UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Defendant character influences mock juror judgments of blame, guilt, and punishment

Permalink

https://escholarship.org/uc/item/8v87v6j9

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors

Cheung, Vanessa
Lagnado, David

Publication Date

2023

Peer reviewed

Defendant Character Influences Mock Juror Judgments of Blame, Guilt, and Punishment

Vanessa Cheung (vanessa.cheung.14@ucl.ac.uk)
Department of Experimental Psychology, University College London,
26 Bedford Way, WC1H 0AP, London, UK

David Lagnado (d.lagnado@ucl.ac.uk)
Department of Experimental Psychology, University College London,
26 Bedford Way, WC1H 0AP, London, UK

Abstract

The present study explored how evaluations of a defendant's character can influence mock jurors' judgments using a belief-updating paradigm. Participants (N=143) were shown a trial transcript in which we manipulated the defendant's character by introducing an irrelevant moral behavior observed before the crime as well as prior conviction. We found that bad defendants were consistently judged to be more deserving of punishment than good defendants. While character information influenced judgments of guilt, blame, and intentionality immediately after it was presented, the effect diminished as participants received more information about the case, and ultimately did not shift their verdicts. In general, participants also mitigated moral judgments for good defendants rather than exacerbate judgments for bad defendants. Thematic analysis of judgment rationales also revealed that participants reasoned about actions, norms, and mental states when evaluating blame and punishment. We discuss the implications of this study in moral and legal decision-making.

Keywords: moral character; moral decision-making; blame attribution; juror decision-making; experimental jurisprudence

Introduction

In everyday moral situations, people are motivated to find explanations to events and hold agents accountable when negative outcomes occur (Alicke, 1992; Weiner, 1995). However, this can be complex as moral reasoning often depends on information beyond observable actions and outcomes, including inferences about agents' mental states (e.g., beliefs, intentions, and desires). Evidence that decision-makers focus on these factors regardless of whether harm occurs (Gromet et al., 2016; Young & Tsoi, 2013) and their preference for coherence between the agent's character, motivations, and actions (Hughes & Trafimow, 2012, 2015) suggests that they place emphasis on agent-related factors when ascribing blame.

Information about moral character plays a principal role in person perception, impression formation, and belief updating (Bos & Dijksterhuis, 2011; Brambilla et al., 2019). It allows us to ascertain whether a certain behavior was a one-off event (brought about by external, situational factors) or a behavioral tendency, which is an important distinction when making inferences about intentions and predicting future behavior.

According to the person-centered approach to moral judgment (Pizarro & Tannenbaum, 2012; Uhlmann et al., 2015), people are motivated to rely on information that is diagnostic of character when attributing blame. They integrate character into their judgments, using it to 'fill in the gaps' when drawing inferences. Indeed, it is reasonable to assume that

someone capable of committing heinous acts is more likely to engage in other questionable behavior (Lagnado & Gerstenberg, 2017). People might also believe that such individuals lack humanity and morality (Hughes, 2017), and should be blamed or punished simply for the kind of person they are.

This poses a clear problem in legal contexts, which aim to prevent defendants from being convicted for the type of person they are rather than for what they did (Anderson, 2012). In the United Kingdom, the Criminal Justice Act of 2003 states that bad character in criminal proceedings is defined as 'evidence of or a disposition towards misconduct' (Section 98, 353), and there are strict limits as to which kinds of character evidence are admissible or how heavily they should be weighed. In the United States, the Federal Rule of Evidence 404 also excludes character evidence and character reasoning. Jurors are instructed not to rely too heavily on character information such as prior conviction, but this is not always effective – they can still infer criminal disposition which then influences their verdict (Eisenberg & Hans, 2009; Givelber & Farrell, 2008; Ogloff & Rose, 2005).

Another concern is that jurors can still draw their own conclusions about the defendant by inference of some circumstantial evidence (Yankah, 2004). For example, in a murder trial, evidence that the defendant is having an affair may be used as evidence to show that he had motive to kill his wife. However, people do not use this information strictly to determine motive; they inevitably label the defendant as the type of person that would cheat on his wife – considering him to be callous, perhaps, or untrustworthy – thus blurring the line between his behavior and character (Nadler & McDonnell, 2011). As such, character becomes intertwined with other evidence because people are motivated to interpret information in a manner that confirms prior beliefs.

