4

4.1 Analyses of Study 4a by batch, and corrections for “peeking”

As described in the main text, for Study 4a, we pre-registered and recruited an initial target \( n = 1200 \) subjects. Upon reaching this initial target, all of our key results were significant. However, while finishing data collection for Study 4b (which recruited Republicans, who are under-represented on MTurk), we decided to direct Democrats to Study 4a, increasing its sample size. After making this decision, we registered an amendment to our Study 4a pre-registration, increasing our target to \( n = 2000 \) subjects. In the main text, we thus report analyses of Study 4a that combine data from both batches of data collection. However, in Table S7, we report our Study 4a results within each individual batch (and also re-report the overall results from both batches combined, to facilitate comparison).

We also pre-registered a plan to correct, in our combined analyses of both batches, for the fact that we “peeked” at first batch of data before deciding to collect a second batch (using the approach of Sagarin, Ambler, & Lee, 2014, Perspectives on Psychological Science). In particular, we planned, for each key result that is significant in our final sample, to report whether the result continues to be significant when accounting for peeking by computing an adjusted alpha threshold that allows us to maintain an actual type-I error rate of .05.

More precisely, instead of computing a single adjusted alpha threshold, we planned to report an adjusted alpha range. This reflects that the required alpha threshold depends on the maximum p-value observed in the initial batch of data collection for which we would have collected more data rather than declaring the initial results non-significant; this could range from a “best-case scenario” of the p-value observed after the initial data collection to a “worst-case scenario” of 1.

Thus, for each result in Table S7, we report a best- and worst-case scenario for the adjusted alpha threshold. Importantly, for each of the three results that are significant in the overall sample, we find that the p-values are smaller than even the worst-case scenario adjusted alpha threshold. Thus, we find that all significant Study 4a results remain significant, even after accounting for peeking.

<table>
<thead>
<tr>
<th>Batch 1 ((n = 1222))</th>
<th>Batch 2 ((n = 752))</th>
<th>Overall ((n = 1974))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Punishment overall</td>
<td>Punishment without looking</td>
<td>Looking overall</td>
</tr>
<tr>
<td>( b = .09 ) ([.03, .15])</td>
<td>( b = .05 ) ([.02, .09])</td>
<td>( b = .12 ) ([.06, .19])</td>
</tr>
<tr>
<td>( t = 3.03, p &lt; .003, n = 815)</td>
<td>( t = 2.05, p = .040, n = 815)</td>
<td>( t = 3.56, p &lt; .001, n = 797)</td>
</tr>
<tr>
<td>( b = .16 ) ([.09, .24])</td>
<td>( b = .07 ) ([.01, .13])</td>
<td>( b = -.01 ) ([-0.09, .08])</td>
</tr>
<tr>
<td>( t = 4.18, p &lt; .001, n = 504)</td>
<td>( t = 2.45, p = .015, n = 504)</td>
<td>( t = 1.11, p = .911, n = 487)</td>
</tr>
<tr>
<td>Overall Adjusted Alpha Threshold: Best case</td>
<td>( b = .12 ) ([0.07, .16])</td>
<td>( b = .06 ) ([0.02, .09])</td>
</tr>
<tr>
<td>Adjusted Alpha Threshold: Worst case</td>
<td>( t = 4.98, p &lt; .001, n = 1319)</td>
<td>( t = 3.13, p = .002, n = 1319)</td>
</tr>
</tbody>
</table>

Table S7. Analyses of Study 4a by batch, and adjusted Alpha Thresholds that account for “peeking” at data between batches.

4.2 Analyses of how making punishment observable influences looking

In our main text analyses of Studies 3-4, when considering the effects of making punishment observable (i.e., when comparing our “Nothing Observable” and “Punishment Observable” conditions), we focus on rates of punishment overall, and punishment without looking, as dependent variables. However, our design also allows us to ask how making punishment observable influences overall rates of looking. Indeed, we pre-registered analyses of
this question, but do not report them in the main text because they are less relevant to our key theoretical questions. Here, however, we report these analyses.

In Study 3, among Democrats, rates of looking were similar in Punishment Observable (35%) and Nothing Observable (34%), \( b = .01 [-.04, .06], t = .40, p = .686, n = 1222. \) Republicans were also similarly likely to look in Punishment Observable (24%) and Nothing Observable (28%), \( b = -.04 [-.09, .01], t = -1.45, p = .148, n = 1214. \)

In Study 4, among Democrats, rates of looking were significantly higher in Punishment Observable (41%) than in Nothing Observable (34%), \( b = .07 [.02, .12], t = 2.73, p = .006, n = 1319. \) In contrast, Republicans were similarly likely to look in Punishment Observable (15%) and Nothing Observable (18%), \( b = -.03 [-.08, .02], t = -1.08, p = .279, n = 763. \)

Thus, we mostly find that making punishment observable has no reliable effect on rates of looking. However, in one case (Democrats in Study 3), making punishment observable increased looking.

We note that, in the discussion section of our main text, we raise the question of whether potential punishers sometimes decline to look because they (i) anticipate a reputational upside to punishing and (ii) therefore wish to avoid being persuaded against punishing by opposing perspectives. At face value, the above-reported analyses seem to provide evidence against this suggestion: creating reputational incentives to punish (by making punishment observable) did not make subjects in our studies significantly less likely to engage with opposing perspectives.

That being said, the above-described logic assumes that potential punishers, before engaging with opposing perspectives, expect such engagement to, with some meaningful probability, decrease their propensity to punish (i.e., they expect the opposing perspectives to be potentially persuasive). If potential punishers do not hold this expectation, a motivation to support punishment should not deter the consideration of opposing perspectives. And if potential punishers hold the opposite expectation (i.e., they expect the opposing perspectives to be unpersuasive), a motivation to support punishment could actually encourage looking.

In our studies, before deciding whether or not to look (i.e., click the links to read the opposing perspective articles), subjects saw headlines for these articles. And because we did not collect any data about subjects’ expectations, on the basis of these headlines, about the persuasiveness of the opposing perspective articles, it is unclear what expectations subjects held. Thus, while our data do not support the proposal that reputational incentives for punishment can discourage the consideration of opposing perspectives, we do not see them as providing a definitive refutation of this hypothesis. It is possible that making punishment observable would have decreased looking if subjects had expected the opposing perspective articles to be more persuasive.

4.3 Analyses of how making looking observable influences punishment

In our main text analyses of Studies 3-4, when considering the effects of making looking observable (i.e., when comparing our “Punishment Observable” and “Both Observable” conditions), we focus on rates of looking overall, and punishment without looking, as dependent measures. However, our design also allows us to ask how making looking observable influences overall rates of punishment. We likewise pre-registered analyses of this question, but do not report them in the main text because they are less relevant to our key theoretical questions. Here, however, we report these analyses.

In Study 3, among Democrats, rates of punishment were similar in Both Observable (24%) and Punishment Observable (23%), \( b = .01 [-.04, .05], t = .27, p = .786, n = 1206. \)
Republicans were also similarly likely to punish in Both Observable (28%) and Punishment Observable (27%), $b = .02 [-.03, .07]$, $t = .72$, $p = .474$, $n = 1197$.

In Study 4, among Democrats, rates of punishment were similar in Both Observable (28%) and Punishment Observable (30%), $b = -.02 [-.07, .03]$, $t = -.77$, $p = .439$, $n = 1284$. Republicans were also similarly likely to punish in Both Observable (25%) and Punishment Observable (25%), $b = -.003 [-.06, .06]$, $t = -.10$, $p = .921$, $n = 779$.

Thus, making looking observable did not influence rates of punishment in our studies.

### 4.4 Analyses of perfect comprehenders

As reported in the main text, our main text analyses include all subjects, regardless of performance on comprehension questions. However, per our pre-registration, below we report secondary analyses that restrict to subjects who correctly answered our comprehension questions on their first try (and produce very similar results).

We note that in Studies 3-4, subjects answered two sets of comprehension questions. The first three questions tested comprehension of the Dictator Game structure, and the second three questions tested comprehension of the extent to which subjects’ punishment and looking decisions were observable to the Evaluator. Importantly, in the below analyses, rather than restricting to subjects who answered all questions correctly on their first try, we merely restrict to subjects who correctly answered the Dictator Game questions on the first try. This decision reflects that in both Study 3 and Study 4, initial performance on each of the three observability questions differed significantly across observability conditions; thus, restricting to subjects who correctly answered all observability questions on their first try would undermine random assignment and thus causal inference.

In particular, in Study 3, a series of chi squared tests reveals significant condition effects on binary variables indicating whether the first (Democrats: $X^2 = 30.37$, $p < .001$; Republicans: $X^2 = 34.10$, $p < .001$), second (Democrats: $X^2 = 174.58$, $p < .001$; Republicans: $X^2 = 178.69$, $p < .001$), and third (Democrats: $X^2 = 6.81$, $p = .033$; Republicans: $X^2 = 13.68$, $p = .001$) observability questions were initially answered correctly. Similarly, in Study 4, condition significantly predicted initial performance on the first (Democrats: $X^2 = 92.94$, $p < .001$; Republicans: $X^2 = 38.76$, $p < .001$), second (Democrats: $X^2 = 204.96$, $p < .001$; Republicans: $X^2 = 146.42$, $p < .001$), and third (Democrats: $X^2 = 13.54$, $p = .001$; Republicans: $X^2 = 6.83$, $p = .033$) observability questions.

