

To appear in: *Journal of Experimental Social Psychology*, vol. 97

Citation: Walker, A. C., Turpin, M. H., Fugelsang, J. A., & Białek, M. (2021). Better the two devils you know, than the one you don't: Predictability influences moral judgments of immoral actors. *Journal of Experimental Social Psychology*, 97, 104220.

<https://doi.org/10.1016/j.jesp.2021.104220>

Better the Two Devils You Know, Than the One You Don't: Predictability Influences Moral
Judgments of Immoral Actors

Alexander C. Walker^{a,*}, Martin Harry Turpin^{a,*}, Jonathan A. Fugelsang^a, and Michał Białek^b

^a Department of Psychology, University of Waterloo, Waterloo, ON, Canada

^b Institute of Psychology, University of Wrocław, Wrocław, Poland

Author Note

All materials and data reported in this manuscript can be accessed via the following link:

<https://osf.io/tc73f/>.

We would like to thank Dr. Igor Grossmann for his insightful feedback regarding various aspects of this project, including his helpful comments on previous versions of this manuscript. We would also like to thank Dr. Marlene Henderson (Associate Editor) and two anonymous reviewers for their helpful comments during the review process.

This research was supported by grants from The Natural Sciences and Engineering Research Council of Canada.

Correspondence concerning this article should be addressed to Alexander C. Walker, Department of Psychology, University of Waterloo, Waterloo, ON, Canada N2L 3G1.

*Notes equal contribution from authors

Email: a24walke@uwaterloo.ca

Abstract

Across six studies ($N = 2,646$), we demonstrate the role that perceptions of predictability play in judgments of moral character, finding that people demonstrate a moral preference for more predictable immoral actors. Participants judged agents performing an immoral action (e.g., assault) for an unintelligible reason as less predictable and less moral than agents performing the same immoral action, along with an additional immoral action (e.g., theft), for a well-understood immoral reason (Studies 1-4). Additionally, agents performing an immoral action for an unintelligible reason were judged as less predictable and less moral compared to agents performing the same immoral act for an unstated reason (Studies 3-5). This moral preference persisted when participants viewed video footage of each agent's immoral action (Study 5). Finally, agents performing immoral actions in an unusual way were judged as less predictable and less moral than those performing the same actions in a more common manner (Study 6). The present research demonstrates how immoral actions performed without a clear motive or in an unpredictable way are perceived to be especially indicative of poor moral character. In revealing peoples' moral preference for predictable immoral actors, we propose that perceptions of predictability play an important, yet overlooked, role in judgments of moral character. Furthermore, we propose that predictability influences judgments of moral character for its ultimate role in reducing social uncertainty and facilitating cooperation with trustworthy individuals and discuss how these findings may be accommodated by person-centered theories of moral judgment and theories of morality-as-cooperation.

Keywords: predictability, moral character, cooperation, person-centered moral judgments, social uncertainty

1. Introduction

“The oldest and strongest emotion of mankind is fear, and the oldest and strongest kind of fear is fear of the unknown” – H. P. Lovecraft (1927)

What makes for a terrifying villain? It can be scary to imagine a person who is driven by extreme greed, or a lust for power, who is willing to do any amount of harm if it pushes them toward achieving their singular goal. However, this sort of person can be placated, bargained with, and most importantly have their actions predicted. Scarier may be the person whose motives are impossible to understand. *No Country for Old Men* (Coen & Coen, 2007) presents the character of Anton Chigurh, a ruthless hitman who at times decides to kill a random person based on the outcome of a coin toss. What makes a character like this seem so immoral and frightening is that such characters cannot be negotiated with and their actions cannot be predicted. Villains like Chigurh appear prominently in fiction, suggesting that people carry the intuition that these unpredictable agents are frightening, immoral, and ought to be avoided. Why might an intuition to avoid unpredictable agents exist?

Evaluating the character of others is important for our social well-being. Individuals able to reliably assess the character of others are able to avoid those most likely to cause them harm (e.g., cheaters and psychopaths), while also benefitting from cooperation with trustworthy individuals. Person-centered theories of moral judgment place character evaluation as fundamental to judgments of morality (Pizarro & Tannenbaum, 2011; Uhlmann et al., 2015). From this perspective, moral judgments are less about the actions of individuals, and more a product of what these actions reveal about a person’s underlying character. Recent work suggests that, contrary to traditional act-based theories (e.g., deontology and utilitarianism; Kant, 1785/1959; Mill, 1861/1998), people view actions not as the endpoint of moral evaluation, but as

signals revealing the moral character of those who perform them (Tannenbaum et al., 2011; Uhlmann et al., 2015). For example, people often distinguish between the morality of an act and the moral character of the person performing the act (Cushman, 2015; Uhlmann & Zhu, 2014; Uhlmann et al., 2014), at times judging an act as morally permissible while simultaneously judging those who perform it as possessing a poor moral character (Uhlmann et al., 2013). Similarly, past work demonstrates instances in which people judge the moral character of those engaging in harmless-but-offensive acts (e.g., masturbating into a dead chicken) more harshly than those engaging in harmful actions (e.g., theft; Uhlmann & Zhu, 2014). While initially puzzling from a harm-based perspective, such judgments are entirely reasonable if such harmless-but-offensive acts are thought to reveal an unstable and anti-social person likely to cause significant harm in the future.

1.1 Perceptions of Predictability and Judgments of Moral Character

While research has revealed various factors relevant to judgments of blame and morality (e.g., intentionality and harm; Alicke, 2000; Gray et al., 2012; Schein & Gray, 2018; Malle et al., 2014), the role that perceptions of predictability play in peoples' assessments of moral character has received considerably less attention (although see Turpin et al., 2021). From a person-centered perspective, there are many reasons why people may judge unpredictable individuals as especially immoral, specifically when these individuals are responsible for causing harm. For example, immoral actions (e.g., assault) done in a seemingly random or unpredictable manner may be viewed as highly indicative of a deficient moral character lacking in key character traits (e.g., empathy). In contrast, the same immoral actions performed for well-understood reasons (e.g., to escape punishment) may be viewed more as matters of circumstance and thus thought to be less revealing of a person's underlying moral character. Furthermore, clear motives can imply

that an offender has a morally neutral goal (e.g., monetary gain) as opposed to one that is highly immoral (e.g., sadistic satisfaction). As such, people may be more likely to interpret harm caused by agents with well-understood motives as indirect (i.e., as a means to an end), as opposed to a consequence of a highly immoral character. Past work has found that people view indirect harms more favorably than direct harms (Royzman & Baron, 2002; Waldmann & Dieterich, 2007) and unintentional harms more favorably than intentional harms (Ames & Fiske, 2013). Therefore, one may prefer more (vs. less) predictable immoral agents on the basis that their actions may be viewed as more indirect and less intentional.

Similarly, predicting the future behavior of unpredictable individuals, acting seemingly without cause, can seem like an impossible task. Being unable to assess a person's future behavior may naturally evoke feelings of social uncertainty. Past research demonstrates peoples' aversion to uncertainty, both within social (Chater & Loewenstein, 2016; FeldmanHall & Shenhav, 2019; Hogg, 2000) and non-social domains (Gneezy et al., 2006; Tversky & Shafir, 1992; van Dijk & Zeelenberg, 2003). Therefore, the difficulty in assessing how an unpredictable person will behave is likely to be experienced as unpleasant and frightening. This may be especially true in cases in which an unpredictable individual is known to have behaved immorally in the past, perhaps in a seemingly random or motiveless way. Overall, the uncertainty surrounding less (vs. more) predictable agents may result in a preference for predictable immoral agents performing calculated immoral actions for well-understood reasons.

1.2 Morality-as-Cooperation: Unpredictable Agents as Bad Co-operators

There is good reason to believe that the sense for morality developed for the purpose of ensuring cooperation among conspecifics (Curry, 2016; Curry et al., 2019; Greene, 2013; Haidt, 2012; Rai & Fiske, 2011; Tomasello & Vaish, 2013), as reflected in theories of morality-as-

cooperation. Being able to easily assess and understand one's motivations facilitates cooperation (Paal & Bereczkei, 2007; Young & Tsoi, 2013; Curry, 2016). When you can assess an individual's motivations you have the choice of cooperating with them when your goals align and are able to avoid conflict when your goals diverge. A person who is ruthlessly motivated toward a singular goal—even if willing to harm others—can usually be relied upon to not engage in senseless harms not in service of their goal. In contrast, an unpredictable agent may betray people at any moment without cause. This second agent represents a poor candidate for cooperating with long-term. If morality is fundamentally underpinned by the need to cooperate, it follows that unpredictable immoral agents with unintelligible motives may be perceived as especially immoral.

1.3 The Present Research

The present research investigates the role perceptions of predictability play in judgments of moral character. In Studies 1-5, we examined peoples' judgments of immoral actors physically assaulting a stranger for either an unintelligible reason, a well-understood immoral reason, or an unstated reason (Studies 3-5). Additionally, in Study 6, participants judged agents performing various immoral actions in either a common (e.g., "Gerald hits a stranger with his fist") or unusual manner (e.g., "Gerald hits a stranger with a frozen fish"). Overall, we hypothesized that participants would display a moral preference for more predictable immoral actors such that agents signalling unpredictability with their actions, either by acting without an intelligible motive or by performing an immoral act in an unusual manner, would be seen as possessing an especially poor moral character. Specifically, in Study 1, we predicted that actors physically assaulting a stranger for an unintelligible reason would be judged as less predictable

and less moral than actors performing the same assault, along with an additional immoral act (e.g., theft), for a well-understood immoral reason.

2. Study 1

3. Method

3.1 Participants

Two hundred participants (35% Female; $M_{\text{age}} = 36.45$, $SD_{\text{age}} = 10.92$; 4% Asian-American; 7% African-American; 88% European-American; 2% mixed ethnicity) were recruited from Amazon Mechanical Turk. Participants received \$1.00 upon completion of an 8-minute online questionnaire for which they were required to be residents of the United States and possess a Mechanical Turk HIT approval rate greater than or equal to 95%. For all studies, we collected our full sample prior to data analyses and report all data exclusions, all manipulations, and all measures used. All studies received approval by a University of Waterloo Research Ethics Committee.

3.2 Materials

To test whether a moral premium exists for predictable agents, we created three vignettes, all of which described a person physically assaulting a stranger. These vignettes differed with regards to the actor's motivations for engaging in the physical assault (see Table 1). Notably, these vignettes were created such that the person who engaged in physical assault for an immoral reason was described as doing objectively more harm (e.g., robbing a bank and assaulting another person) compared to those engaging in physical assault for a moral reason or no reason (i.e., only physical assault). Study 1 featured three different moral and immoral reason vignettes, all of which were constructed similarly yet featured different scenarios (e.g., escape after robbing a bank versus escape from jail). In contrast, our no reason vignette was the same for all

participants. All materials presented in the current study can be viewed in the supplementary materials (Part A).

Table 1

Study 1: Vignette Examples

Motive	Example
No Reason	“Gerald physically assaults Robert for no reason (i.e., Gerald did not benefit in any way from assaulting Robert, who was unknown to him).”
Immoral Reason	“Richard physically assaults Michael in order to escape a bank he had just robbed for \$50,000.”
Moral Reason	“James physically assaults Kevin in order to protect an innocent person who was being assaulted by Kevin.”

Note. Names of the attacker and victim were randomized on each trial. Items were presented in a randomized order.

3.3 Measures

Following the presentation of each vignette, participants were asked to judge the person acting within the vignette on various moral dimensions presented in a randomized order.