In a case where character evidence is admitted, jurors might first form prior beliefs about the defendant's character simply from observation, which they then combine with court-sanctioned character evidence to form an assessment of the defendant's character; this is then used in conjunction with non-character evidence to 'fill in the gaps' of events (Bavli, 2023). As a result, their judgments can be strongly influenced by prior beliefs about the defendant's character.

There is much more to be known about how character information affects and interacts with different criteria of blame in legal decision-making. Although there has been some dis-

1258

course on the influence of character in the court, it has seldom been empirically tested. While the effect of character on moral judgments has been explored in moral psychology, often using moral dilemmas or vignettes (Alicke, 1992; Nadler, 2012; Nadler & McDonnell, 2011), they often lack ecological validity. Furthermore, many prior studies fail to isolate the effect of character – often, the experimental manipulation is of both the agent's character and their motives (e.g., specifying that an agent's motive is to hide cocaine to establish his bad character in Alicke, 1992). It is therefore unclear whether a bad agent was blamed more for a harmful outcome because he is a bad person, or simply because having a bad motive makes a harmful act less justifiable.

To address these issues, we aimed to explore the effect of character when this information is incidental (i.e., not directly relevant) to the act under scrutiny. Basing the experiment on a realistic legal setting would also allow for easier exploration of the complex (and often abstract) factors surrounding moral and legal reasoning. As such, not only would our findings shed light on reasoning processes in legal contexts, they would also be relevant to more general moral judgment and decision-making processes.

Therefore, the present study investigated the effect of defendant character in moral and legal decision-making, specifically on judgments of the defendant's guilt, blameworthiness, intentionality, disposition, as well as people's desire to punish the defendant. To elucidate how people update their judgments as they received more information about the events – and thus proportionately less information about character – we also examined whether and, if so, how people change their judgments as the case progressed.

We investigated whether character information would have an effect on judgments even in a relatively straightforward case, where the defendant's intentions and motivations were unambiguous. We showed participants a shortened courtroom transcript based on a real theft case. As participants read through the transcript, they were asked, on five occasions, to give quantitative judgments on factors relevant to guilt and blame. This repeated-measures paradigm allowed us to assess the influence of experimentally manipulated information about the defendant's character. The defendant's moral character was manipulated between subjects, and information about character was presented at two points in the case, both in the form of information that is not directly causally relevant (i.e., incidental observation and prior conviction).

We hypothesized that participants would give higher ratings of guilt, blame, intentionality, and dispositional judgments, and lower ratings for credibility and situational judgments for bad defendants compared to good and neutral counterparts. Participants should also want to punish bad defendants more and be more likely to convict them of the crime. We predicted the opposite effect would occur in judgments for good defendants. We also predicted that changes in judgment would occur when the character manipulation is first introduced and persist as the study progressed.

Method

Participants and Design

An a priori power analysis found that 126 participants were required to achieve 90% power to detect a moderate effect size ($f = .25$) at an alpha level of 0.05. We recruited 153 participants from a UK sample on Prolific (paid £7.20 per hour). They were pre-screened for age (18-70), highest education level (secondary school/GCSE or above), fluency in English, and approval rate on the platform (95% or above).

Our final sample size was $N = 143$ ($M_{age} = 35.8$, $SD_{age} = 12.8$, $N_{female} = 115$, $N_{male} = 28$), after excluding nine participants from the analysis due to incorrect responses to the attention check and one for choosing the same score for every response.

The experiment had a mixed design, where we manipulated the defendant's character (Good, Bad, Neutral) between conditions, and collected participants' moral judgments at five points throughout the study as a repeated measure.

Materials and Procedure

We showed participants a simplified trial transcript based on a real theft case. A legal expert was consulted regarding the plausibility and presentation of the materials. Full materials are available at https://osf.io/vyx3q/.

The study was completed in Qualtrics. We randomly assigned participants to one of three conditions, where the defendant either had *Good character* (observed to have helped an elderly woman prior to the crime, and had no prior conviction), *Bad character* (observed to have behaved rudely to an elderly woman, and had prior convictions of domestic violence and disorderly conduct), or *Neutral character* (no mention of observed actions or prior conviction). The specific character descriptions were based on traits that people generally agree are morally desirable or undesirable, namely compassion (Helzer et al., 2014) and callousness and lack of consideration for others (Pizarro & Tannenbaum, 2012).