### 4.4.1 Effects of making punishment observable

First, we report effects of making punishment observable (i.e., comparisons between our Nothing Observable and Punishment Observable conditions) among perfect comprehenders.

In Study 3, among Democrats, overall rates of punishment were significantly higher in Punishment Observable (24%) than Nothing Observable (18%), $b = .07 [.02, .12]$, $t = 2.83$, $p = .005$, $n = 1080$. However, rates of punishment without looking did not significantly differ across the Punishment Observable (13%) and Nothing Observable (10%) conditions, $b = .03 [-.01, .07]$, $t = 1.53$, $p = .127$, $n = 1080$. Among Republicans, overall rates of punishment were marginally higher in Punishment Observable (28%) than Nothing Observable (22%), $b = .05 [-.01, .11]$, $t = 1.91$, $p = .056$, $n = 1022$. Furthermore, Republicans were significantly more likely to punish without looking in Punishment Observable (20%) than in Nothing Observable (14%), $b = .06 [.02, .11]$, $t = 2.65$, $p = .008$, $n = 1022$. 
In Study 4, among Democrats, overall rates of punishment were significantly higher in Punishment Observable (31%) than Nothing Observable (18%), \( b = .13 \ [0.08, 0.18], t = 5.22, p < 0.001, n = 1172 \). Moreover, we find significantly higher rates of punishment without looking in Punishment Observable (16%) than Nothing Observable (9%), \( b = .07 \ [0.03, 0.11], t = 3.50, p < 0.001, n = 1172 \). Similarly, Republicans were more likely to punish in Punishment Observable (26%) than Nothing Observable (16%), \( b = .10 \ [0.03, 0.16], t = 3.02, p = .003, n = 659 \). And they were also more likely to punish without looking in Punishment Observable (20%) than in Nothing Observable (13%), \( b = .08 \ [0.02, 0.13], t = 2.59, p = .010, n = 659 \).

4.4.2 Effects of making looking observable

Next, we examine the effects of making looking observable (i.e., comparisons between our Punishment Observable and Both Observable conditions) among perfect comprehenders. In Study 3, among Democrats, overall rates of looking were significantly higher in Both Observable (52%) than Punishment Observable (36%), \( b = .16 \ [0.10, 0.22], t = 5.41, p < 0.001, n = 1056 \). Furthermore, rates of punishment without looking were significantly lower in Both Observable (8%) than Punishment Observable (13%), \( b = -0.05 \ [-0.09, -0.01], t = -2.64, p = 0.009, n = 1056 \). Among Republicans, we likewise observed higher rates of looking in Both Observable (39%) than Punishment Observable (24%), \( b = .15 \ [0.09, 0.21], t = 5.13, p < 0.001, n = 986 \). However, we observed no significant difference between rates of punishment without looking in Both Observable (16%) and Punishment Observable (20%), \( b = -0.04 \ [-0.09, 0.01], t = -1.57, p = 0.116, n = 986 \).

In Study 4, among Democrats, overall rates of looking were significantly higher in Both Observable (50%) than in Punishment Observable (42%), \( b = .07 \ [0.02, 0.13], t = 2.52, p = .012, n = 1137 \). However, we found no significant difference between rates of punishment without looking in Both Observable (13%) versus Punishment Observable (16%), \( b = -0.03 \ [-0.07, 0.01], t = -1.47, p = .142, n = 1137 \). Among Republicans, we likewise saw higher rates of looking in Both Observable (30%) than Punishment Observable (16%), \( b = .14 \ [0.08, 0.20], t = 4.36, p < 0.001, n = 676 \). However, we found no significant difference in rates of punishment without looking between Both Observable (20%) and Punishment Observable (20%), \( b = -0.001 \ [-0.06, 0.06], t = -0.03, p = .972, n = 676 \).

4.5 Analyses of alternative specifications of looking

As described in the main text, per our pre-registration, our primary analyses define “looking” in Studies 3-4 as clicking at least one link to an opposing perspective article. However, for Study 4 (but not Study 3), we also preregistered secondary analyses that define looking as either (i) time spent on the “looking” screen (in which subjects were presented with links to opposing perspective articles, and invited to search the Internet for other opposing perspectives) or (ii) the continuous number of opposing perspective article links that subjects clicked. We report these alternative analyses, which produce very similar results, below.

4.5.1 Looking time (natural-log transformed seconds)

We begin by defining looking in terms of time spent on the “looking” screen. Per our pre-registration, we specifically consider the number of natural-log transformed seconds spent on this screen.

First, we use this new definition of looking to analyze overall rates of looking. In particular, we report the effect of making looking observable on overall rates of looking. In
Study 3, among Democrats, we find that subjects spent more time looking in Both Observable ($M = 3.66$) than Punishment Observable ($M = 3.16$), $b = .50 [.35, .65], t = 6.39, p < .001$. Similarly, Republicans subjects spent more time looking in Both Observable ($M = 3.03$) than Punishment Observable ($M = 3.41$) than Punishment Observable ($M = 3.16$), $b = .34 [.20, .47], t = 4.83, p < .001$. In Study 4, among Democrats, subjects spent more time looking in Both Observable ($M = 3.11$) than Punishment Observable ($M = 2.71$), $b = .39 [.24, .54], t = 5.02, p < .001$.

Next, we use our new definition of looking to redefine punishment without looking as punishing after looking for a below-median amount of time. In doing so, we compute the median (and thus define subjects as being above vs. below the median) separately for Democrats and Republicans. Then, we report the effects of making (i) punishment observable and (ii) looking observable, on punishment without looking.

In Study 3, among Democrats, subjects punished without looking at comparable rates in Punishment Observable (.10) and Nothing Observable (.10), $b = .002 [-.03, .04], t = 1.12, p = .267$, and comparable rates in Both Observable (.08) and Punishment Observable (.10), $b = -.02 [-.06, .01], t = -1.42, p = .157$. Among Republicans, subjects punished without looking at higher rates in Punishment Observable (.14) than Nothing Observable (.10), $b = .05 [.01, .08], t = 2.56, p = .011$, and lower rates in Both Observable (.10) than Punishment Observable (.14), $b = -.04 [-.08, -.01], t = -2.28, p = .023$.

In Study 4, among Democrats, subjects punished without looking at higher rates in Punishment Observable (.14) than Nothing Observable (.08), $b = .06 [.02, .09], t = 3.26, p = .001$, and comparable rates in Both Observable (.11) and Punishment Observable (.14), $b = -.03 [-.06, .01], t = -1.37, p = .170$. Similarly, Republican subjects punished without looking at higher rates in Punishment Observable (.13) than Nothing Observable (.06), $b = .07 [.03, .11], t = 3.38, p = .001$, and comparable rates in Both Observable (.12) and Punishment Observable (.13), $b = -.01 [-.06, .04], t = -.45, p = .651$.

4.5.2 Number of articles

Next, we define looking as the continuous number of opposing perspective articles that subjects clicked the link for. When defining looking this way, in Study 3, we find that Democrats read a larger number of articles in Both Observable ($M = .67$) than Punishment Observable ($M = .44$), $b = .24 [.16, .31], t = 5.96, p < .001$. Similarly, Republicans read a larger number of articles in Both Observable ($M = .61$) than Punishment Observable ($M = .38$), $b = .23 [.15, .32], t = 5.23, p < .001$. In Study 4, among Democrats, subjects read a larger number of articles in Both Observable ($M = .67$) than Punishment Observable ($M = .54$), $b = .13 [.05, .21], t = 3.12, p = .002$. Similarly, Republicans subjects read a larger number of articles in Both Observable ($M = .44$) than Punishment Observable ($M = .20$), $b = .24 [.15, .33], t = 5.12, p < .001$.

We do not use this definition of looking to redefine punishment without looking. This reflects that, in all studies, less than half of subjects read even one article; thus, if we were to mirror our above approach and redefine punishment without looking as punishing after looking at a below-median number of articles, we would wind up with the same definition of punishment without looking that we already used in our primary analyses.
5. Discussion of preregistration deviations

All studies were pre-registered (Study 1a: https://aspredicted.org/blind.php?x=RB5_VT1; Study 1b: https://aspredicted.org/522_B9Z; Study 2a: https://aspredicted.org/43C_4DD; Study 2b: https://aspredicted.org/1B8_B91; Study 3a: https://aspredicted.org/blind.php?x=WM4_C5K; Study 3b: https://aspredicted.org/XX6_24X; Study 4a: https://aspredicted.org/2HX_QQJ; Study 4b: https://aspredicted.org/XRC_RY3). As described in the main text, our analyses closely to our preregistered analysis plans with some minor deviations. Here, we describe these deviations.

5.1 Studies 1-2

We begin by describing deviations in our analyses of our Evaluator studies (i.e., Studies 1-2). We did not deviate from our pre-registrations for Studies 2a-b; thus, this section outlines deviations in our analyses of Studies 1a-b.