3.3.1 Moral Perception. Participants’ moral perception of each hypothetical actor was assessed by calculating their mean rating of each actor on four different moral dimensions (i.e., Goodness, Morality, Peacefulness, and Empathy). Participants judged each actor on each dimension using a 7-point scale with labels surrounding the endpoints of the scale (i.e., Bad/Good, Immoral/Moral, Violent/Peaceful, Merciless/Empathetic). The reliability of this composite was calculated for each item type (No Reason: $\alpha = .88$; Immoral Reason: $\alpha = .92$; Moral Reason: $\alpha = .78$) demonstrating good reliability.

3.3.2 Predictability. Participants judged the predictability of each actor using a 7-point scale that ranged from “*Unpredictable*” to “*Predictable*.”

3.3.3 Intentionality. Participants judged how intentional they viewed the actions of each actor using a 7-point scale that ranged from “*Acted Unintentionally*” to “*Acted Intentionally*.”

3.3.4 Punishment. Participants indicated how much punishment they felt each actor deserved using a 7-point scale that ranged from “*Deserves No Punishment*” to “*Deserves Severe Punishment*.”

3.3.5 Harm. Participants indicated how much harm they felt each actor had caused using a 7-point scale that ranged from “*Caused No Harm*” to “*Caused a Great Deal of Harm*.”

3.4 Design and Procedure

We used a within-subjects design in which all participants were presented with vignettes in which a person physically assaulted a stranger for a moral reason, an immoral reason, and for seemingly no reason (see Table 1). Following the presentation of each vignette, participants evaluated the person acting within the vignette on various moral dimensions (see Measures). Following all moral evaluations, participants were reminded of the no reason vignette and were asked whether they believed that the person acting in this vignette did in fact physically assault someone for no reason or if they made their judgment of this person with an imagined reason in mind.

3.5 Data Preparation

We excluded participants who stated that they imagined a reason for the physical assault described in our no reason item ($n = 32$). This exclusion criterion did not change the interpretation of any significance tests, with the exception that judgments of intentionality no

longer differed between items¹. Note that, for each study applying an identical or similar exclusion criterion (i.e., Studies 1-5), the results of all primary analyses without such exclusions applied can be viewed in the supplementary materials (Part B).

4. Results and Discussion

Participants' mean predictability and moral perception judgments for both no reason and immoral reason items can be viewed in Figure 1. We conducted repeated-measures ANOVAs comparing participants' predictability, moral character, intentionality, punishment, and harm judgments across item type (no reason vs. immoral reason vs. moral reason). Sensitivity power analyses² using G*Power (Faul et al., 2007) indicated 80% power to detect a minimum effect size of $\eta_p^2 = .028$ for the conducted repeated-measures ANOVAs and $d = 0.22$ for follow-up paired samples *t*-tests. Predictability was shown to vary across item type, $F(2,334) = 248.01, p < .001, \eta_p^2 = .598$. As expected, actors engaging in violence for an immoral reason were judged as more predictable ($M = 2.64, 95\% CI = [2.38, 2.90]$) compared to those engaging in violence for seemingly no reason ($M = 1.46, 95\% CI = [1.32, 1.60]$), $t(167) = 8.60, p < .001, d = 0.84$. Moral perceptions also varied across item type, $F(2, 334) = 723.98, p < .001, \eta_p^2 = .813$. Consistent with our hypothesis, actors engaging in violence for an immoral reason were perceived as more moral ($M = 1.65, 95\% CI = [1.51, 1.79]$) compared to those engaging in violence for seemingly no reason ($M = 1.39, 95\% CI = [1.28, 1.50]$), $t(167) = 3.93, p < .001, d = 0.31$.

We observed no difference in judgments of intentionality across item type, $F(2, 334) = 2.65, p = .072, \eta_p^2 = .016$. In contrast, judgments of punishment did vary across item type, $F(2,$

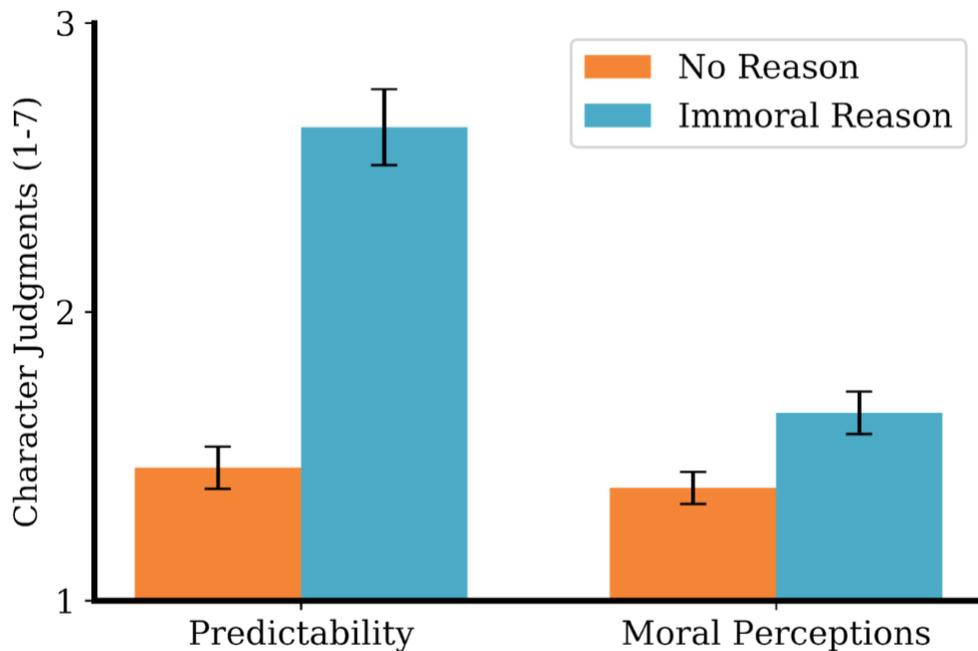
¹ When analyzing our full sample judgments of intentionality varied across item type, $F(2, 398) = 4.38, p = .013, \eta_p^2 = .022$, with actors engaging in violence for an immoral reason being judged as acting more intentionally ($M = 6.30, 95\% CI = [6.12, 6.47]$) than those engaging in violence for seemingly no reason ($M = 5.94, 95\% CI = [5.69, 6.18]$), $t(199) = 2.90, p = .004, d = 0.23$.

² All sensitivity power analyses presented in this manuscript were calculated based on the design used and the sample size analyzed. Additionally, these power analyses were calculated with the assumption of a two-tailed test and an $\alpha = .05$.

334) = 708.96, $p < .001$, $\eta_p^2 = .809$. Nevertheless, this effect was a result of participants unsurprisingly judging actors engaging in violence for a moral reason as less deserving of punishment. Critically, judgments of punishment did not differ between immoral reason ($M = 6.24$, 95% $CI = [6.08, 6.41]$) and no reason ($M = 6.27$, 95% $CI = [6.11, 6.43]$) items, $t(167) = 0.40$, $p = .690$, $d = 0.03$. Lastly, we found that perceptions of harm varied across item type, $F(2, 334) = 214.98$, $p < .001$, $\eta_p^2 = .563$. Despite being described as performing an additional immoral act (e.g., theft), actors engaging in violence for an immoral reason were perceived as causing *less* harm ($M = 6.01$, 95% $CI = [5.83, 6.19]$) compared to those engaging in violence for no reason ($M = 6.30$, 95% $CI = [6.15, 6.45]$), $t(167) = 3.22$, $p = .002$, $d = 0.26$.

Figure 1

Study 1: Judgments of Predictability and Moral Perceptions



Note. Error bars represent +/- 1 standard error.

4.1 Associations between Key Variables

We examined the zero-order associations between key variables within our two critical items: No Reason and Immoral Reason (see Table 2). Consistent with our hypothesis, predictability was positively associated with moral perceptions within both items. Therefore, as participants perceived agents as more predictable they also tended to perceive them as more moral.

Table 2

Study 1: Pearson's Zero-order Correlations

	1	2	3	4	5
1. Predictability	-	-.14	.02	-.16*	.28***
2. Harm	-.39***	-	.27***	.71***	-.64***
3. Intentionality	-.19*	.35***	-	.30***	-.16*
4. Punishment	-.33***	.47***	.27***	-	-.70***
5. Moral Perceptions	.69***	-.44***	-.27***	-.38***	-

Note. Pearson correlations ($N = 168$). Bottom diagonal = No Reason Item. Top diagonal = Immoral Reason Item. *** $p < .001$, ** $p < .01$, * $p < .05$.

4.2 Summary

The results of Study 1 were consistent with the hypothesized moral preference for more predictable immoral actors. Participants judged actors assaulting a stranger for a well-understood immoral reason as more predictable, more moral, and as having caused less harm than actors assaulting a stranger for seemingly no reason. Interestingly, this moral preference did not result from the immoral actions of more predictable agents appearing less intentional, nor did this moral preference result in more predictable agents being viewed as deserving of less punishment. However, the finding that both no-reason and immoral-reason actors were judged to be equally

deserving of punishment may itself reflect a preference for more predictable agents, on account of immoral-reason actors performing an additional immoral act (e.g., a bank robbery). Consistent with this view, we observed a negative association between judgments of predictability and punishment within both no reason and immoral reason items (see Table 2).

5. Study 2

Study 1 demonstrated participants' moral preference for predictable immoral agents acting with clear motives, despite these agents being described as performing an additional immoral act. Nevertheless, this preference could have been driven by participants imagining a more vicious assault when performed without a well-understood motive. In Study 2, we attempted to rule out this possibility by replicating the findings of Study 1 while describing each physical assault, along with its consequences, unambiguously.

6. Method

6.1 Participants

Two hundred US residents (38.5% Female; $M_{\text{age}} = 37.33$, $SD_{\text{age}} = 12.32$; 7% Asian-American; 9% African-American; 81% European-American; 3% mixed ethnicity) were recruited from Amazon Mechanical Turk using the same recruitment criteria as Study 1. Participants received \$0.75 upon completion of a 7-minute online questionnaire. Those who participated in Study 1 were restricted from participating in Study 2. This experiment was preregistered through Open Science Framework (<https://osf.io/bevtp>).

6.2 Materials and Measures

The materials and measures used in Study 2 were identical to those used in Study 1, with two exceptions. First, all vignettes were slightly modified such that each physical assault was described more concretely (i.e., as a punch that broke someone's jaw; see Table 3). Second, we

no longer asked participants to assess how much punishment each actor deserved. Thus, in Study 2, participants evaluated each actor described in our vignettes on our remaining four measures (moral perceptions, predictability, intentionality, and harm).

Table 3

Study 2: Vignette Examples

Motive	Example
No Reason	“Gerald punches Robert for no reason (i.e., Gerald did not benefit in any way from punching Robert, who was unknown to him) breaking Robert’s jaw.”
Immoral Reason	“Richard punches Michael (breaking Michael’s jaw) in order to escape a bank he had just robbed for \$50,000.”
Moral Reason	“James punches Kevin (breaking Kevin’s jaw) in order to protect an innocent person who was being assaulted by Kevin.”

Note. Names of the attacker and victim were randomized on each trial. Items were presented in a randomized order.

6.3 Design and Procedure

Study 2 featured an identical design and procedure to that of Study 1.

6.4 Data Preparation

Consistent with our pre-registered intention, we excluded participants who stated that they imagined a reason for the physical assault described in our no reason item ($n = 24$). We also removed one participant from our analyses that selected the same response option for all judgments. Note that the interpretation of all significance tests remains the same when analyzing our full sample (see supplementary materials Part B).