As part of the belief-updating paradigm (similar to those used in previous research, e.g., Monroe & Malle, 2019), evidence was presented sequentially and participants gave judgments at five time points ('stages') as they received more information about the case. Judgments included a *guilty/not guilty verdict*, *probability of guilt* (What is the probability that the defendant committed the crime?), *blameworthiness* (How much do you blame the defendant for the crime?), *intentionality* (How intentional do you think the defendant's actions were?), as well as ratings of *disposition* and *situation* (To what extent did the defendant commit the crime because he is a bad person / because of the situation he was in?), and *desire to punish* (How much do you want to punish the defendant?). Except for verdict, all judgments were made on a scale of 0% to 100%. At each subsequent stage, participants were shown their previous answer as an anchor.

Stage 1. Initial representations of the case included the prosecution's opening statement and indictment of the charge (theft). Participants made their judgments after reading this.

1259

Stage 2. First half of witness testimony given by the security guard working at the store. As the first part of the character manipulation, the guard was asked why he had noticed the defendant prior to the incident at the till. We varied his response in the scenario based on character condition.

Good vs. Bad character. 'I'd already noticed the defendant before he entered the store because I saw him helping an old lady with her bags [...being rude to an old lady who was blocking his way]. I remember thinking that he seemed to be a nice guy because not many people would have offered [...thinking he was impolite].'

Neutral. 'There just wasn't a lot of people in the store.'

Participants then gave judgments in two additional measures (in this and subsequent stages: *credibility* (How much do you believe that the witness is telling the truth?), and *importance of evidence* (How important was this piece of evidence in making your judgment?).

Stage 3. Second half of witness testimony, in line with facts stated in the prosecution's opening statement.

Stage 4. Defendant's testimony. As the second part of the character manipulation, here the defendant either (in the bad character condition) admitted to having been convicted for domestic violence and disorderly conduct in the past, or (in the good character condition) had no prior convictions. This was included considering it is a common form of 'character information' in legal contexts. The defendant then denied theft saying there had been no dishonest appropriation; he was in a rush and worried he had left a tap running at home.

Stage 5. In the closing statements, both pieces of character information were used in the arguments presented by the prosecution and defense (except in the neutral condition). Judge's directions also reminded jurors of their role in this trial and that character evidence should only be used in judging credibility. Participants gave their final verdict and judgments and described how they made their decision.

Results

Verdict

Of the 143 participants, only 22 gave the final verdict of not guilty (15.38%). Ten of the 22 not-guilty verdicts were given to good defendants (45.5%), seven to bad defendants (31.8%), and five to neutral defendants (22.7%).

Main Analysis of Judgment Measures

To explore the interaction between character and judgment stage, we conducted a mixed ANOVA for each judgment measure using the *afex* package in R (Singmann et al., 2022). To account for variance in baseline measures, we examined between-subject contrasts by comparing the difference between character conditions at each stage with the same difference at baseline (Stage 1) using Holm adjusted p-values. Within-subject contrasts of changes across stages were computed using the *emmeans* package (Lenth, 2022) with Tukey adjusted p-values.

Dispositional and Situational Attributions. For dispositional ratings (i.e., how much the defendant committed the crime because he is a bad person), we found significant main effects of character, $F(2, 140) = 7.22$, $p < .001$, and stage, $F(2.54, 355.85) = 4.45$, $p = .007$, and a significant interaction between character and stage, $F(5.08, 355.85) = 12.13$, $p < .001$. As seen in Figure 1, dispositional judgments increased for the bad defendant ($t(140) = 5.75$, $p < .001$) and decreased for the good defendant ($t(140) = 5.28$, $p < .001$) from baseline to Stage 2, when character information was introduced. The contrast between the two conditions was larger at Stage 2 than at baseline, $t(140) = 7.79$, $p < .001$.

This difference between character conditions persisted until Stage 5, where the difference between good and bad defendants was significantly larger compared to baseline, $t(140) = 7.394$, $p < .001$. (We also found significant differences between bad and neutral conditions, $t(140) = 5.146$, $p < .001$, and good and neutral conditions, $t(140) = 2.21$, $p = .029$.)