First, our Study 1a-b preregistrations describe our between-subject analyses (of evaluations of punishers who did vs. did not look) as primary analyses. However, for reasons of brevity and because these analyses produce results that are similar to our within-subject analyses but afford less statistical power, we report them only in the SI and not in the main text.

Second, our preregistrations describe our analyses of money sent in the Dictator Game as primary. In contrast, analyses of all other dependent variables (i.e., ratings of overall positivity, fairness, competence, and loyalty) are described in our pre-registrations as secondary analyses. Yet our main text does not preferentially focus on money sent. Instead, we report analyses of money sent and also ratings of overall positivity, fairness, competence, and loyalty, and specifically plot positivity ratings in main text Figure 1.

To explain this decision, we begin by noting that we saw money sent in the Dictator Game and ratings of overall positivity and as “global” evaluation variables (i.e., as variables that reflected Evaluators’ overall impressions of Actors). In contrast, we saw ratings of fairness, competence, and loyalty as more specific evaluation variables. Originally, we planned to privilege money sent as our “primary” global evaluation variable. However, upon seeing our data, we noticed that many Evaluators chose either to share none or to exactly half of their endowment, and almost no Evaluators shared more than half of their endowment. This distributional feature likely reflects that money shared in the Dictator Game tapped both (i) the extent to which the Evaluator had a positive impression of the Actor and (ii) the extent to which the Evaluator was selfish vs. generous, or committed to avoiding inequality, etc. For example, Evaluators who were motivated to maximize their own payoffs may have never shared any money with any Actors, regardless of how positively they regarded them. And Evaluators who were committed to fairness may have always shared half of their money with all Actors, regardless of how positively they regarded them.

In contrast, ratings of overall positivity were more continuously distributed, perhaps reflecting that this global evaluation variable was a “purer” reflection of positive regard for the Actor. Thus, we ultimately felt that overall positivity was the more informative global evaluation variable. Moreover, we also ultimately felt that analyses of perceived fairness, competence, and loyalty provided critical insight into why Evaluators formed the global evaluations of Actors that they did. For these reasons, we report analyses of all DVs in our main text, and plot overall positivity in Figure 1 (while reproducing this figure with sharing in SI Figure S1). Furthermore, in Studies 2a-b, we pre-registered positivity ratings our as our primary DV.

Third, our pre-registrations describe our analyses of Evaluations of punishers who do vs. do not look as primary, but our analyses of Evaluations of punishers vs. non-punishers (in the
absence of looking information) as secondary. However, our main text prominently features both sets of analyses. This decision reflects that, while our primary theoretical focus was on evaluations of punishers who did vs. did not look, we felt that understanding whether punishment was evaluated positively in general was critical for interpreting Evaluators’ preferences for punishment with vs. without looking.

Fourth, we report mediation analyses that were not pre-registered. While exploratory, we felt that these analyses were informative (and pre-registered them in Studies 2a-b).

5.3 Studies 3-4

Next, we describe deviations in our analyses of our Actor studies (i.e., Studies 3-4). First, our Study 3-4 preregistrations describe as primary two analyses that we ultimately report only in the SI (and not in our main text): (i) analyses of the effect of making punishment observable on looking, and (ii) analyses of the effect of making looking observable on punishment. As noted in the main text, we ultimately saw these analyses as less central to our theoretical questions; we therefore felt that they would distract from the flow of the main text and obfuscate the relationship between our analyses and the hypotheses we sought to test.

Second, for Study 4 but not Study 3, we preregistered a plan to report secondary analyses that used alternative definitions of looking. However, for completeness, in this SI we report these secondary analyses for both Study 4 and Study 3.

6. Appendix: Experimental stimuli

Here, we describe and illustrate with screenshots our experimental stimuli for all studies. Additionally, on OSF (https://osf.io/3es2k/?view_only=e272c31aa7304eca841f084d582187fb), we provide .qsf and PDF export files of the Qualtrics surveys used for each study, as well as PDFs containing the full texts of all opposing perspective articles (that subjects in Studies 3-4 viewed if they clicked the relevant links).

6.1 Studies 1a-b

After reporting their Prolific IDs and providing informed consent, subjects began by reporting their age, gender, and political party affiliation, and answering two attention check questions:
Age:

Which gender do you identify with more closely?

- Male
- Female
- Non-binary / other identity

Do you think of yourself as more of a Democrat or a Republican?

If you instead consider yourself an Independent, please indicate whether, as of today, you lean more Democratic or Republican.

<table>
<thead>
<tr>
<th>Democrat</th>
<th>Republican</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Thank you. Please carefully read the following story about a woman named Sarah.

Sarah works at a local grocery store. At the store, Sarah's job is to serve as the cashier. Normally, Sarah works Monday-Friday but does not work weekends. However, last week Sarah's coworker Ben asked her to cover his Saturday shift. So this Saturday, Sarah has to work a 7-hour shift.

What is Sarah's job at the grocery store?

- Manager
- Cashier
- Stocker
- Customer service
- It was not specified in the story

The meal pictured above consists of multiple food items. From the options below, please select the food that is probably not one of these food items.

- ketchup
- onion
- mustard
- bread
- olives
- relish
- hotdog
Next, we introduced the Dictator Game (described as the “Sharing Game”), informed subjects that they would be interacting with another participant (described as “Player 2”) who shares their political party, and presented a set of comprehension questions about the game:

In this study, you will also participate in an interactive game with another participant.

The game is called the Sharing Game, and it has two players: Player 1 and Player 2. You will be Player 1.

Another participant will be Player 2. Like you, this participant ALSO identifies as a Democrat. And, just like we just told you that Player 2 is a Democrat, we will also tell Player 2 that you are a Democrat.

In the Sharing Game, you (Player 1) start with 50 cents. You then choose how many cents, if any, to share with Player 2.

Please answer the following questions, to make sure you understand the Sharing Game.

Imagine that you are deciding how much to share with Player 2.

Which decision will result in **you** earning the most money?

- You deciding to share 0 cents
- You deciding to share 25 cents
- You deciding to share 50 cents

Imagine that you are deciding how much to share with Player 2.

Which decision will result in **Player 2** earning the most money?

- You deciding to share 0 cents
- You deciding to share 25 cents
- You deciding to share 50 cents

What political affiliation is Player 2?

- Player 2 is a Republican
- Player 2 is a Democrat

Note: the above screenshot shows the screen for Democrats in Study 1a. For Republicans in Study 1b, references to Democrats were replaced with references to Republicans.

Next, we explained that Player 2 signed a petition, and described the petition. Below we show how this looked for each of the three petitions. Recall that in Study 1a, Democrats were randomly assigned to either the Moore or Negy petition; in Study 1b, Republicans were always assigned to the Amazon petition.
Democrats, Moore Petition:

On this page, we would like you to tell you a bit more about Player 2.

We already recruited Player 2 to complete a study, which had an additional component.

In addition to participating in the Sharing Game, Player 2 also had the opportunity to sign a petition, hosted on Change.org.

The petition calls for the LAPD to fire Chief Michael Moore. It argues that Chief Moore should be fired following comments he made that blamed protestors for George Floyd's death.

On the next page, we will provide you with the full text of the petition and ask you to read it.

We also showed this full text to Player 2, who then decided whether or not to help get Chief Moore fired by signing the petition.

Democrats, Negy Petition:

On this page, we would like you to tell you a bit more about Player 2.

We already recruited Player 2 to complete a study, which had an additional component.

In addition to participating in the Sharing Game, Player 2 also had the opportunity to sign a petition, hosted on Change.org.

The petition calls for the University of Central Florida to fire professor Charles Negy. The petition alleges that professor Negy should be fired on the basis of abhorrent racist comments he has made on his personal Twitter account.

On the next page, we will provide you with the full text of the petition and ask you to read it.

We also showed this full text to Player 2, who then decided whether or not to help get professor Negy fired by signing the petition.

Republicans, Amazon Petition:

On this page, we would like you to tell you a bit more about Player 2.

We already recruited Player 2 to complete a study, which had an additional component.

In addition to participating in the Sharing Game, Player 2 also had the opportunity to sign a petition, hosted on Change.org.

The petition calls for the remove of "Blue Lives Murder" merchandise from Amazon. It alleges that the merchandise is hateful and must be immediately removed.

On the next page, we will provide you with the full text of the petition and ask you to read it.

We also showed this full text to Player 2, who then decided whether or not to help remove "Blue Lives Murder" merchandise from Amazon by signing the petition.
Next, we presented a screenshot of the petition.

Democrats, Moore Petition:

Below is a screenshot of the petition that we showed to Player 2. Over 300,000 people have already signed the petition, as highlighted in the screenshot below.

Again, the petition calls for the LAPD to fire Chief Michael Moore. It argues that Chief Moore should be fired following comments he made that blamed protestors for George Floyd’s death.

On this screen, please read through the petition.

LAPD is a corrupt department who claims the peaceful protests all around Los Angeles are civil unrests and criminal acts. This man should be fired.