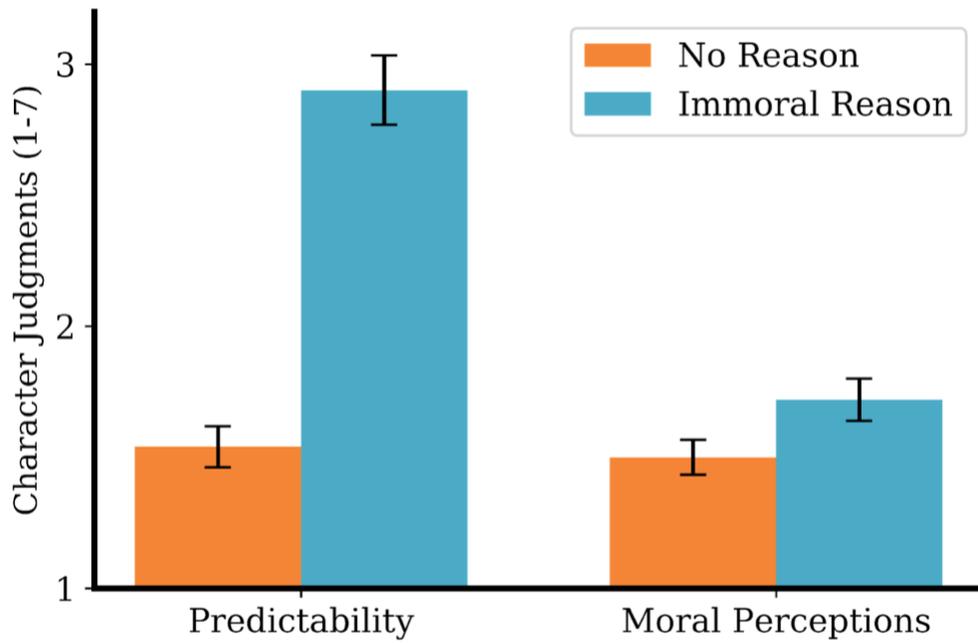
7. Results and Discussion

Participants’ mean predictability and moral perception judgments for both no reason and immoral reason items can be viewed in Figure 2. We conducted repeated-measures ANOVAs comparing participants’ predictability, moral character, intentionality, and harm judgments

across item type (no reason vs. immoral reason vs. moral reason). Sensitivity power analyses indicated 80% power to detect a minimum effect size of $\eta_p^2 = .027$ for the conducted repeated-measures ANOVAs and $d = 0.21$ for follow-up paired samples t -tests. Predictability was found to vary across item type, $F(2, 348) = 196.33, p < .001, \eta_p^2 = .530$. As expected, actors engaging in violence for an immoral reason were judged as more predictable ($M = 2.90, 95\% CI = [2.64, 3.16]$) compared to those engaging in violence for an unintelligible reason ($M = 1.54, 95\% CI = [1.39, 1.69]$), $t(174) = 10.90, p < .001, d = 0.91$. Moral perceptions also varied across item type, $F(2, 348) = 582.58, p < .001, \eta_p^2 = .770$. Replicating the results of Study 1, actors engaging in violence for an immoral reason were perceived as more moral ($M = 1.72, 95\% CI = [1.56, 1.88]$) compared to those engaging in violence for seemingly no reason ($M = 1.50, 95\% CI = [1.37, 1.63]$), $t(174) = 3.34, p = .001, d = 0.22$. We observed no difference in judgments of intentionality across item type, $F(2, 348) = 1.57, p = .210, \eta_p^2 = .009$. Conversely, perceptions of harm did vary across item type, $F(2, 348) = 118.98, p < .001, \eta_p^2 = .406$. As in Study 1, actors engaging in violence for an immoral reason were perceived as causing *less* harm ($M = 6.18, 95\% CI = [6.02, 6.34]$) compared to those engaging in violence for no reason ($M = 6.39, 95\% CI = [6.28, 6.50]$), $t(174) = 3.02, p = .003, d = 0.22$, despite being described as performing an additional immoral act.

Figure 2

Study 2: Judgments of Predictability and Moral Perceptions



Note. Error bars represent +/- 1 standard error.

7.1 Associations Between Key Variables

We examined the zero-order associations between key variables within our two critical items: No Reason and Immoral Reason (see Table 4). As in Study 1, judgments of predictability and moral character were positively associated within no reason and immoral reason items.

Table 4

Study 2: Pearson's Zero-order Correlations

	1	2	3	4
1. Predictability	-	-.24**	-.19*	.42***
2. Harm	-.41***	-	.31***	-.62***
3. Intentionality	-.19*	.23**	-	-.47***
4. Moral Perceptions	.63***	-.64***	-.32***	-

Note. Pearson correlations ($N = 175$). Bottom diagonal = No Reason Item. Top diagonal = Immoral Reason Item. *** $p < .001$, ** $p < .01$, * $p < .05$.

7.2 Summary

Replicating the results of Study 1, actors assaulting a person for a well-understood immoral reason were once again judged as more predictable, more moral, and as having caused less harm than actors performing the same assault for seemingly no reason. Thus, inconsistent with the idea that this moral preference resulted from participants imagining actors described in our no reason vignette as carrying out a more vicious assault, we replicated the results of Study 1 while describing each physical assault—along with its consequences—unambiguously (i.e., as a punch that broke a person's jaw). Second, although we observed a negative association between judgments of predictability and intentionality within both no reason and immoral reason items, judgments of intentionality once again did not differ across item type. Hence, the moral preference for more predictable immoral actors observed in Studies 1 and 2 cannot be explained by participants perceiving the immoral actions of less predictable actors as having been performed with greater intent.

8. Study 3

In Study 3 we introduced a control item to assess peoples' baseline intuitions when information regarding an agent's motives was left unstated. We hypothesized that actors described as performing an immoral act for seemingly no reason would be judged as less predictable and less moral than those whose motives for performing an immoral act were immoral or left unstated, despite such actors performing either an additional (e.g., theft) or the identical (i.e., assault) immoral action.

9. Method

9.1 Participants

Four hundred and one US residents (8% Asian-American; 9% African-American; 80% European-American; 2% mixed ethnicity; 1% Native American) were recruited from Amazon Mechanical Turk using the same recruitment criteria as Studies 1 and 2. Participants received \$0.50 upon completion of a 5-minute online questionnaire. Those who participated in Studies 1 or 2 were restricted from participating in Study 3. We excluded data from 25 participants who reported responding randomly at some point during the experiment, leaving data from 376 participants (43% Female; $M_{age} = 39.51$, $SD_{age} = 12.35$) to be analyzed. This experiment was preregistered through Open Science Framework (<https://osf.io/avnug>).

9.2 Measures and Materials

We replaced moral reason items from Studies 1 and 2 with a control item that did not mention a reason (or lack of reason) for the described assault (e.g., "Kevin punches James [breaking James' jaw].") Apart from this change, Study 3 used the same measures and materials as Study 2.

9.3 Design and Procedure

Study 3 used the same procedure and within-subjects design as Study 2. However, to further improve data quality, Study 3 concluded with the following item: “Is there any reason that we shouldn't use your data (e.g., did you randomly select responses at any point during the survey)?” Participants who responded “Yes” to this question were removed from our analyses ($n = 25$).

9.4 Data Preparation

Once again, consistent with our pre-registered intention, we excluded participants who stated that they imagined a reason for the physical assault described in our no reason item ($n = 62$). Note that the interpretation of all significance tests remains the same when including these participants in our sample (see supplementary materials Part B).

10. Results and Discussion

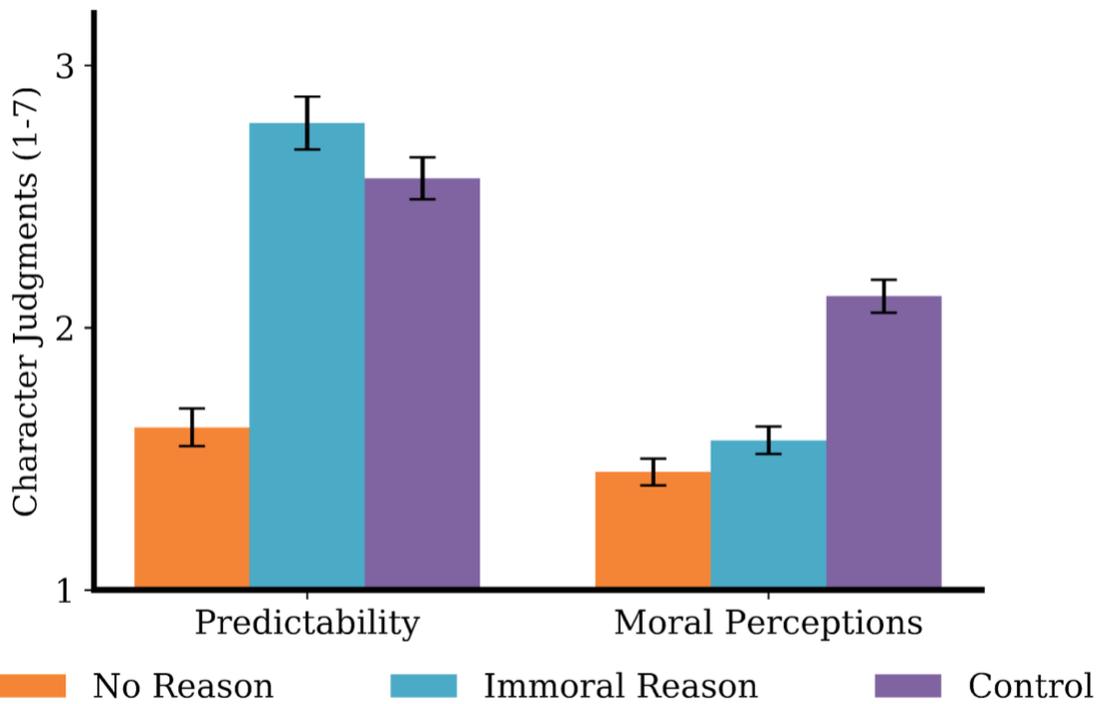
Participants' mean predictability and moral perception judgments for no reason, immoral reason, and control items can be viewed in Figure 3. We compared participants' judgments of predictability, moral character, intentionality, and harm across three items (no reason vs. immoral reason vs. control) using repeated-measures ANOVAs. Sensitivity power analyses revealed 80% power to detect a minimum effect size of $\eta_p^2 = .015$ for the conducted repeated-measures ANOVAs and $d = 0.16$ for follow-up paired samples t -tests. Predictability was found to vary across item type, $F(2, 626) = 91.41, p < .001, \eta_p^2 = .226$. As expected, actors engaging in violence for seemingly no reason were perceived as less predictable ($M = 1.62, 95\% CI = [1.48, 1.76]$) than those engaging in violence for an immoral ($M = 2.78, 95\% CI = [2.58, 2.98]$), $t(313) = 12.22, p < .001, d = 0.73$, or unstated reason ($M = 2.57, 95\% CI = [2.41, 2.73]$), $t(313) = 11.46, p < .001, d = 0.71$. Additionally, moral perceptions varied across item type, $F(2, 626) = 99.04, p$

$< .001$, $\eta_p^2 = .240$. Replicating the results of Studies 1 and 2, actors engaging in violence for an immoral reason were perceived as more moral ($M = 1.57$, 95% $CI = [1.47, 1.67]$) than those engaging in violence for seemingly no reason ($M = 1.45$, 95% $CI = [1.35, 1.55]$), $t(313) = 3.09$, $p = .002$, $d = 0.13$. Furthermore, actors assaulting a stranger for an unstated reason were judged as more moral ($M = 2.12$, 95% $CI = [2.00, 2.24]$) than those performing the same assault for seemingly no reason, $t(313) = 12.08$, $p < .001$, $d = 0.66$.