On the other hand, when participants were asked to give judgments on how situational the agent's behavior was, results only showed a significant main effect of stage $F(2.28, 319.60) = 8.79$, $p < .001$. No significant differences were found both between and within character conditions.
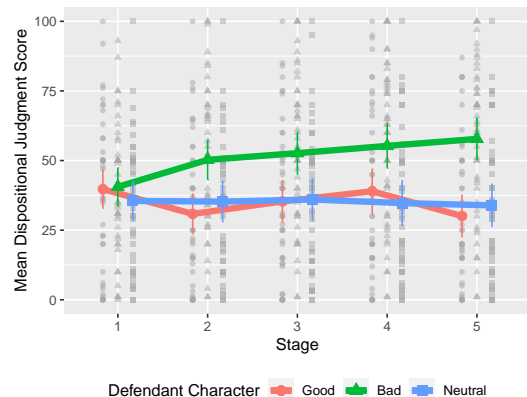

Figure 1: Dispositional judgments between character conditions and by stage. Error bars indicate 95% CI.

Punishment. We found consistent differences in punishment ratings between character conditions as the case progressed. There was a significant main effect of character, $F(2, 140) = 6.18$, $p = .003$, of stage, $F(2.09, 293.03) = 13.20$, $p < .001$, and a significant interaction of character and stage, $F(2.09, 293.03) = 2.58$, $p = .035$. More specifically, participants consistently wanted to punish bad defendants more and good defendants less.

As seen in Figure 2, the contrast between good and bad character conditions at Stage 2 was significantly larger than the same comparison at baseline, $t(140) = 3.992$, $p < .001$.

Differences between good and bad defendants also occurred at Stage 3, $t(140) = 2.48$, $p = .043$, and again at Stage 5, $t(140) = 3.36$, $p = .002$. There were no significant within-
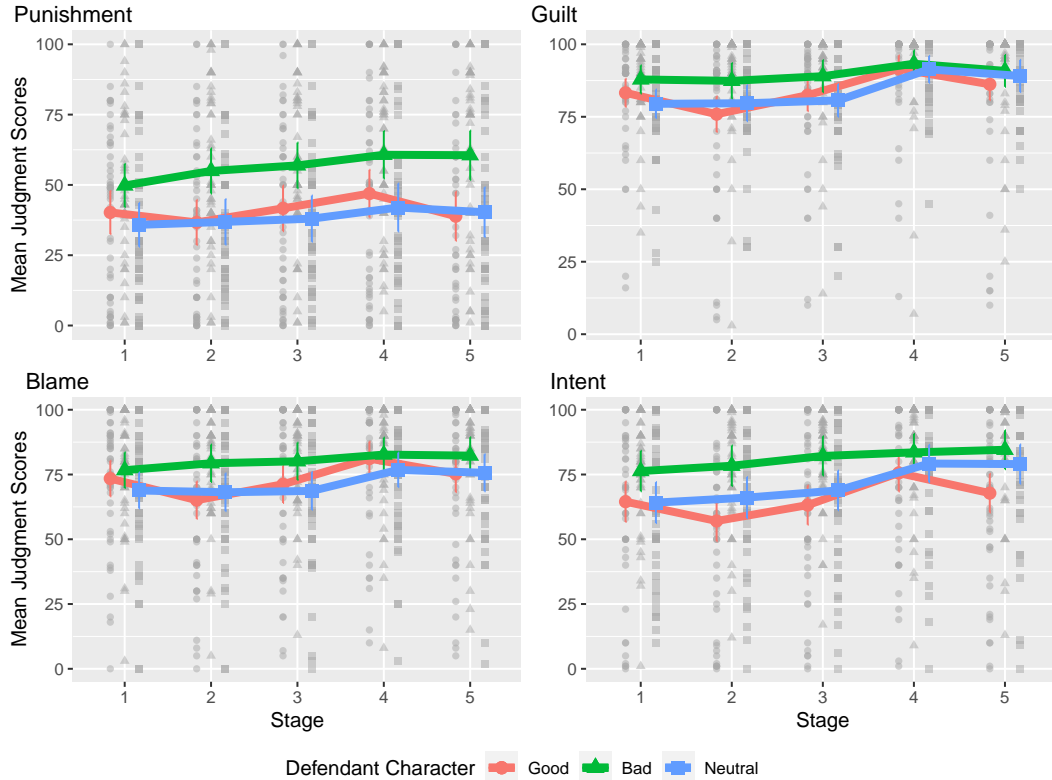
Figure 2: Judgments of punishment, guilt, blame, and intent between character conditions and by stage.

subject changes in punishment judgments, except for a decrease in ratings for the good defendant between Stages 4 and 5 ($t(140) = 5.78$, $p < .001$).