On June 1st, Moore stated that George Floyd’s death is on protestors and looters hands. He is belittling the struggles and ways of the Black Lives Matter movement. This is only the shadow of an underlying racist and discriminating department.

Let LAPD Police Chief Michael Moore know how you feel about his blaming protestors for George Floyd’s death.
Democrats, Negy Petition:

Below is a screenshot of the petition that we showed to Player 2. Over 30,000 people have already signed the petition, as highlighted in the screenshot below.

Again, the petition calls for the University of Central Florida to fire professor Charles Negy. It alleges that professor Negy should be fired on the basis of abhorrent racist comments he has made on his personal Twitter account.

On this screen, please read through the petition.

We are calling on the University of Central Florida to dismiss psychology professor Charles Negy due to abhorrent racist comments he has made and continue to make on his personal Twitter account. In addition to racism, Negy has engaged in perverse transphobia and sexism on his account, which is just as reprehensible. While he has a right to free speech, he does not have a right to dehumanize students of color and other minority groups, which is a regular occurrence in his classroom. By allowing him to continue in his position, UCF would simply be empowering another cog in the machine of systemic racism.

UCF is a diverse, welcoming campus of 60,000 students from all across the country and the world. Our diversity is our strength, and we should have faculty who understand and promote that— not the opposite. College is a place to learn and be exposed to new ideas, but for so many Black students, racism is not just an idea, but something they have experienced for their entire life, and the same goes for the LGBT community with trans and homophobia. They deserve better, and so does the entire campus community.

I once again want to reiterate my love for this university. We are a community which always has each other's back. As a student coming from a place of privilege, this is me having the backs of my fellow Knights.

Please dismiss Professor Negy for the good of the entire campus community.
Republicans, Amazon Petition:

Below is a screenshot of the petition that we showed to Player 2. Over 115,000 people have already signed the petition, as highlighted in the screenshot below.

Again, the petition calls for the removal of “Blue Lives Murder” merchandise from Amazon. It argues that the merchandise is hateful and must be immediately removed.

On this screen, please read through the petition.

![Petition](image-url)
Next, we explained to subjects that Player 2 had the chance to “look” at opposing perspectives, and show an example headline from an opposing perspective article.

Democrats, Moore petition:

After viewing this petition, Player 2 decided whether or not to sign it.

Additionally, we gave Player 2 an opportunity to take some time—if Player 2 wanted it—to consider OPPOSING perspectives before deciding whether to sign. It was completely up to Player 2 whether to use this time, and how.

For Player 2's convenience, we provided links to some specific articles that may provide opposing perspectives pertaining to the petition. For example, we provided a link to the below article published by the LA Times describing how elected politicians in Los Angeles, including the mayor, continue to support Police Chief Moore.

Player 2 also had the option of taking time to search the Internet for other opposing perspectives.

And, importantly, Player 2 also had the option NOT to read any opposing perspectives before deciding whether or not to sign the petition.

Democrats, Negy petition:

After viewing this petition, Player 2 decided whether or not to sign it.

Additionally, we gave Player 2 an opportunity to take some time—if Player 2 wanted it—to consider OPPOSING perspectives before deciding whether to sign. It was completely up to Player 2 whether to use this time, and how.

For Player 2's convenience, we provided links to some specific articles that may provide opposing perspectives pertaining to the petition. For example, we provided a link to the below article, published by the Orlando Sentinel, describing professor Negy's claim that he is the subject of a "witch hunt".

Player 2 also had the option of taking time to search the Internet for other opposing perspectives.

And, importantly, Player 2 also had the option NOT to read any opposing perspectives before deciding whether or not to sign the petition.
Republicans, Amazon Petition:

After viewing this petition, Player 2 decided whether or not to sign it.

Additionally, we gave Player 2 an opportunity to take some time—If Player 2 wanted it—to consider OPPOSING perspectives before deciding whether to sign. It was completely up to Player 2 whether to use this time, and how.

For Player 2’s convenience, we provided links to some specific articles that may provide opposing perspectives pertaining to the petition. For example, we provided a link to the below article, published by a local Houston news outlet, reporting on the perspective of an activist who does not think “Blue Lives Murder” merchandise constitutes hate speech.

Player 2 also had the option of taking time to search the internet for other opposing perspectives.

And, importantly, Player 2 also had the option NOT to read any opposing perspectives before deciding whether or not to sign the petition.

Next, we explained that Player 2s varied with respect to the amount of time they spent looking:

Importantly, participants assigned to the role of Player 2 varied in the amount of time they spent considering opposing perspectives before deciding whether or not to sign the petition.

Some Player 2s spent no or very little time considering opposing perspectives before deciding whether or not to sign the petition.

In contrast, other Player 2s spent a lot of time considering opposing perspectives before deciding whether or not to sign the petition.

Next, we collected our dependent variables, for a set of different Player 2s with different behavioral profiles (i.e., a set of Player 2s about whom we provided different information about their punishment and looking behavior).

The first two Player 2s that subjects evaluated were punishers who did vs. did not look. We manipulated, between subjects, the order in which these Player 2s were presented. After the first but before the second was presented, subjects saw a transition screen.
This was the screen subjects used to evaluate a punisher who did not look:

Now, we would like you to make a Sharing Game decision.

Another participant, in the role of Player 2, spent a BELOW-AVERAGE amount of time considering opposing perspectives, and then chose TO sign the petition.

After this study is completed, we will match Player 1s with Player 2s and compute bonuses for all players. You may be matched with this Player 2.

If you are paired with this Player 2, how many cents, if any, would you like to share?

<table>
<thead>
<tr>
<th>0</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>35</th>
<th>40</th>
<th>45</th>
<th>50</th>
</tr>
</thead>
</table>

Note: this Player 2 is a REAL participant who really did make the decisions described above. If you are paired with this Player 2, your decision really will be used to compute your bonus and their bonus in this study.

Please also answer the following questions about your impression of this Player 2.

**How positive is your evaluation of this Player 2?**

<table>
<thead>
<tr>
<th>Very negative</th>
<th>Neutral</th>
<th>Very positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>100</td>
</tr>
</tbody>
</table>

**To what extent do you think this Player 2 is a loyal supporter of Black Lives Matter?**

<table>
<thead>
<tr>
<th>Not a supporter at all</th>
<th>A moderate supporter</th>
<th>A very loyal supporter</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>100</td>
</tr>
</tbody>
</table>

**How competent do you think this Player 2 is?**

<table>
<thead>
<tr>
<th>Very incompetent</th>
<th>Neutral</th>
<th>Very competent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>100</td>
</tr>
</tbody>
</table>

**How fair do you think this Player 2 is?**

<table>
<thead>
<tr>
<th>Very unfair</th>
<th>Neutral</th>
<th>Very fair</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>100</td>
</tr>
</tbody>
</table>

Note: the above screenshot shows the screen for Democrats in Study 1a, where we measured loyalty towards Black Lives Matter; for Republicans in Study 1b, we instead measured loyalty towards Blue Lives Matter. We also randomized between-subjects the order in which we presented the questions about how competent, fair, and loyal Player 2 was; subjects were randomly assigned to one order and then that order was consistently applied across all Player 2s that they evaluated.

The screen that subjects used to evaluate a punisher who did look was identical, except that we replaced “below-average” with “above-average” in the second sentence on the screen.
This was the transition screen that subjects saw (between evaluating a punisher who did vs. did not look):

Thank you for making that decision.

On the subsequent screens, you will make a series of decisions about how much to share with OTHER potential Player 2s that you could be paired with.

These Player 2s are all REAL participants, who, like you, identify as Democrats. They all had the opportunity to sign the petition about Professor Charles Negy. And they all chose how long to spend considering opposing perspectives before deciding whether or not to sign the petition. However, these Player 2s varied in terms of the decisions that they made in the study.

On the subsequent screens, please decide how much you would like to share with each of them, if you are ultimately paired with them.

Note: the above screenshot shows screen for Democrats who were assigned to the Negy petition; the text was adapted to reference the correct party and petition for all subjects.

The next two Player 2s that subjects evaluated were non-punishers who did vs. did not look. Subjects who evaluated a punisher who did look before a punisher who did not look correspondingly evaluated a non-punisher who did look before evaluating a non-punisher who did not look, and vice versa. The screens that subjects used to evaluate non-punishers were identical to the screens they used to evaluate punishers, except that we replaced “chose TO sign the petition” with “chose NOT to sign the petition”.

The next two Player 2s that subjects evaluated were punishers and non-punishers, with no information provided about looking. Before presenting these two Player 2 profiles, we showed subjects this transition screen:

Thank you.

On the next screens, you will make two more sharing game decisions about Player 2s that you might be matched with.

For these decisions, we will tell you whether or not Player 2 decided to sign the petition.

However, we will NOT tell you anything about how long Player 2 spent considering opposing perspectives before making this decision.

The screens that subjects used to evaluate Player 2s without looking information were identical to the screens they used to evaluate Player 2s with looking information, except that we did not provide information about looking. So, the non-punisher screen began like this:

Now, we would like you to make another Sharing Game decision.

Another participant, in the role of Player 2, chose NOT to sign the petition.