Judgments of intentionality also differed across item type, $F(2, 626) = 32.66$, $p < .001$, $\eta_p^2 = .094$. Agents performing an immoral act without an easily accessible motive were judged as having acted less intentionally ($M = 6.21$, 95% $CI = [6.04, 6.38]$) than those acting with a clear immoral motive ($M = 6.45$, 95% $CI = [6.32, 6.58]$), $t(313) = 2.64$, $p = .009$, $d = 0.17$, but more intentionally than those whose motive was left unstated ($M = 5.68$, 95% $CI = [5.51, 5.85]$), $t(313) = 4.97$, $p < .001$, $d = 0.34$. Lastly, perceptions of harm varied across item type, $F(2, 626) = 12.75$, $p < .001$, $\eta_p^2 = .039$. Despite their actions being identical, actors engaging in violence for an unstated reason were perceived as causing less harm ($M = 6.20$, 95% $CI = [6.07, 6.33]$) compared to those engaging in violence for seemingly no reason ($M = 6.47$, 95% $CI = [6.36, 6.58]$), $t(313) = 4.81$, $p < .001$, $d = 0.25$. Unlike in Studies 1 and 2, we observed no difference in perceived harm when comparing agents acting for an immoral reason ($M = 6.42$, 95% $CI = [6.31, 6.53]$) and no reason, $t(313) = 0.88$, $p = .382$, $d = 0.05$.

Figure 3

Study 3: Judgments of Predictability and Moral Perceptions



Note. Error bars represent +/- 1 standard error.

10.1 Associations Between Key Variables

In Study 3, we once again examined the zero-order associations between key variables of interest (see Tables 5 and 6). As in Studies 1 and 2, we observed a positive association between perceptions of predictability and morality within all item types.

Table 5

Study 3 Pearson's Zero-order Correlations: No Reason and Immoral Reason Items

	1	2	3	4
1. Predictability	-	-.16**	-.02	.36***
2. Harm	-.15**	-	.47***	-.44***
3. Intentionality	-.07	.32***	-	-.26***
4. Moral Perceptions	.61***	-.39***	-.20***	-

Note. Pearson correlations ($N = 314$). Bottom diagonal = No Reason Item. Top diagonal = Immoral Reason Item. *** $p < .001$, ** $p < .01$, * $p < .05$.

Table 6

Study 3 Pearson's Zero-order Correlations: Control Item

	1	2	3	4
1. Predictability	-			
2. Harm	-.20***	-		
3. Intentionality	-.25***	.37***	-	
4. Moral Perceptions	.64***	-.38***	-.37***	-

Note. Pearson correlations ($N = 314$). *** $p < .001$, ** $p < .01$, * $p < .05$.

10.2 Summary

The results of Study 3 replicated the primary findings of Studies 1 and 2. Immoral actors assaulting an individual for seemingly no reason were judged as less predictable and less moral than actors performing the same assault, along with an additional immoral act (e.g., theft), for a well-understood immoral reason. Furthermore, no-reason actors were judged as less predictable and less moral compared to those whose reason for assaulting a person was left unstated. Thus, describing an immoral actor as lacking an intelligible motive for their actions resulted in

especially harsh judgments of their moral character compared to when information about an actor's motive was left unstated. Lastly, the observed moral preference for more predictable immoral actors was once again not explained by judgments of intentionality. In fact, despite being judged as more moral, immoral-reason actors were judged as having acted more intentionally compared to actors described in our no reason vignette.

11. Study 4

In Studies 1-3, participants morally preferred actors who assaulted a person for a well-understood immoral reason compared to actors who performed an identical assault for seemingly no reason. One possible explanation for this moral preference is that in describing an individual as assaulting another person for no reason, participants inferred that this individual took pleasure in the pain and suffering that they caused.³ Such sadistic motives may reasonably be thought to be indicative of a violent person possessing an especially poor moral character. In Study 4, we introduced a sadistic item, in which an individual was described as assaulting a person in order to enjoy their pain and suffering. If such sadistic motives are readily inferred for the actor described in our no reason vignette, we would expect that participants' judgments of both actors would be difficult to distinguish. Conversely, if participants view the actor described in our no reason vignette as unpredictable, yet not sadistic, then participants' judgments of both actors should be easy to differentiate, specifically with regards to assessments of their predictability.

12. Method

12.1 Participants

A sample of 602 participants was recruited from the online crowdsourcing platform Prolific. Participants were recruited under the condition that they were residents of either the US

³ Although note that, in Studies 1-3, we excluded participants who stated that they imagined a reason (including sadistic pleasure) for the assault described in our no reason vignette.

or Canada and maintained a Prolific approval rating of 95% or greater. Participants received £0.60 (approximately \$0.75 USD) upon completion of a 6-minute online questionnaire. We excluded data from 78 participants who reported responding randomly at some point during the experiment or failed to pass a bot-detection item, leaving data from 524 participants (55% Female; $M_{\text{age}} = 33.63$, $SD_{\text{age}} = 13.50$) to be analyzed⁴. This experiment was preregistered through Open Science Framework (<https://osf.io/gjcwf>).

12.2 Materials

Study 4 introduced a sadistic vignette for which an individual was described as assaulting another person for a sadistic reason (e.g., “Harvey punches Ben (breaking Ben's jaw) because he enjoys watching people in pain and suffering.”). Apart from the addition of this vignette, Study 4 used the same materials as Study 3.

12.3 Measures

Following the presentation of each vignette, participants judged the person acting within the vignette on various moral dimensions presented in a randomized order. In Study 4, we introduced a new measure (i.e., trustworthiness) in order to assess the role that perceptions of predictability play in judgments of trust, a critical component of cooperation. We also introduced a measure of intelligibility, which we expected to share a strong association with assessments of predictability and a similar relation to judgments of morality. Finally, we no longer asked participants to judge how intentional they viewed the actions of each actor on account of these judgments failing to explain participants' moral preference for more predictable actors in Studies 1-3. In Study 4, all character judgments were elicited on a 100-point slider (as opposed to a 7-point scale) to allow participants to provide more exact judgments of each described actor.

⁴ Unlike Mechanical Turk, Prolific did not allow us to view and report the ethnic breakdown of our sample.

Furthermore, our measure of moral perception was also modified such that it was now assessed with a single item, as opposed to a multi-item composite (see below).

12.3.1 Moral Perception. Participants judged the moral character of each hypothetical actor using a 100-point slider with the endpoints of “0” and “100” labelled as “*Very Immoral*” and “*Very Moral*” respectively. They were prompted by the question “How would you judge [Name’s] moral character?” and were instructed to let “0” represent the most immoral/bad person they could imagine and “100” the most moral/good person they could imagine.

12.3.2 Trustworthiness. Participants judged the trustworthiness of each actor using a 100-point slider with endpoints labelled as “*Very Untrustworthy*” and “*Very Trustworthy.*” This judgment was prompted by the question “How would you judge [Name’s] trustworthiness?” In making this judgment participants were instructed to let “0” represent the most untrustworthy person they could imagine and “100” the most trustworthy person they could imagine.

12.3.3 Predictability. Participants judged the predictability of each actor using a 100-point slider with endpoints labelled as “*Very Unpredictable*” and “*Very Predictable.*” They were prompted by the question “How would you judge [Name’s] predictability?” and were instructed to let “0” represent the most unpredictable person they could imagine and “100” the most predictable person they could imagine.

12.3.4 Intelligibility. Participants were asked “How difficult is it to understand the motivation(s) for [Name’s] actions (e.g., understand why he acted the way he did)?” They responded to this question using a 100-point slider with endpoints labelled as “*Not at all Difficult*” and “*Very Difficult.*”

12.3.5 Harm. Participants were asked “How much harm would you say [Name] caused?” and responded to this question using a 100-point slider with the endpoints labelled as “*Caused No Harm*” and “*Caused a Great Deal of Harm.*”

12.4 Design and Procedure

Study 4 used the same within-subjects design as Study 3. The procedure of Study 4 also mirrored that of Study 3 with one exception. In Study 4, three additional items were presented following the randomized presentation of four vignettes (no reason, immoral reason, control, and sadistic) and the moral judgment of actors described in these vignettes. All three of these additional items were designed to directly assess participants’ beliefs regarding how a person’s predictability relates to their morality (e.g., “In general, do you think that people who are more predictable are more moral?”).

12.5 Data Preparation

As in Studies 1-3 and consistent with our pre-registered intention, we excluded participants who stated that they imagined a reason for the assault described in our no reason vignette ($n = 70$). Note that the interpretation of all significance tests remains the same when including these participants in our sample (see supplementary materials Part B).

13. Results and Discussion

First, we compared participants’ judgments of moral character, trustworthiness, predictability, intelligibility, and harm for no-reason and sadistic actors using paired-samples t -test. Sensitivity power analyses indicated 80% power to detect a minimum effect size of $d = 0.13$ for these analyses. Inconsistent with the idea that participants imagined sadistic motives for the assault described in our no reason vignette, actors described within this vignette were judged as more moral ($M = 9.77$, 95% $CI = [8.52, 11.02]$), $t(453) = 9.22$, $p < .001$, $d = 0.46$, and more

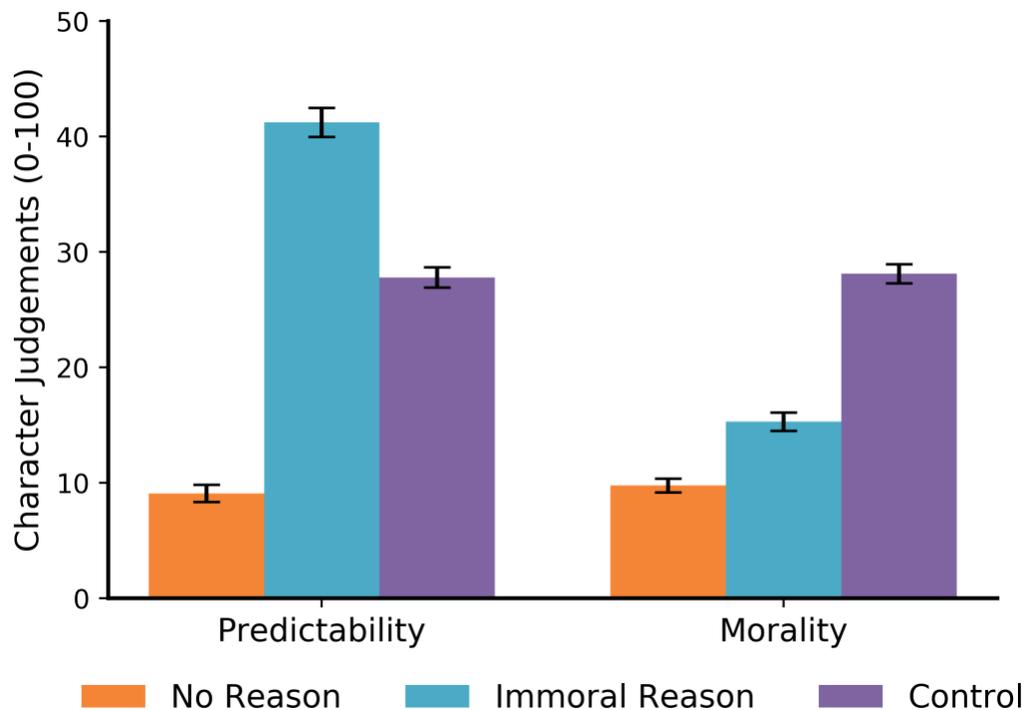
trustworthy ($M = 8.13$, 95% $CI = [6.85, 9.42]$), $t(453) = 3.71$, $p < .001$, $d = 0.19$, compared to actors described as having sadistic motives ($M = 4.14$, 95% $CI = [3.17, 5.11]$ and $M = 5.55$, 95% $CI = [4.36, 6.74]$, respectively). Similarly, actors described as assaulting a person for no reason were judged to have caused less harm ($M = 89.04$, 95% $CI = [87.83, 90.24]$), than those with sadistic motives ($M = 90.67$, 95% $CI = [89.50, 91.85]$), $t(453) = 3.49$, $p < .001$, $d = 0.13$. Additionally, participants judged sadistic actors as more predictable ($M = 40.64$, 95% $CI = [37.43, 43.85]$), $t(453) = 18.41$, $p < .001$, $d = 1.13$, and as possessing motives that were less difficult to interpret ($M = 58.34$, 95% $CI = [54.78, 61.90]$), $t(452) = 18.20$, $p < .001$, $d = 1.13$, compared to no-reason actors ($M = 9.09$, 95% $CI = [7.50, 10.69]$ and $M = 92.30$, 95% $CI = [90.87, 93.73]$, respectively). Therefore, it appears that the moral preference for more predictable immoral actors observed in Studies 1-3 was not simply a consequence of participants inferring sadistic motives for the unpredictable actor described in our no reason vignette.