Guilt. The effect of character on guilt judgments was less prominent. We found a significant main effect of stage, $F(359.16, 189.54) = 23.27$, $p < .001$, and marginally significant interaction of character and stage, $F(359.16, 189.54) = 2.31$, $p = .043$. As seen in Figure 2, participants mitigated guilt judgments for the good defendant from baseline to Stage 2 ($t(140) = 3.54$, $p = .039$). This led to contrasts between character conditions to be significantly larger at Stage 2 than baseline for both good and bad defendants ($t(140) = 2.34$, $p = .041$) and good and neutral defendants ($t(140) = 2.56$, $p = .035$). At the later stages, judgments converged between the three character conditions.

Blame. Blame judgments were also less sensitive to information about the defendant's character. There was a significant main effect of stage, $F(2.40, 336.53) = 15.29$, $p < .001$, and a significant interaction of character and stage, $F(2.40, 336.53) = 2.47$, $p = .034$. As seen in Figure 2, participant mitigated blame ratings for the good defendant ($t(140) = 4.59$, $p < .001$), causing contrasts between character conditions to become significantly larger at Stage 2 compared to baseline for both good and bad defendants ($t(140) = 4.28$, $p < .001$) and good and neutral defendants ($t(140) = 2.97$, $p = .007$). No significant differences were found be-

tween conditions at subsequent stages, although blame judgments for the good defendant increased again from Stage 3 to Stage 4 ($t(140) = 3.86$, $p = .014$), thus converging with the other conditions.

Intent. Participants showed consistent differences in intent ratings between character conditions. There was a significant main effects of character, $F(2, 140) = 5.26$, $p = .006$, of stage, $F(2.23, 311.66) = 21.27$, $p < .001$, and their interaction, $F(2.23, 311.66) = 2.76$, $p = .023$. As seen in Figure 2, contrasts between character conditions were significantly larger at Stage 2 than baseline, for both good and bad defendants ($t(140) = 2.82$, $p = .016$) and the good and neutral defendants ($t(140) = 2.78$, $p = .006$). There were no such differences in the later stages, although judgments for the good defendant increased from Stage 3 to Stage 4 ($t(140) = 4.26$, $p = .003$), and then decreased again at Stage 5 ($t(140) = 4.42$, $p = .002$).

Credibility. We found no significant differences between mean credibility judgments for the witness and defendant across character conditions. However, the defendant was generally considered less credible than the witness, $F(1, 140) = 117.22$, $p < .001$.

Correlations between judgment measures. Table 1 shows the correlations between mean judgments for each measure. There were notable positive correlations between judgments of blame and guilt ($r = .73$, $p < .001$), blame and

intentionality ($r = .72$, $p < .001$), and guilt and intentionality ($r = .62$, $p < .001$). Punishment ratings were strongly correlated with dispositional attributions ($r = .71$, $p < .001$) and moderately correlated with blame ($r = .54$, $p < .001$). Dispositional attributions were moderately associated with blame ($r = .50$, $p < .001$) and intentionality ($r = .52$, $p < .001$), but not with situational attributions ($r = .06$).

Table 1: Correlational Matrix of Moral Judgments.

	Blame	Guilt	Int.	Disp.	Sit.	Pun.
Blame		.73***	.72***	.50***	.15***	.54***
Guilt	.73***		.62***	.38***	.19***	.45***
Int.	.72***	.62***		.52***	.11**	.54***
Disp.	.50***	.38***	.52***		.06	.71***
Sit.	.15***	.19***	.11**	.06		.06
Pun.	.54***	.45***	.54***	.71***	.06**	

* p<.05, ** p<.01, *** p<.001

Qualitative Analysis of Judgment Rationales

To further explore participants' decision-making process, we conducted a thematic analysis of their judgment rationales, which was an open response to the question 'please use a few lines to describe your reasoning.' The first author coded responses blind to character condition and developed a codebook in the process. A research assistant used the codebook to code the responses independently. They discussed any discrepancies in code application, which were resolved after reaching consensus. The first author then arranged post-agreement codes into different themes on how participants reasoned about the case. Across all codes, the average inter-coder reliability was 97.85%.

Theme 1: Facts of the case. Participants often included facts of the case in their judgments, such as a description of the defendant's actions, the evidence presented by the prosecution, and the defendant's admission to his actions. Some focused exclusively on the facts, remarking that the defendant's intentions or mental states did not matter (*'[His] intention did not sway my opinion [. . .] the act of leaving a store without paying for your items first is theft,'* P65).