And punisher screen began identically, except that we replaced “chose NOT to” with “chose TO”.

---

Please also answer the following questions about our impression of this Player 2.

Did you like them? 

If you are paired with this Player 2, how many cents, if any, would you give each of them, if you were paid for each decision.

Another participant, in the role of Player 2, chose NOT to sign the petition.

On the next screens, you will make two more sharing game decisions about Player 2s that you might be matched with.

Thank you.
Next, subjects evaluated a Player 2 about whom they had no information:

Thank you for making these decisions.

Finally, how many cents would you like to share with Player 2, if given no information about Player 2’s decisions?

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>35</th>
<th>40</th>
<th>45</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: this decision may really be used to compute your bonus and the bonus of a Player 2 in this study.

Given no information about Player 2’s decisions, how positive is your impression of Player 2?

<table>
<thead>
<tr>
<th>Very negative</th>
<th>Neutral</th>
<th>Very positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
</tr>
</tbody>
</table>

Given no information about Player 2’s decisions, to what extent do you think Player 2 is a loyal supporter of Black Lives Matter?

<table>
<thead>
<tr>
<th>Not a supporter at all</th>
<th>A moderate supporter</th>
<th>A very loyal supporter</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>100</td>
</tr>
</tbody>
</table>

Given no information about Player 2’s decisions, how competent do you think Player 2 is?

<table>
<thead>
<tr>
<th>Very incompetent</th>
<th>Neutral</th>
<th>Very competent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>100</td>
</tr>
</tbody>
</table>

Given no information about Player 2’s decisions, how fair do you think Player 2 is?

<table>
<thead>
<tr>
<th>Very unfair</th>
<th>Neutral</th>
<th>Very fair</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>100</td>
</tr>
</tbody>
</table>

Finally, subjects completed the following post-experimental survey:

Thank you for making those decisions.

Now, we would like you to provide your own opinion about the petition. We have included a screenshot of the petition below, for your reference.

Please rate your agreement with the following statement: "I strongly support the petition and the underlying cause behind it."

<table>
<thead>
<tr>
<th>1 - Strongly disagree</th>
<th>2</th>
<th>3 - Somewhat disagree</th>
<th>4</th>
<th>5 - Neither agree nor disagree</th>
<th>6</th>
<th>7 - Somewhat agree</th>
<th>8</th>
<th>9 - Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Below this question, we re-printed the screenshot of the petition]

Please describe how you made your choices in this study.

[Page break]

Is there anything you’d like to share with us about your impression of this study?

[Page break]
To what extent have you previously participated in other studies like this one?

<table>
<thead>
<tr>
<th>1 - Nothing like this scenario</th>
<th>2</th>
<th>3 - Something like this scenario</th>
<th>4</th>
<th>5 - Exactly this scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

To what extent do you believe that the petition is real?

<table>
<thead>
<tr>
<th>1 - Very skeptical that it is real</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7 - Very confident that it is real</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

To what extent do you believe that you really will be matched with a real Player 2?

<table>
<thead>
<tr>
<th>1 - Very skeptical that I will be matched with a real Player 2</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7 - Very confident that I will be matched with a real Player 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Page break]
Please choose the category that describes the total amount of income you earned in 2020. Consider all forms of income, including salaries, tips, interest and dividend payments, scholarship support, student loans, parental support, social security, alimony, and child support, and others.

- Under $5,000
- $5,000-$10,000
- $10,001-$15,000
- $15,001-$25,000
- $25,001-$35,000
- $35,001-$50,000
- $50,001-$65,000
- $65,001-$80,000
- $80,001-$100,000
- Over $100,000

Please specify your race. (Choose one or more categories)

- White/Caucasian (Anglo/Euro) American
- Black or African American
- Asian or Asian American
- American Indian or Alaska Native
- Native Hawaiian or other Pacific Islander
- Hispanic/Latino
- Multicultural

You indicated that you lean Democrat. Which is more true of you?

- I consider myself a strong Democrat, who strongly supports the party
- I consider myself a weak Democrat, who only leans towards the party

On the below scale, how strong of a Democrat do you consider yourself?

1 - Not very strong Democrat
2
3
4 - Moderately strong Democrat
5
6
7 - Very strong Democrat

In general, how conservative or liberal do you consider yourself to be?

1 - Very conservative
2
3
4 - Neither liberal nor conservative
5
6
7 - Neither liberal nor conservative
8
9 - Very liberal

When it comes to SOCIAL policy, how conservative or liberal do you consider yourself to be?

1 - Very conservative
2
3
4 - Neither liberal nor conservative
5
6
7 - Neither liberal nor conservative
8
9 - Very liberal

When it comes to ECONOMIC policy, how conservative or liberal do you consider yourself to be?

1 - Very conservative
2
3
4 - Neither liberal nor conservative
5
6
7 - Neither liberal nor conservative
8
9 - Very liberal

To what extent do you support Black Lives Matter?

1 - Strongly oppose
2
3
4 - Neither support nor oppose
5
6
7 - Neither support nor oppose
8
9 - Strongly support

Note: the above screenshot shows the screen for Democrats in Study 1a, where we measured support for Black Lives Matter; for Republicans in Study 1b, we referenced Republicans (rather than Democrats) and measured support for Blue Lives Matter (“To what extent do you support "Blue Lives Matter" (a countermovement, started in response to Black Lives Matter, advocating that those who are prosecuted and convicted of killing law enforcement officers should be sentenced under hate crime statutes)?”).
6.2 Studies 2a-b

Studies 2a-b were extremely similar to Studies 1a-b. Here, we show screenshots that highlight the few differences.

First, the screen introducing the Dictator Game was identical to Studies 1a-b for all Study 2 conditions except the “other participant” treatment from Study 2a. In that treatment, the screen looked different because we did not say that Player 2 was a co-partisan, but rather described Player 2 merely as “another participant”, as follows:

In this study, you will also participate in an interactive game with another participant.

The game is called the Sharing Game, and it has two players: Player 1 and Player 2. You will be Player 1.

Player 2 is another participant.

In the Sharing Game, you (Player 1) start with 50 cents. You then choose how many cents, if any, to share with Player 2.

Please answer the following questions, to make sure you understand the Sharing Game.

Imagine that you are deciding how much to share with Player 2.
Which decision will result in you earning the most money?
- You deciding to share 0 cents
- You deciding to share 25 cents
- You deciding to share 50 cents

Imagine that you are deciding how much to share with Player 2.
Which decision will result in Player 2 earning the most money?
- You deciding to share 0 cents
- You deciding to share 25 cents
- You deciding to share 50 cents

Second, before subjects began evaluating Player 2, we showed a screen that (i) provided subjects in the treatments that “added information” with some additional information about the Player 2 they were paired with, and (ii) asked all subjects to write a paragraph reflecting on their initial impression of Player 2.

This was the relevant screen in the baseline conditions:

In light of all of the information that you have received about Player 2, what is initial your impression of them? Please write a few sentences in the box below.
This was the screen in the “other participant” treatment from Study 2a:

On the following screens, you will answer some questions about Player 2. Remember, Player 2 is a REAL other participant.

In light of all of the information that you have received about Player 2, what is initial your impression of them? Please write a few sentences in the box below.

This was the screen in the “Democrat + Conflict of Interest (Privacy)” treatment from Study 2a:

On the following screens, you will answer some questions about Player 2. Remember, Player 2 is a REAL other participant.

First, however, we’d like to provide some additional information about the Player 2 you have been paired with. In particular, below we have provided a screenshot of some questions that we asked Player 2, and the answers that Player 2 provided.

On this page, we have printed all instructions for you in GREEN, to help you distinguish the screenshot from the instructions for you.

Please carefully consider the questions we asked Player 2, and the answers they provided, before advancing the screen.

Do you think of yourself as more of a Democrat or a Republican?

If instead you consider yourself an Independent, please indicate whether, as of today, you lean more Democratic or Republican.

Democrat
Republican

Before you decide whether to sign the petition, we’d like you to reflect on whether there might be any reason that signing the petition would not be in your personal self-interest.

In other words, do you have any "conflict of interest" with the petition? Is there any reason it would be bad for YOU for this petition to receive support?

As a reminder, the petition calls for the LAPD to fire Chief Michael Moore. It argues that Chief Moore should be fired following comments he made that blamed protestors for George Floyd’s death.

In the box below, please describe any reason(s) you can think of that signing the petition would not be in your personal self-interest.

Please explain in at least 1-2 sentences.

Signing the petition might make some of my personal information public. This could lead to me receiving spam. This could become more severe as the petition receives more support and attention.

In light of all of the information that you have received about Player 2, what is initial your impression of them? Please write a few sentences in the box below.
In the other treatments that added information, the screen was identical to the screen shown above, except that we featured a different Player 2 response screenshot.

This was the Player 2 response screenshot in the “Democrat + Ignore injustice” treatment from Study 2a:

*Do you think of yourself as more of a Democrat or a Republican?*

If you instead consider yourself an independent, please indicate whether, as of today, you lean more Democratic or Republican.

- Democrat
- Republican

In this section of the HIT, we'd like you to think about a time that you witnessed racism or racial injustice towards somebody else. Is this something you have witnessed before? If so, please try to think of what happened and share it below.