Next, we compared participants' judgments of moral character, trustworthiness, predictability, intelligibility, and harm across our three critical items (no reason vs. immoral reason vs. control) using repeated-measures ANOVAs. Sensitivity power analyses indicated 80% power to detect a minimum effect size of $\eta_p^2 = .011$ for these analyses. Participants' mean predictability and moral perception judgments for no reason, immoral reason, and control items can be viewed in Figure 4. Predictability was found to vary across item type, $F(2, 904) = 283.32$, $p < .001$, $\eta_p^2 = .385$. Consistent with Studies 1-3, actors described as assaulting a person for no reason were perceived as less predictable ($M = 9.09$, 95% $CI = [7.50, 10.69]$) than those performing an assault for an immoral ($M = 41.23$, 95% $CI = [38.60, 43.87]$), $t(452) = 22.78$, $p < .001$, $d = 1.34$, or unstated reason ($M = 27.80$, 95% $CI = [25.96, 29.64]$), $t(453) = 17.03$, $p < .001$, $d = 1.00$. Intelligibility also varied across item type, $F(2, 906) = 1017.89$, $p < .001$, $\eta_p^2 =$

.692. Consistent with assessments of each actors' predictability, participants had greater difficulty assessing the motives of actors engaging in violence for seemingly no reason ($M = 92.30$, 95% $CI = [90.87, 93.73]$) compared to those engaging in violence for an immoral ($M = 30.76$, 95% $CI = [27.94, 33.57]$), $t(453) = 38.15$, $p < .001$, $d = 2.54$, or unstated reason ($M = 83.04$, 95% $CI = [81.17, 84.91]$), $t(453) = 8.66$, $p < .001$, $d = 0.51$.

Figure 4

Study 4: Judgments of Predictability and Moral Perceptions



Note. Error bars represent +/- 1 standard error.

Moral perceptions varied across item type, $F(2, 904) = 175.56$, $p < .001$, $\eta_p^2 = .280$.

Replicating the results of Studies 1-3, actors engaging in violence for an immoral reason were perceived as more moral ($M = 15.30$, 95% $CI = [13.60, 17.00]$) than those engaging in violence for seemingly no reason ($M = 9.77$, 95% $CI = [8.52, 11.02]$), $t(453) = 5.81$, $p < .001$, $d = 0.34$.

Furthermore, actors assaulting a stranger for an unstated reason were judged as more moral ($M =$

28.11, 95% $CI = [26.34, 29.87]$) compared to those performing the same assault for seemingly no reason, $t(452) = 19.87, p < .001, d = 1.10$. Judgments of trustworthiness also varied across item type, $F(2, 906) = 170.51, p < .001, \eta_p^2 = .273$. Consistent with moral perceptions, actors engaging in violence for an immoral reason were perceived as more trustworthy ($M = 14.24, 95\% CI = [12.45, 16.04]$) than those described as engaging in violence for no reason ($M = 8.13, 95\% CI = [6.85, 9.42]$), $t(453) = 6.33, p < .001, d = 0.36$. Additionally, actors assaulting a stranger for an unstated reason were judged as more trustworthy ($M = 27.20, 95\% CI = [25.37, 29.04]$) than those described as performing the same assault for no reason, $t(453) = 19.52, p < .001, d = 1.10$. Lastly, perceptions of harm varied across item type, $F(2, 906) = 21.17, p < .001, \eta_p^2 = .045$. As in Studies 1-2, actors engaging in violence for an immoral reason were perceived as causing *less* harm ($M = 86.45, 95\% CI = [84.90, 88.00]$) than those engaging in violence for no reason ($M = 89.04, 95\% CI = [87.83, 90.24]$), $t(453) = 4.01, p < .001, d = 0.17$, despite being described as performing an additional immoral act. Moreover, actors assaulting an individual for an unstated reason were perceived as causing less harm ($M = 85.05, 95\% CI = [83.67, 86.44]$) than those described as performing the same assault for no reason, $t(453) = 7.80, p < .001, d = 0.28$.

Study 4 also allowed us to assess participants' beliefs about the morality of predictable and unpredictable people (see Table 7). Despite participants demonstrating a moral preference for more predictable immoral actors, only a minority of participants endorsed the belief that predictable people are more moral than unpredictable people (with a majority selecting a neutral option). Nevertheless, a majority of participants (85%) did endorse predictable people as superior cooperation partners, consistent with their judgments of unpredictable "no reason" actors as

particularly untrustworthy and immoral. Thus, assessments of predictability may often play an explicit role when deciding whether to cooperate with another individual.

Table 7

Study 4 Belief Items

	Endorsed predictable people as favorable	Endorsed neither as favorable	Endorsed unpredictable people as favorable
Immoral Action	88 (19.4%)	322 (70.9%)	44 (9.7%)
Cooperation	384 (84.6%)	67 (14.8%)	3 (0.7%)
Morality	115 (25.3%)	330 (72.7%)	9 (2.0%)

Note. Each cell contains the number of participants selecting a particular response option with the percentage of participants selecting the response option presented in parentheses. *Immoral Action* = “Who do you believe is more moral? A predictable person performing immoral actions or an unpredictable person performing immoral actions?”; *Cooperation* = “Who do you believe makes a better cooperation partner? A predictable person or an unpredictable person?”; *Morality* = “In general, do you think that people who are more predictable are more moral?”

13.1 First Trial Judgments

In Studies 1-4, participants demonstrated a moral preference for actors assaulting a person for a well-understood immoral reason as opposed to seemingly no reason, despite immoral-reason actors being described as performing an additional immoral act. However, Studies 1-4 all utilized a within-subjects design, in which participants judged both no-reason and immoral-reason actors. Thus, whether this moral preference would be observed in a between-subjects design, absent the comparison between actors, remains unknown. In order to answer this question, we conducted exploratory analyses assessing only participants’ judgments of the first actor described. A full report of these exploratory between-subjects analyses for Studies 1-4 can be viewed in the supplementary materials (Part C). Additionally, here we report the results of

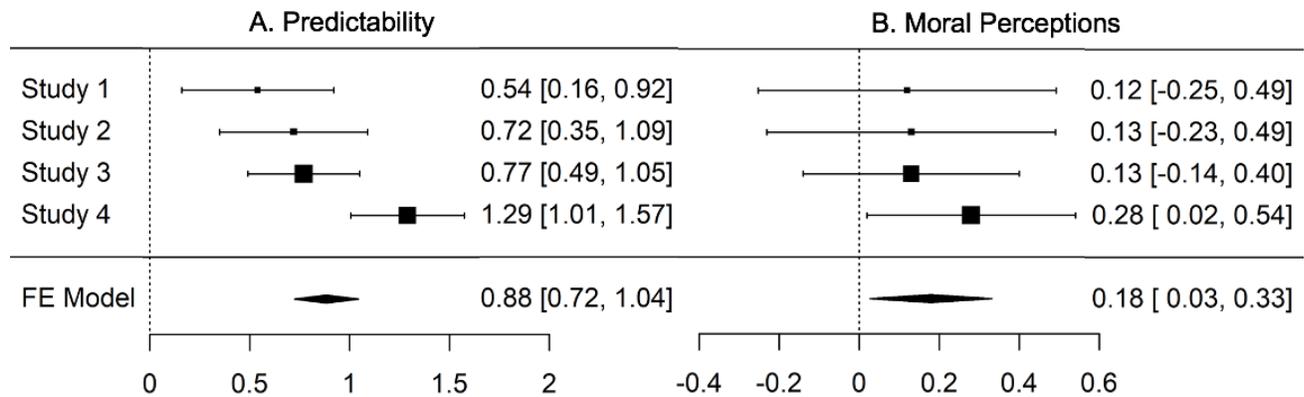
internal meta-analyses examining the effect sizes obtained when comparing participants' first trial judgments of immoral-reason and no-reason actors in Studies 1-4⁵.

In order to estimate the average effect size when comparing first trial judgments of immoral-reason and no-reason actors, we conducted internal meta-analyses using a fixed-effects method for judgments of predictability (Figure 5A) and moral perceptions (Figure 5B). These analyses revealed that, across Studies 1-4, participants judged immoral-reason actors as more predictable, $d = 0.88$, 95% *CI* [0.72; 1.04], $z = 10.86$, $p < .001$, and more moral, $d = 0.18$, 95% *CI* [0.03; 0.33], $z = 2.31$, $p = .021$, than no-reason actors. Furthermore, we also conducted an internal meta-analysis for judgments of harm, as these judgments were also elicited consistently across Studies 1-4. Consistent with judgments of morality, participants judged immoral-reason actors as causing less harm than no-reason actors, $d = 0.31$, 95% *CI* [0.16; 0.46], $z = 3.97$, $p < .001$. Overall, examining only participants' first trial judgments, we find evidence when aggregating across Studies 1-4 that immoral-reason actors were judged as more predictable, more moral, and as having caused less harm compared to no-reason actors. Nevertheless, it is worth noting that the observed moral preference for immoral-reason actors appeared to be somewhat reduced when assessed between-subjects ($d = 0.18$ vs. $d = 0.26$ within-subjects), and that first trial judgments of morality did not reliably differ within Studies 1-3 (see Figure 5B). Thus, the observed moral preference for more predictable immoral actors, behaving immorally for a well-understood reason, may be strengthened by (yet not contingent on) judgments of immoral-reason and no-reason actors being elicited within a within-subjects design.

⁵ Note that the presented immoral reason and no reason vignettes were virtually identical across Studies 1-4, facilitating these analyses.

Figure 5

Internal Meta-Analyses for First Trial Judgments of Predictability and Moral Perceptions



Note. Forest plots depicting the results of fixed effects method internal meta-analyses for first trial judgments of predictability (Panel A) and moral perceptions (Panel B). Values depicted on the right-hand side of each panel reflect effect sizes (Cohen’s *d*) and the 95% confidence intervals surrounding these effect sizes (presented in brackets and depicted as error bars). Positive effect sizes reflect participants’ judgments of immoral-reason actors as more predictable and more moral than no-reason actors.

13.2 Associations Between Key Variables

We once again examined the zero-order associations between key variables of interest (see Tables 8 and 9) in Study 4. Consistent with Studies 1-3, we observed a positive association between perceptions of predictability and morality within no reason, immoral reason, and control items. However, this association disappeared within our sadistic item, further distinguishing this item from our no reason vignette. Additionally, we observed a positive association between perceptions of predictability and trustworthiness within all item types.

Table 8

Study 4 Pearson's Zero-order Correlations: No Reason and Immoral Reason Items

	1	2	3	4	5
1. Predictability	-	-.26***	-.09	.18***	.15**
2. Intelligibility	-.25***	-	.12**	-.12*	-.21***
3. Harm	-.26***	.23***	-	-.43***	-.49***
4. Trustworthiness	.38***	-.20***	-.29***	-	.68***
5. Moral Perceptions	.35***	-.23***	-.40***	.52***	-

Note. Pearson correlations ($N = 454$). Bottom diagonal = No Reason Item. Top diagonal = Immoral Reason Item. *** $p < .001$, ** $p < .01$, * $p < .05$.