Theme 2: Inference of defendant's mental states. Participants frequently made inferences about the defendant's mental states (e.g., intentions, knowledge, and emotions). While some were quick to establish strong intent, others suggested that his actions seemed less intentional and more impulsive, or chose to give him the benefit of the doubt (*'[he] made a gross error of judgment rather than set off to intentionally commit a crime,'* P94). Many considered the defendant's testimony when they reasoned about intentionality, though some found this to not be believable due to lack of evidence.

Participants also inferred about the defendant's thoughts and emotions, for example that he was not in a good state (*'preoccupied,'* P30; or *'under a lot of stress [or] exhaustion,'* P94). Some made inferences of knowledge, suggesting that the defendant acted knowingly (*'[he] was aware of his*

actions and the consequences,' P37), or that he knew what he was doing was wrong.

Theme 3: Counterfactual reasoning. The most frequently mentioned counterfactual was that the defendant could or should have done differently as there were many possible alternative courses of action (*'he could have just left [the items] at the till and returned to the store later,'* P16). Participants also considered how reasonable the defendant's actions were by comparing them to what a reasonable person would have done (*'he should and would have known that taking the items home without paying was theft,'* P92). Some used counterfactual reasoning in the defendant's favor, suggesting that his actions were situational (*'he may not have committed the offence if his shopping trip had run smoothly,'* P118).

Theme 4: Evaluation of defendant. Participants evaluated the defendant's credibility, with some pointing out that he had the motivation to lie. They also shared their attitudes toward the defendant, with some blaming his *'immoral'* (P77) actions, while some expressed empathy towards him and did not want him to be punished harshly, as this could (*'bring more detrimental consequences than good,'* P103).

Some participants also referenced the defendant's moral character. For the bad defendant, participants considered both his behavior and prior convictions (*'his attitude towards a customer and staff was also very negative and aggressive,'* P84). Some made dispositional judgments of his behavior (*'after finding out his domestic violence and disorderly conduct charges though, as well as the witness' statement on his character, it seemed more likely to be a moral flaw,'* P117). People made opposite remarks about the good defendant (*'he seemed like a kind helpful person,'* P30). However, good character did not always mitigate judgments (*'just because he helped a lady with her bags doesn't mean he's incapable of making a poor decision and stealing,'* P55).

Discussion

This study set out to explore the influence of the defendant's character in juror decision-making, specifically whether incidental character information would affect judgments in a relatively unambiguous case. We hypothesized that bad defendants would be judged more harshly than good or neutral counterparts, receiving higher ratings of guilt, blame, intentionality, and dispositional attributions, and jurors would be more likely to give a guilty verdict and report stronger desires to punish. These changes in judgments would occur when incidental character information is first introduced and persist as the case progressed.

These predictions were partially supported by our findings: character information affected people's judgments of blame, guilt, and intentionality immediately after it was presented. At first glance, it seems concerning that such simple, irrelevant information could affect decision-making in a straightforward case. However, character did not affect final judgments of guilt and blame or shift final verdicts. In other words, while good and bad defendants were blamed differ-

ently immediately after the presentation of character information, differences in judgments eventually converged across conditions such that initial evaluations did not dominate subsequent judgments – participants tended to correct for this bias as more information became available.

One key novel finding was that punishment ratings were much more sensitive to incidental character information than judgments of blame and guilt, and differences between character conditions persisted throughout the study. Participants consistently judged bad defendants to be more deserving of punishment and were more likely to attribute their actions to dispositional factors throughout the study, even when they were not blamed differently than good or neutral defendants. This has significant implications on the broader context of moral judgment as it explores the question of why we punish. It makes pragmatic sense to punish repeat offenders for the sake of deterrence and social desirability – or do we punish bad people simply because we want them to suffer? Prior work has found that people largely subscribe to punishment for the purpose of retributive justice (Carlsmith & Darley, 2008; Carlsmith et al., 2002), or because they desire revenge and equality (Bone & Raihani, 2015).

Our findings partially support the person-centered approach on moral judgment (Uhlmann et al., 2015), mainly how people focus on information that is diagnostic of character even when it is not directly relevant to the action under judgment. While bad character had limited effects on final judgments of guilt and blame, it strongly influenced dispositional judgments, suggesting that people used character information to build a picture of the agent and integrate this into their reasoning, especially when details are not yet clear. However, these evaluations did not necessarily translate to final judgments of blame and guilt.