In particular, please describe what you witnessed in at least 2-3 sentences.

I saw a person in a store get followed by employees just because of their skin color. The person was black and it was a luxury designer store, the workers followed the customer everywhere.

In the situation you just described, did you do anything to stop or address the racism or injustice? If so, what did you do?

I did not do anything

Looking back on the situation that you described, what more or else could you have done to stop or address the injustice?

I could have stood up for the person and told the workers to stop harassing the customer.
This was the screenshot in the “Democrat + Conflict of interest (Police)” treatment from Study 2a:

Do you think of yourself as more of a Democrat or a Republican?

If instead you consider yourself an Independent, please indicate whether, as of today, you lean more Democratic or Republican.

Democrat

Republican

Before you decide whether to sign the petition, we’d like you to reflect on whether there might be any reason that signing the petition would not be in your personal self-interest.

In other words, do you have any “conflict of interest” with the petition? Is there any reason it would be bad for you for this petition to receive support?

As a reminder, the petition calls for the LAPD to fire Chief Michael Moore. It argues that Chief Moore should be fired following comments he made that blamed protestors for George Floyd’s death.

In the box below, please describe any reason(s) you can think of that signing the petition would not be in your personal self-interest.

Please explain in at least 1-2 sentences.

I have an uncle who was a police officer.

This was the screenshot in the “Democrat + Previously Independent” treatment from Study 2a:

Do you think of yourself as more of a Democrat or a Republican?

If instead you consider yourself an Independent, please indicate whether, as of today, you lean more Democratic or Republican.

Democrat

Republican

Prior to the way you currently identify politically, which political party did you consider yourself to be a member of (if applicable)?

Democrat

Republican

Independent/Other (please specify)

Please describe, in a few sentences, the history of the way that you have identified politically over time.

If you have not always identified politically the way that you do now, please comment on how you used to identify, and what prompted the change. What lead you to your current political identity?

I used to praise Trump, and then COVID hit, and Trump showed his true colors, and I became a Democrat after that.

How many of the last four presidential or midterm elections (2020, 2018, 2016, 2014) have you cast a vote in?

0 1 2 3 4
This was the screenshot in the “Democrat + Responded to injustice” treatment from Study 2b:

Do you think of yourself as more of a Democrat or a Republican?

If you instead consider yourself an independent, please indicate whether, as of today, you lean more Democratic or Republican.

Democrat

Republican

In this section of the HIT, we'd like you to think about a time that you witnessed racism or racial injustice towards somebody else.

Is this something you have witnessed before? If so, please try to think of what happened and share it below.

In particular, please describe what you witnessed in at least 2-3 sentences.

I have witnessed racism in my former organization where coworkers of color or anyone that wasn't white didn't get the same opportunities white workers got. They would get little to no promotions and less salary wages

In the situation you just described, did you do anything to stop or address the racism or injustice? If so, what did you do?

Yes I stood up against my former employer and encouraged other colleagues to do so as well

After subjects reflected on their initial impression of Player 2, we also showed one more screen (that was not included in Studies 1a-b) before subjects began evaluating Player 2. In this screen, we explained that subjects would be making multiple decisions about how much to share with Player 2, depending on how Player 2 had previously behaved. This was the relevant screen:

On the subsequent screens, you will make a series of decisions about how much to share with Player 2, given different information about how Player 2 behaved in the previous study.

Importantly, after this study is completed, we will use ONE of your decisions to compute your bonus and pay you. Specifically, we will use the decision you make that corresponds to how Player 2 ACTUALLY behaved in the previous study.

However, because you do not yet know how Player 2 actually behaved, you do not yet know which of your decisions will be used to compute your bonus. So, all of the decisions you will make are important.

Then, like in Studies 1a-b, we collected our dependent variables, for the same set of Player 2 behavioral profiles that we featured in Studies 1a-b. However, while Studies 1a-b described each behavioral profile as corresponding to a different Player 2, Studies 2a-b described each behavioral profile as corresponding to a different way that a single Player 2 (i.e., the Player 2 that the subject was matched with) might have behaved. Moreover, we did not show a “transition screen” after the first behavior profile (like we did in Studies 1a-b) because we had already explained upfront that subjects would be making multiple decisions.
Here is an example of a Study 2 behavioral profile screen (in this case, corresponding to Player 2 choosing to punish without looking):

Now, we would like you to make a Sharing Game decision.

We will use the decision you make on this screen if the Player 2 you are paired with...

...spent a BELOW-AVERAGE amount of time considering opposing perspectives, and then chose TO sign the petition.

If Player 2 behaved in this manner, how many cents, if any, would you like to share with them?

0  5  10  15  20  25  30  35  40  45  50

Note: After this study is completed, we will use ONE of your decisions. Specifically, we will use the decision you make that corresponds to how Player 2 ACTUALLY behaved. Then, we will use this decision to compute bonuses for both players. So, the decision you make on this screen may actually be used.

Please also answer the following questions about your impression of Player 2, if Player 2 spent a BELOW-AVERAGE amount of time considering opposing perspectives, and then chose TO sign the petition.

In this scenario, how **positive** is your evaluation of Player 2?

<table>
<thead>
<tr>
<th>Very negative</th>
<th>Neutral</th>
<th>Very positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>25</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In this scenario, how **competent** do you think Player 2 is?

<table>
<thead>
<tr>
<th>Very incompetent</th>
<th>Neutral</th>
<th>Very competent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>25</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In this scenario, how **fair** do you think Player 2 is?

<table>
<thead>
<tr>
<th>Very unfair</th>
<th>Neutral</th>
<th>Very fair</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>25</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In this scenario, to what extent do you think Player 2 is a **loyal supporter** of Black Lives Matter?

<table>
<thead>
<tr>
<th>Not a supporter at all</th>
<th>A moderate supporter</th>
<th>A very loyal supporter</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>25</td>
<td>75</td>
</tr>
</tbody>
</table>
6.3 Studies 3-4
After reporting their Prolific IDs and providing informed consent, subjects began by reporting their age, gender, and political party affiliation, and answering two attention check questions; these measures were all identical to those presented at the beginning of Study 1. So, see Study 1 in this document for screenshots.

Next, we introduced the Dictator Game (described as the “Sharing Game”), informed subjects that they would be interacting with another participant (described as “Player 1”) who shares their political party, and presented a set of comprehension questions about the game.

For Democrats in Study 3a, Player 1 was described as relatively less ideological:

In this HIT, you will participate in an interactive game with another MTurk worker.

The game is called the Sharing Game, and it has two players: Player 1 and Player 2. **You will be Player 2.**

Another Mturk worker will be Player 1. **Like you, this worker ALSO indicated to us in a survey that they identify as a Democrat.**

More specifically, Player 1 indicated that they identify as a weak Democrat, who only leans towards the party.

And, just like we just told you that Player 1 is a Democrat, we will also tell Player 1 that you are a Democrat.

In the Sharing Game, Player 1 starts with 50 cents. Player 1 then chooses how many cents, if any, to share with you (Player 2).

Please answer the following questions, to make sure you understand the Sharing Game.

Imagine that Player 1 is deciding how much to share with to you.

Which decision will result in **Player 1** earning the highest payoff?

- Player 1 deciding to share 0 cents
- Player 1 deciding to share 25 cents
- Player 1 deciding to share 50 cents

Imagine that Player 1 is deciding how much to share with to you.

Which decision will result in **you** earning the highest payoff?

- Player 1 deciding to share 0 cents
- Player 1 deciding to share 25 cents
- Player 1 deciding to share 50 cents

What information do you know about Player 1?

- Player 1 identifies as a Republican
- Nothing
- Player 1 identifies as a weak Democrat, who only leans towards the party
For Democrats in Study 4a, Player 1 was described as relatively more ideological. Thus, the screen instead began as follows:

In this HIT, you will also participate in an interactive game with another MTurk worker.

The game is called the Sharing Game, and it has two players: Player 1 and Player 2. You will be Player 2.

Another Mturk worker will be Player 1. Like you, this worker ALSO indicated to us in a survey that they identify as a Democrat.

More specifically, Player 1 indicated that they identify as a strong Democrat who strongly supports Black Lives Matter.

And, just like we just told you that Player 1 is a Democrat, we will also tell Player 1 that you are a Democrat.

In the Sharing Game, Player 1 starts with 50 cents. Player 1 then chooses how many cents, if any, to share with you (Player 2).

The comprehension questions that followed were identical to those presented in Study 3a, except that the final answer chance for last question instead read “Player 1 identifies as a strong Democrat who strongly supports Black Lives Matter”.

For Republicans in Study 3b, the screen was identical to what we showed in Study 3a, except that we referenced Republicans instead of Democrats.

For Republicans in Study 4b, the screen was identical to what we showed in Study 4a, except that we (i) referenced Republicans instead of Democrats and (ii) the sentence starting with “more specifically” instead read “More specifically, Player 1 indicated that they identify as a strong Republican who strongly supports “Blue Lives Matter” (a countermovement, started in response to Black Lives Matter, advocating that those who are prosecuted and convicted of killing law enforcement officers should be sentenced under hate crime statutes).” Furthermore, the final answer choice for last question read “Player 1 identifies as a strong Republican who strongly supports Blue Lives Matter”.