Table 9

Study 4 Pearson's Zero-order Correlations: Control and Sadistic Items

	1	2	3	4	5
1. Predictability	-	-.39***	-.02	.09*	.08
2. Intelligibility	-.15**	-	.11*	-.06	-.03
3. Harm	-.26***	.15**	-	-.27***	-.28***
4. Trustworthiness	.58***	-.17***	-.34***	-	.69***
5. Moral Perceptions	.58***	-.20***	-.35***	.73***	-

Note. Pearson correlations ($N = 454$). Bottom diagonal = Control Item. Top diagonal = Sadistic Item. *** $p < .001$, ** $p < .01$, * $p < .05$.

13.3 Summary

A goal of Study 4 was to assess whether the observed moral preference for more predictable immoral actors could be explained by participants inferring sadistic motives for the assault described in our no reason vignette. To this end, we found that judgments of no-reason actors were consistently distinguished from judgments of sadistic actors assaulting a person in

order to enjoy their pain and suffering. Thus, the results of Study 4 were inconsistent with the idea that participants imagine sadistic motives for the assault described in our no reason vignette. Additionally, the results of Study 4 replicated the primary findings of Studies 1-3 using slightly modified measures. Specifically, participants judged actors assaulting a person for seemingly no reason as less predictable and less moral than those performing the same assault for an immoral or unstated reason. Similarly, no-reason actors were perceived as possessing more difficult to interpret motives and as being less trustworthy than actors described within immoral reason and control vignettes.

14. Study 5

Participants in Studies 1-4 judged actors assaulting a person for seemingly no reason as less predictable and less moral than those performing an identical assault for either a well-understood immoral reason or an unstated reason. Nevertheless, participants in Studies 1-4 did not witness the actions of the individuals they judged, instead being described these actions as well as each actor's motives. In Study 5, we obtained real-world video footage of physical assaults that lacked an apparent motive and described these assaults as occurring for no reason, an unstated reason, or an intelligible immoral reason. Despite participants witnessing each assault and more predictable immoral-reason actors being described as performing an additional immoral act (i.e., theft), we hypothesized that actors seen assaulting a person for seemingly no reason would be judged as less predictable and less moral compared to those assaulting a person for a well-understood immoral reason. Similarly, we predicted that no-reason actors would be judged as less predictable and less moral than actors seen assaulting an individual for an unstated reason.

Furthermore, while participants displayed a moral preference for more predictable immoral actors in Studies 1-4, it is unclear whether this preference was guided more by affect or deliberative reasoning. Thus, in Study 5, we explored the extent to which participants felt that their moral judgments were guided by emotion or a careful and reasoned assessment of each actor's character. Finally, we introduced a new measure of punishment, asking participants to assume the role of a judge and sentence each immoral actor on the basis of their actions. Thus, along with judging each immoral actor on various dimensions, participants were asked to make a hypothetical real-world judgment regarding the amount of punishment (i.e., months in jail) they felt each actor deserved.

15. Method

15.1 Participants

Five hundred participants (10% Asian-American; 6% African-American; 81% European-American; 2% mixed ethnicity) were recruited from Amazon Mechanical Turk. Those who participated in Studies 1-3 or Study 6 were restricted from participating in Study 5. Participants received \$1.25 upon completion of a 10-minute online questionnaire for which they were required to be residents of the United States and possess a Mechanical Turk HIT approval rate greater than or equal to 99%. We excluded data from 17 participants who reported responding randomly at some point during the experiment, leaving data from 483 participants (53% Female; $M_{\text{age}} = 43.15$, $SD_{\text{age}} = 13.66$) to be analyzed. This experiment was preregistered through Open Science Framework (<https://osf.io/em4tw>).

15.2 Materials

Study 5 featured three brief videos depicting a man physically assaulting another person. These videos were obtained via an internet search and depicted real-world assaults caught on

camera. Each assault was such that it featured a male attacker and male victim, lacked a readily apparent motive, and did not contain visible injury (e.g., blood). Presented with each video was a brief description of the assault (see Table 10). These descriptions were similar to the no reason, immoral reason, and control vignettes of Studies 3 and 4, yet were adapted to fit the real-world footage used⁶. In order to avoid presenting assaults of different severity across item types, we allowed each video to be paired with each description across participants. This resulted in the need for each video to show an assault that could plausibly be described as occurring for no reason. As such, the theft described in our immoral reason description was not shown in the presented video footage. Furthermore, unlike Studies 1-4, no-reason actors were explicitly stated to not have performed the additional immoral act (i.e., theft) done by the immoral-reason actor.

⁶ Based on our inability to find real-world footage of assaults matching the immoral reason vignette presented in Studies 1-4 (e.g., “Richard punches Michael in order to escape a bank he had just robbed for \$50,000.”), we modified this vignette in Study 5 such that the described immoral-reason actor no longer assaulted another individual to escape punishment for a robbery he committed but rather to commit the robbery itself.

Table 10

Study 5: Video Descriptions

Motive	Description
No Reason	“Gerald was caught on camera physically assaulting James for no reason (i.e., Gerald did not benefit in any way from assaulting James, who was unknown to him). The assault was brief, ending a few seconds after the captured video footage. Notably, Gerald did not steal any of James' belongings (e.g., his phone or wallet). Fortunately, James suffered only minor injuries as a result of the attack.”
Immoral Reason	“Richard was caught on camera physically assaulting Michael in order to steal his phone and wallet. The assault was brief, ending a few seconds after the captured video footage. Immediately after the attack Richard stole Michael’s phone and wallet (not shown on video). Fortunately, Michael suffered only minor injuries as a result of the attack.”
Control	“Kevin was caught on camera physically assaulting Robert. The assault was brief ending a few seconds after the captured video footage. Fortunately, Robert suffered only minor injuries as a result of the attack.”

Note. Each description was presented with one of three videos displaying an assault that featured a male attacker and male victim, lacked a readily apparent motive, and did not contain visible injury. The pairing of videos and descriptions as well as the order of presentation of these pairs was counterbalanced across participants. Names of the attacker and victim were randomized on each trial.

15.3 Measures

The measures used in Study 5 mirrored those used in Study 4. Participants judged each hypothetical actor on five dimensions (moral perceptions, trustworthiness, predictability, intelligibility, and harm) using a 100-point slider. As in Study 4, these judgments were made in a randomized order. However, Study 5 also featured the addition of two new measures (punishment and emotion/reason), which are described below.

15.3.1 Punishment. Participants were asked to imagine they were a judge responsible for sentencing each actor. They were told that based on the crimes committed by this person, they can be sentenced to serve anywhere from 0 to 60 months (5 years) in prison. Participants were then asked to judge how much punishment each hypothetical actor deserved by responding to the

question: “How many months in prison would you sentence [Name] to serve for his actions in the described scenario?” Responses to this question were made on a 60-point slider with the endpoints of “0” and “60” labelled as “0 months” and “60 months” respectively.

15.3.2 Emotion/Reason. We assessed the extent to which participants reported relying on emotion and reason when judging each actor in Study 5. Specifically, participants were asked to indicate their agreement with the following statements using a 7-point scale that ranged from 1 (Strongly Disagree) to 7 (Strongly Agree): “When judging [Name] I relied heavily on the emotions that his actions brought about” and “When judging [Name] I relied heavily on a careful and reasoned assessment of his actions.” Next, participants were asked “Did you rely more on emotions or reasoning when judging [Name]?” Responses to this question were made using a 100-point slider with the endpoints of “0” and “100” labelled as “*Relied More on Emotions*” and “*Relied More on Reasoning*” respectively.

15.4 Design and Procedure

Study 5 featured the same within-subjects design as Studies 1-4. Participants judged three actors who they watched physically assault another individual. The motives of each actor were described in a brief vignette which accompanied each presented video and varied across item type (i.e., no reason, immoral reason, and control; see Table 10). The order in which items were presented, along with the specific pairing of videos and descriptions was counterbalanced across participants. Following the evaluation of each actor on various dimensions (i.e., moral perception, trustworthiness, predictability, intelligibility, harm, punishment, and emotion/reason), participants were reminded of the presented no reason vignette and were asked if they imagined a reason for this actor’s assault when making their judgments (as in Studies 1-4). Next, mirroring the procedure of Study 4, participants concluded Study 5 by responding to

three belief items (e.g., “In general, do you think that people who are more predictable are more moral?”) and a data check question.

15.5 Data Preparation

Consistent with our pre-registered intention (and Studies 1-4), we excluded participants who stated that they imagined a reason for the assault depicted in our no reason vignette ($n = 87$). Additionally, we removed two participants from our analyses on account that these participants provided incoherent responses (e.g., judged an actor as causing a great deal of harm and as very moral). Note that these exclusions did not change the interpretation of any significance tests, with the exception that judgments of trustworthiness no longer differed between no-reason and immoral-reason actors⁷ (see supplementary materials Part B).

16. Results and Discussion

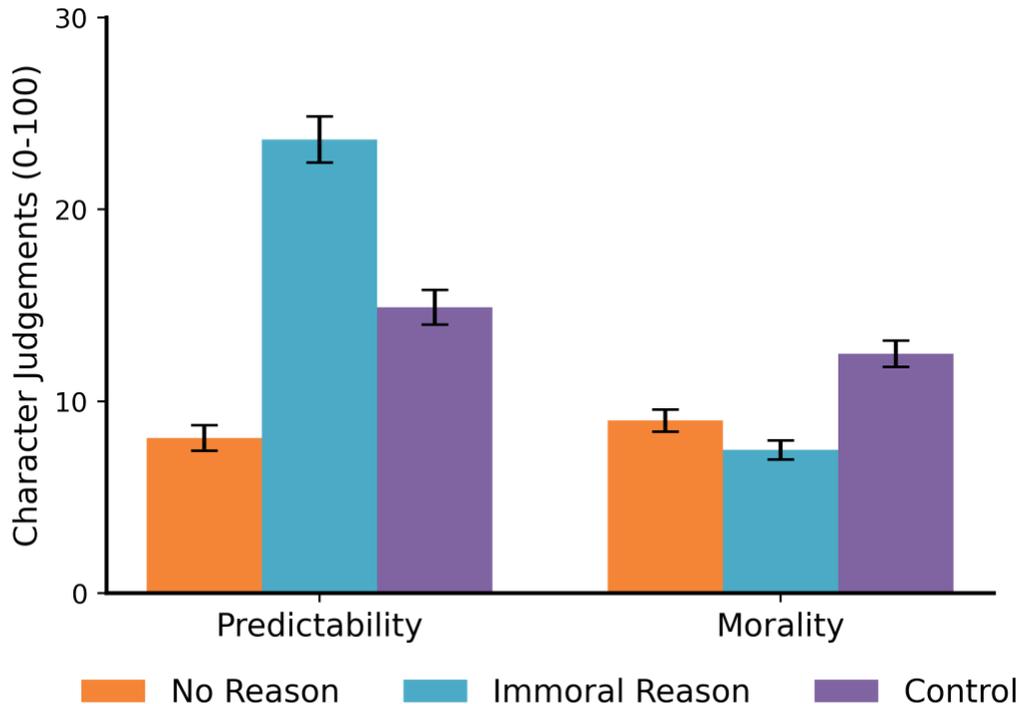
Participants’ mean predictability and moral perception judgments for no reason, immoral reason, and control items can be viewed in Figure 6. We once again compared participants’ judgments of morality, trustworthiness, predictability, intelligibility, harm, and punishment across item types (no reason vs. immoral reason vs. control) using repeated-measures ANOVAs. Sensitivity power analyses revealed 80% power to detect a minimum effect size of $\eta_p^2 = .012$ for the conducted repeated-measures ANOVAs and $d = 0.14$ for follow-up paired samples t -tests. Predictability was found to vary across item type, $F(2, 786) = 104.43, p < .001, \eta_p^2 = .210$. As expected, actors described as assaulting a person for no reason were judged as less predictable ($M = 8.09, 95\% CI = [6.78, 9.40]$) than those performing the assault for an immoral ($M = 23.65, 95\% CI = [21.28, 26.02]$), $t(393) = 13.40, p < .001, d = 0.77$, or unstated reason ($M = 14.91, 95\%$

⁷ When analyzing our full sample, participants’ judged actors assaulting an individual for an immoral reason as less trustworthy ($M = 5.84, 95\% CI = [4.98, 6.71]$) than those assaulting an individual for seemingly no reason ($M = 7.46, 95\% CI = [6.39, 8.52]$), $t(482) = 3.06, p = .002, d = 0.15$.