Thus, contrary to blame-early models (e.g., Alicke, 2000), participants did not seem especially motivated by a desire to blame, nor did they engage in a blame-validation mode of processing. Had that been the case, we would have found a larger distinction between character conditions at Stage 4, as information about a bad defendant's prior conviction would have exacerbated pre-existing blame. Since we found no between-subject differences near the end of the study, this suggests that people focused on the evidence and were less influenced by character as more information came to light. This is in line with recent research showing that people draw inferences from circumstantial evidence and incorporate them into the reasoning process (Monroe & Malle, 2017).

Furthermore, we found that participants were more sensitive to information about good character – the differences between character conditions in judgments of blame and guilt were primarily propelled by mitigating judgments for good defendants, while judgments for bad defendants were more stable. Therefore, rather than engaging in blame validation, we suggest that people expected a certain degree of 'badness' inherent in criminal acts, and only mitigated judgments when they had reason to (i.e., when the defendant was good).

Our qualitative findings also support the observation that participants focused on drawing inferences about actions, norms, and the defendant's mental states when reasoning about blame and punishment. Notably, they also used counterfactual reasoning to consider whether the defendant was causally and morally responsible for his actions and whether they were reasonable. These counterfactuals also mostly focused on mutating actions rather than mental states, possibly because the defendant had some level of knowledge and intent behind his actions (Gilbert et al., 2015; Kirfel & Lagnado, 2021). However, the role of character is still ambiguous – only some participants mentioned that character played a role in their reasoning.

This study has several limitations. First, though not statistically significant, there was some variance in participants' baseline judgments. We had accounted for this in the analysis by comparing the interaction between each stage and baseline for between-subject contrasts. Second, we only showed how incidental character information could influence judgments when it was presented early in the case. It remains unclear whether it would affect judgments in the same way once more information was made available. It is also unclear whether other kinds of character information (e.g., prior conviction) would have the same effect when presented earlier. Furthermore, since Stage 4 included both information about prior conviction and the defendant's testimony, we were unable to disentangle their effects on judgments.

Finally, we had chosen a relatively unambiguous case as we wanted to construct a situation that limited the type of inferences participants made, but this induced a ceiling effect in guilt judgments and an overwhelming proportion of guilty verdicts. Thus, we were unable to determine whether character influences jurors' verdicts. In a future study, we aim to reduce this ceiling effect by varying the ambiguity of the case or the defendant's intentions and behavior. Based on existing theories (Bavli, 2022), we predict that the more ambiguous the situation, the more likely jurors are to use character information when drawing inferences about the defendant's mental states and behavior.

Information about moral character plays an important role in how people reason about and attribute blame. Our findings extend present work by showing that incidental character information can influence mock jurors' judgments in a legal case, but this effect diminishes for judgments of blame and guilt as more information about the case becomes available, suggesting that these inferences are often used reasonably. Notably, we found that people tended to mitigate blame for good defendants rather than exacerbate them for bad defendants. We also found that people wanted to punish bad defendants more than good or neutral defendants even when they did not blame them more. Future research on how people evaluate and use character in their judgments is needed to extend our knowledge on the mechanisms underlying different kinds of moral reasoning, both in juror decision-making and in everyday contexts.

Acknowledgments

References

Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology*, *63*(3), 368–378.

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, *126*(4), 556–574.

Anderson, B. J. (2012). Recognizing character: A new perspective on character evidence. *Yale Law Journal*, *121*, 1912–1968.

Bavli, H. J. (2022). An aggregation theory of character evidence. *Journal of Legal Studies*, *51*(1), 39–62.

Bavli, H. J. (2023). Character evidence as a conduit for implicit bias. *U.C. Davis Law Review*, *56*, 1019–1084.

Bone, J. E., & Raihani, N. J. (2015). Human punishment is motivated by both a desire for revenge and a desire for equality. *Evolution and Human Behavior*, *36*(4), 323–330. https://doi.org/10.1016/j.evolhumbehav.2015.02.002

Bos, M. W., & Dijksterhuis, A. (2011). Unconscious thought works bottom-up and conscious thought works top-down when forming an impression. *Social Cognition*, *29*(6), 727–737. https://doi.org/10.1521/soco.2011.29.6.727

Brambilla, M., Carraro, L., Castelli, L., & Sacchi, S. (2019). Changing impressions: Moral character dominates impression updating. *Journal of Experimental Social Psychology*, *82*, 64–73. https://doi.org/10.1016/j.jesp.2019.01.003

Carlsmith, K. M., & Darley, J. M. (2008). Psychological aspects of retributive justice. *Advances in Experimental Social Psychology*, *40*, 193–236. https://doi.org/10.1016/S0065-2601(07)00004-4

Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish? deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology*, *83*(2), 284–299. https://doi.org/10.1037//0022-3514.83.2.284

Eisenberg, T., & Hans, V. P. (2009). Taking a stand on taking the stand: The effect of a prior criminal record on the decision to testify and on trial outcomes. *Cornell Law Review*, *94*, 1353–1390.