If subjects answered any comprehension questions incorrectly, they were presented with the screen involving the questions a second time, along with the text: “On the previous page, you answered one or more questions incorrectly. You MUST answer ALL questions correctly to continue on with the study. Please carefully re-read the below instructions and answer the questions again”. Here, they were not allowed to precede to the next screen until they answered all questions correctly.

Next, we told subjects that they had the opportunity to sign a petition. Below we show the text we presented for each of the three petitions. Below this text, we displayed a screenshot of the petition. These screenshots were identical to those presented in Study 1; so, see Study 1 in this document for screenshots.
Democrats, Moore petition (Study 3a):

This HIT will also have another component.

In addition to participating in the Sharing Game, you will also have the opportunity to sign a petition, hosted on Change.org. Below, we have provided a screenshot of the petition. Later in this HIT, we will provide a link to the petition. Over 300,000 people have already signed the petition, as highlighted in the screenshot below.

The petition calls for the LAPD to fire Chief Michael Moore. It argues that Chief Moore should be fired following comments he made that blamed protestors for George Floyd’s death.

On this screen, please read through the petition. Then, later in this HIT, you will have the chance to decide whether you’d like to help get Chief Moore fired by signing the petition.

You can choose to sign or not to sign. One thing you should keep in mind is that we will tell Player 1 whether or not you choose to sign.

Democrats, Negy petition (Study 4a):

This HIT will also have another component.

In addition to participating in the Sharing Game, you will also have the opportunity to sign a petition, hosted on Change.org. Below, we have provided a screenshot of the petition. Later in this HIT, we will provide a link to the petition. Over 30,000 people have already signed the petition, as highlighted in the screenshot below.

The petition calls for the University of Central Florida to fire professor Charles Negy. It argues that professor Negy should be fired on the basis of abhorrent racist comments he has made on his personal Twitter account.

On this screen, please read through the petition. Then, later in this HIT, you will have the chance to decide whether you’d like to help get professor Negy fired by signing the petition.

You can choose to sign or not to sign. One thing you should keep in mind is that we will tell Player 1 whether or not you choose to sign.

Note: the above two screenshots show the screens from the “Punishment Observable” and “Both Observable” conditions, where punishment was observable.

Republicans, Amazon petition (Studies 3b and 4b):

This HIT will also have another component.

In addition to participating in the Sharing Game, you will also have the opportunity to sign a petition, hosted on Change.org. Below, we have provided a screenshot of the petition. Later in this HIT, we will provide a link to the petition. Over 115,000 people have already signed the petition, as highlighted in the screenshot below.

The petition calls for the removal of “Blue Lives Murder” merchandise from Amazon. It argues that the merchandise is hateful and must be immediately removed.

On this screen, please read through the petition. Then, later in this HIT, you will have the chance to decide whether you’d like to help remove “Blue Lives Murder” merchandise from Amazon by signing the petition.

You can choose to sign or not to sign. One thing you should keep in mind is that your decision will be completely private: we will NOT tell Player 1 whether or not you choose to sign.

Note: the above screenshot shows the screen from the “Nothing Observable” condition, where punishment was not observable.
Next, we told subjects that they would have a chance to look at opposing perspective articles, and showed an example headline from an opposing perspective article.

Democrats, Negy Petition (Study 4a):

Additionally, we will give you an opportunity to take some time—if you would like—to consider OPPOSING perspectives before deciding whether to sign. It is completely up to you whether to use this time, and how.

For your convenience, we will provide you with links to some specific articles that may provide opposing perspectives pertaining to the petition. For example, we will provide a link to the below article, published by the Orlando Sentinel, describing professor Negy’s claim that he is the subject of a “witch hunt”. You can also choose to use the time to search the Internet for other opposing perspectives.

Importantly, you can also choose NOT to read any opposing perspectives and instead move forward with the survey.

You can take as much or as little time as you would like to consider opposing perspectives. One thing you should keep in mind is that we will tell Player 1 how long you spend considering opposing perspectives before deciding whether to sign.

Democrats, Moore petition (Study 3a):

Additionally, we will give you an opportunity to take some time—if you would like—to consider OPPOSING perspectives before deciding whether to sign. It is completely up to you whether to use this time, and how.

For your convenience, we will provide you with links to some specific articles that may provide opposing perspectives pertaining to the petition. For example, we will provide a link to the below article published by the LA Times describing how elected politicians in Los Angeles, including the mayor, continue to support Police Chief Moore. You can also choose to use the time to search the Internet for other opposing perspectives.

Importantly, you can also choose NOT to read any opposing perspectives and instead move forward with the survey.

You can take as much or as little time as you would like to consider opposing perspectives. One thing you should keep in mind is that your decision will be completely private: we will NOT tell Player 1 how long you spend considering opposing perspectives before deciding whether to sign.
Republicans, Amazon Petition (Studies 3b and 4b):

Additionally, we will give you an opportunity to take some time—if you would like—to consider OPPOSING perspectives before deciding whether to sign. It is completely up to you whether to use this time, and how.

For your convenience, we will provide you with links to some specific articles that may provide opposing perspectives pertaining to the petition. For example, we will provide a link to the below article, published by a local Houston news outlet, reporting on the perspective of an activist who does not think “Blue Lives Murder” merchandise constitutes hate speech. You can also choose to use the time to search the Internet for other opposing perspectives.

Importantly, you can also choose NOT to read any opposing perspectives and instead move forward with the survey.

You can take as much or as little time as you would like to consider opposing perspectives. One thing you should keep in mind is that your decision will be completely private: we will NOT tell Player 1 how long you spend considering opposing perspectives before deciding whether to sign.

Note: the above two screenshots show the screens from the “Nothing Observable” and “Punishment Observable” conditions, where looking was not observable.
Next, we summarized for subjects the relevant information about their observability condition, and presented a second set of comprehension questions about what was observable to Player 1. This screen looked different across conditions.

Nothing Observable:

Thank you. On this page, we’d like to review how the petition and Sharing Game components of this HIT relate to each other.

In the Sharing Game, before deciding how much to share with you, Player 1 will see the full text of the petition and learn that you had the opportunity to sign it. Furthermore, Player 1 will learn that you were given time to consider opposing perspectives before committing to signing the petition.

Importantly, however, we will NOT tell Player 1 how long you spend considering opposing perspectives, OR whether or not you ultimately decide to sign the petition.

In other words, your signing decision—and the time you spend considering opposing perspectives—will both remain completely private.

Please answer the following questions, to make sure you understand.

Before Player 1 decides how much money to share with you…

Will Player 1 learn that you have the opportunity to sign the petition?

☐ No
☐ Yes, but Player 1 will not get to read the full text of the petition
☐ Yes, and Player 1 will also get to read the full text of the petition

Will Player 1 learn that you were given time to consider opposing perspectives before committing to signing?

☐ No
☐ Yes

What will Player 1 learn about your behavior?

☐ Nothing. Player 1 will not find out how long I spend considering opposing perspectives, or whether I ultimately sign the petition
☐ Player 1 will only find out whether I ultimately sign the petition (but not how long I spend considering opposing perspectives)
☐ Player 1 will both find out how long I spend considering opposing perspectives, and whether I ultimately sign the petition
In Punishment Observable, we presented the same questions, but introduced them as follows:

Thank you. On this page, we’d like to review how the petition and Sharing Game components of this HIT relate to each other.

In the Sharing Game, before deciding how much to share with you, Player 1 will see the full text of the petition and learn that you had the opportunity to sign it. Furthermore, Player 1 will learn that you were given time to consider opposing perspectives before deciding whether to sign the petition.

Importantly, we will NOT tell Player 1 how long you spend considering opposing perspectives. We WILL, however, tell Player 1 whether or not you ultimately decide to sign the petition.

In other words, while the time you spend considering opposing perspectives will remain completely private, your signing decision will be shared with Player 1.

In Both Observable, we likewise presented the same questions, but introduced them as follows:

Thank you. On this page, we’d like to review how the petition and Sharing Game components of this HIT relate to each other.

In the Sharing Game, before deciding how much to share with you, Player 1 will see the full text of the petition and learn that you had the opportunity to sign it. Furthermore, Player 1 will learn that you were given time to consider opposing perspectives before deciding whether to sign the petition.

And, importantly, we will tell Player 1 how long you spend considering opposing perspectives, AND whether or not you ultimately decide to sign the petition.

In other words, your signing decision—and the time you spend considering opposing perspectives—will both be shared with Player 1.

Like with the first set of comprehension questions, if subjects answered any questions on this screen incorrectly, they were presented with the screen a second time, along with the text: “On the previous page, you answered one or more questions incorrectly. You MUST answer ALL questions correctly to continue on with the study. Please carefully re-read the below instructions and answer the questions again”. Here, they were not allowed to precede to the next screen until they answered all questions correctly.

Next, we informed subjects that they were about to precede to the page where they could look at opposing perspectives:

Thank you.