$CI = [13.12, 16.69]$), $t(393) = 7.88$, $p < .001$, $d = 0.42$. Intelligibility also varied across item type, $F(2, 786) = 410.44$, $p < .001$, $\eta_p^2 = .511$. Consistent with assessments of predictability, participants reported greater difficulty assessing the motives of actors assaulting an individual for seemingly no reason ($M = 89.12$, 95% $CI = [87.34, 90.90]$) compared to those assaulting an individual for an immoral ($M = 45.93$, 95% $CI = [42.40, 49.45]$), $t(393) = 23.22$, $p < .001$, $d = 1.50$, or unstated reason ($M = 82.46$, 95% $CI = [80.22, 84.71]$), $t(393) = 6.20$, $p < .001$, $d = 0.32$.

Figure 6

Study 5: Judgments of Predictability and Moral Perceptions



Note. Error bars represent +/- 1 standard error.

Moral perceptions varied across item type, $F(2, 786) = 37.85$, $p < .001$, $\eta_p^2 = .088$.

Inconsistent with the results of Studies 1-4, actors engaging in violence for an immoral reason were judged as less moral ($M = 7.47$, 95% $CI = [6.49, 8.46]$) than those engaging in violence for seemingly no reason ($M = 9.00$, 95% $CI = [7.87, 10.13]$), $t(393) = 2.83$, $p = .005$, $d = 0.14$.

Nevertheless, replicating the results of Studies 3 and 4, actors assaulting an individual for an unstated reason were judged as more moral ($M = 12.49$, 95% $CI = [11.14, 13.84]$) than those performing the same assault for seemingly no reason, $t(393) = 5.56$, $p < .001$, $d = 0.28$.

Judgments of trustworthiness also differed across item type, $F(2, 786) = 27.12$, $p < .001$, $\eta_p^2 = .065$. As predicted, actors assaulting an individual for an unstated reason were judged as more trustworthy ($M = 9.40$, 95% $CI = [8.15, 10.64]$) than those assaulting an individual for seemingly no reason ($M = 6.26$, 95% $CI = [5.28, 7.24]$), $t(393) = 5.68$, $p < .001$, $d = 0.27$. In contrast to judgments of morality, judgments of trustworthiness did not differ between immoral-reason ($M = 5.58$, 95% $CI = [4.65, 6.52]$) and no-reason actors, $t(393) = 1.36$, $p = .175$, $d = 0.07$.

Judgments of harm also varied across item type, $F(2, 786) = 52.15$, $p < .001$, $\eta_p^2 = .117$. Actors engaging in violence for an immoral reason were judged to have caused more harm ($M = 74.14$, 95% $CI = [72.08, 76.21]$) compared to those engaging in violence for seemingly no reason ($M = 66.63$, 95% $CI = [64.24, 69.02]$), $t(393) = 7.91$, $p < .001$, $d = 0.33$. Perceptions of harm did not differ when comparing judgments of actors described in control ($M = 65.52$, 95% $CI = [63.12, 67.92]$) and no reason vignettes, $t(393) = 1.23$, $p = .220$, $d = 0.05$. Finally, judgments of punishment varied across item type, $F(2, 786) = 56.57$, $p < .001$, $\eta_p^2 = .126$. Participants sentenced immoral-reason actors to more months in prison ($M = 27.95$, 95% $CI = [26.05, 29.86]$) than no-reason actors ($M = 23.48$, 95% $CI = [21.52, 25.44]$) when asked to make a hypothetical sentencing judgment, $t(393) = 6.48$, $p < .001$, $d = 0.23$. Additionally, participants sentenced actors assaulting an individual for seemingly no reason to more months in prison than actors performing the same assault for an unstated reason ($M = 20.88$, 95% $CI = [19.05, 22.70]$), $t(393) = 4.08$, $p < .001$, $d = 0.14$.

In Study 5, we explored the extent to which participants reported relying on emotion and reason when judging each presented immoral actor. The results of a repeated-measures ANOVA revealed that participants' self-reported use of emotions did not vary by item type, $F(2, 786) = 1.30, p = .274, \eta_p^2 = .003$. However, participants' use of reason did vary by item type, $F(2, 786) = 7.40, p = .001, \eta_p^2 = .018$. Specifically, participants endorsed using reason to a greater extent when judging actors engaging in violence for an immoral reason ($M = 5.92, 95\% CI = [5.81, 6.03]$) compared to when judging those engaging in violence for an unstated reason ($M = 5.77, 95\% CI = [5.66, 5.88]$), $t(393) = 3.49, p = .001, d = 0.14$, or no reason ($M = 5.77, 95\% CI = [5.66, 5.88]$), $t(393) = 3.07, p = .002, d = 0.13$. Furthermore, when asked to indicate whether they relied more on emotions or reason, participants reported relying more on reason when judging each immoral actor (No Reason: $M = 72.25, 95\% CI = [69.71, 74.79]$; Immoral Reason: $M = 75.44, 95\% CI = [73.09, 77.80]$; Control: $M = 72.87, 95\% CI = [70.36, 75.38]$)⁸. Additionally, one-sample t -tests revealed that participants' responses to this question reliably differed from the mid-point value (i.e., 50, representing equal reliance on emotion and reason) within all item types, No Reason: $t(393) = 17.23, p < .001, d = 0.87$, Immoral Reason: $t(393) = 21.23, p < .001, d = 1.07$, Control: $t(393) = 17.91, p < .001, d = 0.90$. Thus, participants reported using reason, more so than emotion, when judging the character of immoral actors presented in Study 5, particularly when judging actors who assaulted a person for a well-understood immoral reason.

As in Study 4, we assessed participants' beliefs regarding the morality of predictable and unpredictable individuals (see Table 11). Consistent with the results of Study 4, a majority of participants (88%) endorsed the belief that predictable people make better cooperation partners,

⁸ Participants responded to this question using a 100-point slider that ranged from 0 "Relied More on Emotions" to 100 "Relied More on Reasoning." Thus, higher values were indicative of a stronger self-reported reliance on reason when judging immoral actors (with values above 50 indicating a greater reliance on reason than emotion).

once again suggesting that assessments of predictability may often play an explicit role in peoples’ decisions regarding whether to cooperate with other individuals. Also mirroring the results of Study 4, only a minority of participants endorsed predictable individuals as more moral than their less predictable counterparts (with a majority selecting a neutral option). Thus, the observed moral preference for more predictable actors engaging in violence for a well-understood immoral reason in Studies 1-4, along with participants’ moral preference for more predictable actors engaging in violence for an unstated reason (as opposed to no reason) in Studies 3-5, was once again not reflected in the majority of participants’ explicit beliefs regarding predictable and unpredictable individuals.

Table 11

Study 5 Belief Items

	Endorsed predictable people as favorable	Endorsed neither as favorable	Endorsed unpredictable people as favorable
Immoral Action	90 (22.8%)	272 (69.0%)	32 (8.2%)
Cooperation	348 (88.3%)	44 (11.2%)	2 (0.5%)
Morality	140 (35.5%)	245 (62.2%)	9 (2.3%)

Note. Each cell contains the number of participants selecting a particular response option with the percentage of participants selecting the response option presented in parentheses. *Immoral Action* = “Who do you believe is more moral? A predictable person performing immoral actions or an unpredictable person performing immoral actions?”; *Cooperation* = “Who do you believe makes a better cooperation partner? A predictable person or an unpredictable person?”; *Morality* = “In general, do you think that people who are more predictable are more moral?”

16.1 First Trial Judgments

In Study 5, we conducted first trial between-subjects analyses utilizing only participants’ judgments of the first actor whose assault they witnessed and whose motive was described. Specifically, we compared participants’ first trial judgments of moral character, trustworthiness,

predictability, intelligibility, harm, and punishment across item type (no reason [$n = 127$] vs. immoral reason [$n = 135$] vs. control [$n = 132$]) using one-way ANOVAs and follow-up two-tailed independent samples t -tests⁹. Sensitivity power analyses revealed 80% power to detect a minimum effect size of $\eta_p^2 = .024$ for the conducted one-way ANOVAs and $d = 0.35$ for follow-up independent samples t -tests. Judgments of predictability varied across item type, $F(2, 391) = 15.68, p < .001, \eta_p^2 = .074$. As expected, actors engaging in violence for seemingly no reason were judged as less predictable ($M = 10.13, 95\% CI = [7.47, 12.80]$) than those engaging in violence for an immoral reason ($M = 23.33, 95\% CI = [19.46, 27.19]$), $t(260) = 5.49, p < .001, d = 0.68$, or an unstated reason ($M = 16.40, 95\% CI = [13.24, 19.56]$), $t(257) = 2.99, p = .003, d = 0.37$. Likewise, judgments of intelligibility varied across item type, $F(2, 391) = 78.79, p < .001, \eta_p^2 = .287$. Consistent with first trial judgments of predictability, no-reason actors were perceived as possessing more difficult to interpret motives ($M = 88.20, 95\% CI = [85.13, 91.26]$) compared to actors described in immoral reason ($M = 51.68, 95\% CI = [45.55, 57.81]$), $t(260) = 10.34, p < .001, d = 1.29$, and control vignettes ($M = 83.64, 95\% CI = [80.32, 86.95]$), $t(257) = 2.00, p = .047, d = 0.25$.

Judgments of moral character varied across item type, $F(2, 391) = 16.27, p < .001, \eta_p^2 = .077$. As predicted, actors assaulting an individual for seemingly no reason were judged to be less moral ($M = 10.31, 95\% CI = [8.22, 12.41]$) than those performing the same assault for an unstated reason ($M = 15.82, 95\% CI = [13.33, 18.31]$), $t(257) = 3.33, p < .001, d = 0.42$.

However, contrary to our hypothesis, no-reason actors were judged to be more moral than immoral-reason actors ($M = 7.46, 95\% CI = [5.82, 9.10]$), $t(260) = 2.14, p = .033, d = 0.26$.

Judgments of trustworthiness also varied across item type, $F(2, 391) = 6.42, p = .002, \eta_p^2 = .032$.

⁹ While we pre-registered the use of one-tailed t -tests, we use two-tailed t -tests here in order to report differences between judgments of immoral-reason and no-reason actors that ran contrary to our hypotheses.

Consistent with judgments of morality, actors seen assaulting a person for seemingly no reason were judged to be less trustworthy ($M = 7.43$, 95% $CI = [5.56, 9.29]$) than actors seen performing an identical assault for an unstated reason ($M = 10.47$, 95% $CI = [8.14, 12.80]$), $t(257) = 2.01$, $p = .046$, $d = 0.25$. Nevertheless, judgments of trustworthiness did not differ between immoral-reason ($M = 5.30$, 95% $CI = [3.42, 7.19]$) and no-reason actors, $t(260) = 1.58$, $p = .116$, $d = 0.19$.