Gilbert, E. A., Tenney, E. R., Holland, C. R., & Spellman, B. A. (2015). Counterfactuals, control, and causation: Why knowledgeable people get blamed more. *Psychological Bulletin*, *41*(5), 643–658. https://doi.org/10.1177/0146167215572137

Givelber, D., & Farrell, A. (2008). Judges and juries: The defense case and differences in acquittal rates. *Law Social Inquiry*, *33*(1), 31–52.

Gromet, D. M., Goodwin, G. p., & Goodman, R. A. (2016). Pleasure from another's pain: The influence of a target's hedonic states on attributions of immorality and evil. *Personality and Social Psychology Bulletin*, *42*(8), 1077–1091. https://doi.org/10.1177/0146167216651408

Helzer, E. G., Furr, R. M., Hawkins, A., Barranti, M., Blackie, L. E., & Fleeson, W. (2014). Agreement on the perception of moral character. *Personality and Social Psychology Bulletin*, *40*(12), 1698–1710. https://doi.org/10.1177/0146167214554957

Hughes, J. S. (2017). In a moral dilemma, choose the one you love: Impartial actors are seen as less moral than partial ones. *British Journal of Social Psychology*, *56*(3), 561–577. https://doi.org/10.1111/bjso.12199

Hughes, J. S., & Trafimow, D. (2012). Inferences about character and motive influence intentionality attributions about side effects. *British Journal of Social Psychology*, *51*(4), 661–673. https://doi.org/10.1111/j.2044-8309.2011.02031.x

Hughes, J. S., & Trafimow, D. (2015). Mind attributions about moral actors: Intentionality is greater given coherent cues. *British Journal of Social Psychology*, *54*(2), 220–235. https://doi.org/10.1111/bjso.12077

Kirfel, L., & Lagnado, D. (2021). Changing minds — epistemic interventions in causal reasoning. *Cognition*, *212*, 104721. https://doi.org/10.1016/j.cognition.2021.104721

Lagnado, D. A., & Gerstenberg, T. (2017). Causation in legal and moral reasoning. In *Oxford handbook of causal reasoning* (pp. 565–602). Oxford University Press.

Lenth, R. V. (2022). *Emmeans: Estimated marginal means, aka least-squares means* [R package version 1.8.1-1]. https://CRAN.R-project.org/package=emmeans

Monroe, A. E., & Malle, B. F. (2017). Two paths to blame: Intentionality directs moral information processing along two distinct tracks. *Journal of Experimental Psychology: General*, *146*(1), 123–133. https://doi.org/10.1037/xge0000234

Monroe, A. E., & Malle, B. F. (2019). People systematically update moral judgments of blame. *Journal of Personality and Social Psychology*, *116*(2), 215–256. https://doi.org/10.1037/pspa0000137

Nadler, J. (2012). Blaming as a social process: The influence of character and moral emotion on blame. *Law and Contemporary Problems*, *75*(2), 1–31.

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

Ogloff, J. R., & Rose, V. G. (2005). The comprehension of judicial instructions. In N. Brewer & K. D. Williams (Eds.), *Psychology and law: An empirical perspective* (pp. 407–444). Guilford Press.

Pizarro, D. A., & Tannenbaum, D. (2012). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In *The social psychology of morality: Exploring the causes of good and evil* (pp. 91–108). American Psychological Association. https://doi.org/https://doi.org/10.1037/13091-005

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2022). *Afex: Analysis of factorial experi-*

ments [R package version 1.1-1]. https://CRAN.R-project. org/package=afex

Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science*, *10*(1), 72–81. https://doi.org/10. 1177/1745691614556679

Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. Guilford Press.

Yankah, E. N. (2004). Good guys and bad guys: Punishing character, equality and the irrelevance of moral character to criminal punishment. *Cardozo Law Review*, *25*, 1019–1067.

Young, L., & Tsoi, L. (2013). When mental states matter, when they don't, and what that means for morality. *Social and Personality Psychology Compass*, *7*(8), 585–604. https://doi.org/10.1111/spc3.12044