On the next screen, we will give you the opportunity to take some time—if you would like—to consider opposing perspectives before deciding whether to sign the petition.

As a reminder, it is completely up to you whether to use this time, and how. We will provide you with links to some specific articles that may provide opposing perspectives pertaining to the petition. You can also choose to use the time to search the Internet for other opposing perspectives.

Importantly, you can also choose NOT to read any opposing perspectives and instead move forward with the survey.

Remember, we will tell Player 1 how long you choose to spend considering opposing perspectives before deciding whether to sign the petition. So Player 1 will learn whether you make a quick decision about whether to sign the petition, or choose to extensively consider opposing perspectives first.

Note: the above screenshot shows the screen from “Both Observable”. In the other two conditions, the last paragraph instead read: “Remember, we will NOT tell Player 1 how long you choose to spend considering opposing perspectives before deciding whether to sign the petition. So Player 1 will NOT learn whether you make a quick decision about whether to sign the petition, or choose to extensively consider opposing perspectives first”.
Next, we measured looking by presenting the following screens, and measuring which, if any, links subjects clicked on. Subjects were not informed that their link-clicking would be tracked.

Democrats, Negy petition (Study 4a):

Please take as much or as little time as you would like to consider opposing perspectives.

Remember, we will tell Player 1 how long you spend on this page considering opposing perspectives.

When you have decided whether to sign the petition, please advance the screen.

Previously, we showed you the headline of the below article, published by the Orlando Sentinel, describing professor Negy's claim that he is the subject of a "witch hunt". The link to view this article is here.

Below is the headline of another article you may consider, published by the National Association of Scholars, arguing that firing professor Negy would be a violation of his rights. The link to view this article is here.

Note: the above screenshot shows the screen from the “Both Observable” condition, where looking was observable.
Democrats, Moore petition (Study 3a):

Please take as much or as little time as you would like to consider opposing perspectives.

Remember, we will NOT tell Player 1 how long you spend on this page considering opposing perspectives.

When you have decided whether to sign the petition, please advance the screen.

Previously, we showed you the headline of the below article, published by the LA Times, describing the support Chief Moore has maintained in Los Angeles political circles since the controversy erupted. The link to view this article is [here].

Below is the headline of another article you may consider, published by FOX 11 Los Angeles, reporting on Los Angeles Mayor Eric Garcetti's statement of support for Chief Moore. The link to view this article is [here].
Republicans, Amazon Petition (Studies 3b and 4b):

Please take as much or as little time as you would like to consider opposing perspectives.

Remember, we will NOT tell Player 1 how long you spend on this page considering opposing perspectives.

When you have decided whether to sign the petition, please advance the screen.

Previously, we showed you the headline of the below article, published by Cick2Houston, reporting on the perspective of an activist who does not think “Blue Lives Murder” merchandise constitutes hate speech. The link to view this article is [here](#).

Below is the headline of another article you may consider, published by FIRE, about what constitutes hate speech and the ways that it is protected by the first amendment. The link to view this article is [here](#).

Note: the above two screenshots show the screens from the “Nothing Observable” and “Punishment Observable” conditions, where looking was not observable.
Next, we measured punishment by presenting the following screens, and measuring whether subjects clicked the link to the petition. Subjects were not informed that their link-clicking would be tracked.

On the next page, you will decide whether or not to sign the petition.

If you choose to sign, we will NOT collect your identifying information. Instead, we will ask you to show us that you really did sign another way.

Specifically, immediately after you sign, you will be redirected to a new screen. At the top of this new screen, you will see the 1-2-3 image below. However, instead of seeing the black box, you will see some text. Specifically, you will see a simple phrase, written at the top of the page in relatively small font.

When you see the 1-2-3 image, you will be DONE signing the petition. You do NOT need to take any further action. Instead, please stay on THIS PAGE and pay attention to the phrase that is written above the 1-2-3 image (in place of the black box).

If you sign the petition, we will ask you to report that phrase back to us, to show that you really did sign.

[Page break]

Thank you. Now, you will decide whether or not to sign the petition.

Remember, Player 1 will learn about the petition (and will get to read the full text of the petition). And we will tell Player 1 whether or not you ultimately chose to sign the petition.

If you would like to sign, please do so now by clicking here.

Remember, if you sign, please stay on the subsequent screen and pay attention to the phrase written above the 1-2-3 image!

Did you choose to sign the petition?

☐ Yes
☐ No

If you DID sign the petition, you should be able to see the 1-2-3 image, pictured above, on the next screen. To show us that you signed, please type the phrase written above the 1-2-3 image (instead of the black box).

If you did NOT sign the petition, please instead just type “no”.

Note: If you DID sign the petition, but are confused by the above instructions, please just type any phrase that you saw after signing the petition.

Please also rate your agreement with the following statement:

"I am strongly committed to supporting the petition and the underlying cause behind it."

1 - Strongly disagree
2
3 - Somewhat disagree
4
5 - Neither agree nor disagree
6
7 - Somewhat agree
8
9 - Strongly agree

Note: the above screenshot shows the screens from the “Punishment Observable” and “Both Observable” conditions, where punishment was observable. In “Nothing Observable”, we replaced the sentence that read “And we will tell Player 1 whether or not you ultimate choose to sign the petition” with “But we will NOT tell Player 1 whether or not you sign the petition”.
Finally, subjects completed the following post-experimental survey:

Thank you for your decision. Now, we would like you to answer a few questions about the petition. (Unlike your signing decision, your responses to these questions will NOT be shown to Player 1.)

<table>
<thead>
<tr>
<th>Not at all</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>How moral do you think the petition is?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>To what extent do you agree with the petition?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>How comfortable are you with the petition's approach?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>To what extent do you think the petition's approach is proportionate and appropriate?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

[Note: the above screenshot shows the screens from the “Punishment Observable” and “Both Observable” conditions, where punishment was observable. In “Nothing Observable”, the parenthetical instead read: (As was the case for your signing decision, your responses to these questions will NOT be shown to Player 1).]

[Page break]

Thank you. For our records, on the screen where you had the chance to consider opposing perspectives, did you do any of the things listed below?

If so, please check all that apply. If not, that's totally fine; please leave this question blank and advance to the next page.

Note: your response to this question will NOT be shown to Player 1. It is only for our records.

☐ I read the Herald Mail Media article (headline: UCF professor behind tweets deemed racist says he is subject of witch hunt)
☐ I read the National Association of Scholars article (headline: Let the tweeter beware)
☐ I searched the Internet for other perspectives that would oppose the petition
☐ I searched the Internet for other perspectives that would support the petition

[Note: the above screenshot shows the screen from Study 4a, where Democrats saw the Negy petition. The text was adapted to reference the correct party and petition for all subjects.]

[Page break]

Thank you.

Before completing this HIT, how much did you know about the events described in the petition?

<table>
<thead>
<tr>
<th>1 - I had never heard of them</th>
<th>2</th>
<th>3 - I had heard of them, but knew only some details</th>
<th>4</th>
<th>5 - I had been closely following the situation and knew many details</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

Please describe how you made your choices in this HIT.

[Page break]
Now, we would like to ask you a few questions regarding the amount of time you decided to spend considering opposing perspectives.

To what extent did you make your decision (regarding how long to spend considering opposing perspectives) because...

1 - Not at all  2  3  4  5  6  7 - Entirely

...you personally felt that it was truly the right decision? ○ ○ ○ ○ ○ ○ ○
...you wanted to see yourself as a good person? ○ ○ ○ ○ ○ ○ ○
...you wanted others to see you as a good person? ○ ○ ○ ○ ○ ○ ○
...you wanted Player 1 to see you as a good person? ○ ○ ○ ○ ○ ○ ○

Now, we would like to ask you a few questions about your decision regarding whether to sign the petition.

To what extent did you make your decision (regarding whether to sign the petition) because...

1 - Not at all  2  3  4  5  6  7 - Entirely

...you personally felt that it was truly the right decision? ○ ○ ○ ○ ○ ○ ○
...you wanted to see yourself as a good person? ○ ○ ○ ○ ○ ○ ○
...you wanted others to see you as a good person? ○ ○ ○ ○ ○ ○ ○
...you wanted Player 1 to see you as a good person? ○ ○ ○ ○ ○ ○ ○

To what extent do you believe that the petition and articles providing opposing perspectives are real?

1 - Very skeptical that they are real  2  3  4  5  6  7 - Very confident that they are real

...they are real ○ ○ ○ ○ ○ ○ ○

To what extent do you believe that Player 1 and the sharing game are real?

1 - Very skeptical that they are real  2  3  4  5  6  7 - Very confident that they are real

...they are real ○ ○ ○ ○ ○ ○ ○
Note: the above screenshot shows the screen for Democrats in Studies 3a and 4a, where we measured support for Black Lives Matter; for Republicans in Studies 3b and 4b, we referenced the Republican party (rather than the Democrat party) and measured support for Blue Lives Matter (“To what extent do you support "Blue Lives Matter" (a countermovement, started in response to Black Lives Matter, advocating that those who are prosecuted and convicted of killing law enforcement officers should be sentenced under hate crime statutes)?”).