Judgments of harm varied across item type, $F(2, 391) = 28.57$, $p < .001$, $\eta_p^2 = .127$. Actors seen engaging in violence for seemingly no reason were judged to have caused more harm ($M = 71.21$, 95% $CI = [67.59, 74.83]$) than actors engaging in violence for an unstated reason ($M = 57.42$, 95% $CI = [53.08, 61.75]$), $t(257) = 4.82$, $p < .001$, $d = 0.60$. However, no-reason actors were judged to have caused less harm than immoral-reason actors ($M = 76.90$, 95% $CI = [73.74, 80.06]$), $t(260) = 2.35$, $p = .019$, $d = 0.29$. Judgments of punishment also varied across item type, $F(2, 391) = 15.83$, $p < .001$, $\eta_p^2 = .075$. When asked to make a hypothetical sentencing judgment, participants sentenced immoral-reason actors to more months in prison ($M = 28.73$, 95% $CI = [25.64, 31.82]$) than no-reason actors ($M = 18.28$, 95% $CI = [15.06, 21.49]$), $t(260) = 4.64$, $p < .001$, $d = 0.57$. Conversely, judgments of punishment for actors described in no reason and control vignettes ($M = 17.67$, 95% $CI = [14.63, 20.72]$) did not differ, $t(257) = 0.27$, $p = .788$, $d = 0.03$.

16.2 Associations Between Key Variables

We once again examined the zero-order associations between key variables of interest (see Tables 12 and 13) in Study 5. As in Studies 1-4, we observed a positive association between perceptions of predictability and morality within no reason, immoral reason, and control items.

Additionally, consistent with Study 4, we observed a positive association between perceptions of predictability and trustworthiness within each item.

Table 12

Study 5 Pearson’s Zero-order Correlations: No Reason and Immoral Reason Items

	1	2	3	4	5	6
1. Predictability	-	-.43***	.22***	.21***	-.11*	-.12*
2. Intelligibility	-.35***	-	-.08	-.03	.08	.08
3. Moral Perceptions	.42***	-.22***	-	.62***	-.35***	-.26***
4. Trustworthiness	.48***	-.18***	.68***	-	-.22***	-.20***
5. Harm	-.08	.06	-.30***	-.18***	-	.30***
6. Punishment	-.06	.00	-.30***	-.20***	.37***	-

Note. Pearson correlations ($N = 394$). Bottom diagonal = No Reason Item. Top diagonal = Immoral Reason Item. *** $p < .001$, ** $p < .01$, * $p < .05$.

Table 13

Study 5 Pearson’s Zero-order Correlations: Control Item

	1	2	3	4	5	6
1. Predictability	-					
2. Intelligibility	-.38***	-				
3. Moral Perceptions	.37***	-.10	-			
4. Trustworthiness	.35***	-.08	.72***	-		
5. Harm	-.12*	.08	-.36***	-.28***	-	
6. Punishment	-.09	.04	-.34***	-.30***	.41***	-

Note. Pearson correlations ($N = 394$). *** $p < .001$, ** $p < .01$, * $p < .05$.

16.3 Summary

The results of Study 5 were mixed. While participants judged actors assaulting a person for seemingly no reason as less predictable and less moral than those performing the same assault for an unstated reason, no-reason actors were judged as less predictable and *more* moral than actors assaulting a person for a well-understood immoral reason. Thus, while a moral preference for more predictable immoral actors was observed when comparing actors performing the same single immoral act (i.e., an assault), unlike in Studies 1-4, participants did not show a moral preference for more predictable immoral-reason actors performing an additional immoral act (i.e., assault and theft). Of course, the introduction of video footage portraying the assault performed by each actor in Study 5 provides one explanation for these divergent findings. Nevertheless, additional methodological changes (e.g., vignette modifications) leave open the possibility of alternative explanations for why participants displayed a moral preference for immoral-reason actors in Studies 1-4 and no-reason actors in Study 5 (see Section 20.2 for a discussion of all methodological changes and their possible relation to the aforementioned divergent findings). Additionally, contrasting the results of Study 1, judgments of punishment were consistent with judgments of moral character in Study 5. That is, no-reason actors were judged to be deserving of more punishment (i.e., months in jail) than those assaulting a person for an unstated reason but less punishment than immoral-reason actors performing an additional immoral act. Lastly, in Study 5 we explored the extent to which participants' self-reported using emotion or reason when judging each of the presented immoral actors. While the majority of participants endorsed using both emotion and reason to some extent, participants reported relying primarily on reason when judging each actor.

17. Study 6

Study 6 represented a more general test of our predictability hypothesis. Instead of focusing on the stated motives for an action as the signal of unpredictability, we tested whether introducing superficial elements that signal unpredictability would make otherwise equivalent actions seem less moral. We hypothesized that those performing actions in an unusual way would be judged as less predictable and less moral than those performing equivalent acts in a more typical manner.

18. Method

18.1 Participants

One thousand and five US residents (8% Asian-American; 11% African-American; 79% European-American; 2% mixed ethnicity) were recruited from Amazon Mechanical Turk using the same recruitment criteria as Studies 1-3. In Study 6, we increased the size of our sample to accommodate switching to a between-subjects design. Participants received \$0.50 upon completion of a 5-minute online questionnaire. Those who participated in Studies 1, 2, or 3 were restricted from participating in Study 6. We excluded data from 142 participants who failed to pass both a comprehension check and data check question, leaving data from 863 participants (49% Female; $M_{\text{age}} = 40.29$, $SD_{\text{age}} = 12.97$) to be analyzed. This experiment was preregistered through Open Science Framework (<https://osf.io/zjen7>).

18.2 Materials

Participants were presented four vignettes, each of which depicted a hypothetical agent performing an action (e.g., assault) in either a common or unusual way (see Table 14). These vignettes were designed with the goal of equating common and unusual act items on all morally relevant dimensions (e.g., harm) other than predictability.

Table 14

Study 6: Vignettes

	Common Act	Unusual Act
Assault	Gerald hits a stranger with his fist (breaking his jaw).	Gerald hits a stranger with a frozen fish (breaking his jaw).
Robbery	Richard breaks into a stranger's home and steals \$100.	Richard breaks into a stranger's home and steals \$100 worth of underwear.
Sex Act	Kevin has sexual intercourse with a sex toy.	Kevin covers himself from head to toe in peanut butter and has sexual intercourse with a sex toy.
Noise Complaint	John wakes up his neighbors at 2 am by talking to his friends loudly in his backyard.	John wakes up his neighbors at 2 am by howling at the moon in his backyard.

Note. Items were presented in a randomized order.

18.3 Measures

The measures used in Study 6 (i.e., moral perceptions, predictability, intentionality, and harm) were identical to those used in Studies 2 and 3.

18.4 Design and Procedure

Study 6 featured a between-subjects design in which participants were randomly assigned to either a Common Act or Unusual Act condition. Those in the Common Act condition were presented with four common act vignettes (see Table 14) and judged each of the four described agents on various dimensions. Similarly, those in the Unusual Act condition were presented with four unusual act vignettes and judged four agents on the same set of dimensions. Following all judgments, participants responded to both a data check and comprehension check question.

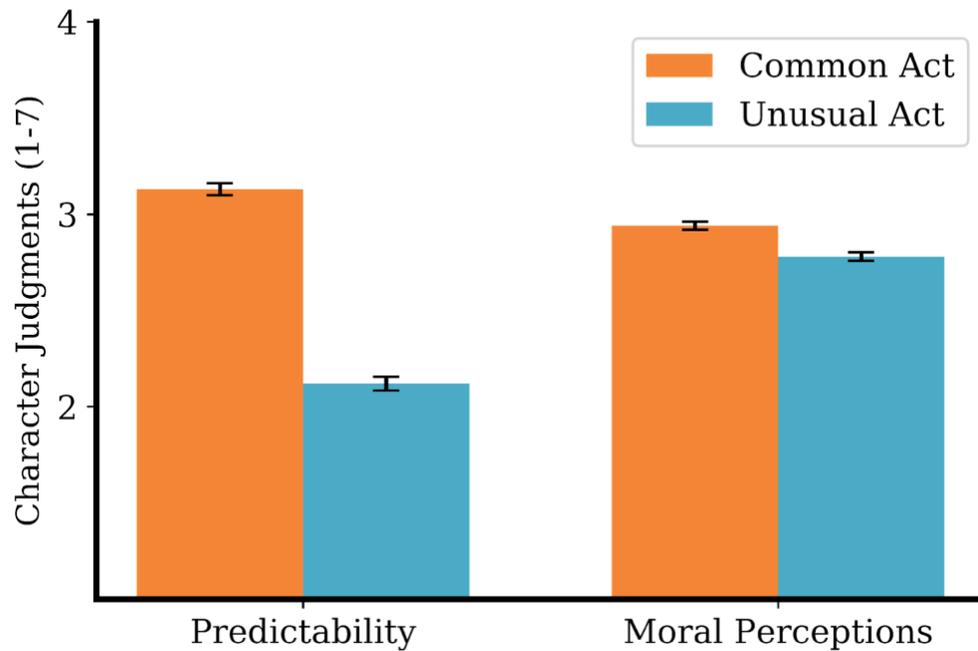
19. Results and Discussion

Participants' mean predictability and moral perception judgments for both Unusual Act and Common Act conditions can be viewed in Figure 7. We conducted independent-samples *t*-tests comparing judgments of predictability, moral perceptions, intentionality, and harm across

Common Act and Unusual Act conditions. A sensitivity power analysis indicated 80% power to detect a minimum effect size of $d = 0.19$ for these analyses. First, as expected, participants' perceived agents described in the Common Act condition as more predictable ($M = 3.13$, 95% $CI = [3.07, 3.19]$) compared to those in the Unusual Act condition ($M = 2.12$, 95% $CI = [2.05, 2.19]$), $t(861) = 15.07$, $p < .001$, $d = 1.02$. Furthermore, as predicted, participants judged agents performing actions in a common manner as more moral ($M = 2.94$, 95% $CI = [2.90, 2.98]$) than those performing the same actions in an unusual way ($M = 2.78$, 95% $CI = [2.74, 2.82]$), $t(861) = 3.56$, $p < .001$, $d = 0.26$. Additionally, agents described as performing actions in an unusual way were judged as acting more intentionally ($M = 6.26$, 95% $CI = [6.20, 6.32]$) compared to those performing the same actions in a more common manner ($M = 5.74$, 95% $CI = [5.68, 5.80]$), $t(861) = 8.33$, $p < .001$, $d = 0.57$. Lastly, unlike judgments of predictability, judgments of harm did not align with participants' moral perceptions, with agents performing actions in an unusual way being judged as causing *less* harm ($M = 4.34$, 95% $CI = [4.29, 4.39]$) compared to those performing the same actions in a more common manner ($M = 4.51$, 95% $CI = [4.46, 4.56]$), $t(861) = 3.49$, $p = .001$, $d = 0.23$.

Figure 7

Study 6: Judgments of Predictability and Moral Perceptions



Note. Error bars represent +/- 1 standard error.

19.1 Associations Between Key Variables

We once again examined the zero-order associations between key variables of interest (see Table 15) in Study 6. Consistent with Studies 1-5, we observed a positive association between predictability judgments and moral perceptions within both conditions.

Table 15

Study 6: Pearson's Zero-order Correlations

	1	2	3	4
1. Predictability	-	-.08	-.05	.45***
2. Harm	.02	-	.24***	-.29***
3. Intentionality	-.28***	.13**	-	-.17***
4. Moral Perceptions	.50***	-.29***	-.24***	-

Note. Pearson correlations ($N = 863$). Bottom diagonal = Unusual Act Condition. Top diagonal = Common Act Condition. *** $p < .001$, ** $p < .01$, * $p < .05$.

19.2 Summary

Consistent with the hypothesized moral preference for more predictable immoral actors, agents performing actions in an unusual way were judged as less predictable and less moral than those performing the same actions in a more common manner. This moral preference for more predictable agents described in the Common Act condition was observed despite these agents being judged as having caused *more* harm than those in the Unusual Act condition. Notably, agents described as performing an act in an unusual way were perceived as acting more intentionally than those performing the same actions in a common manner, perhaps offering one explanation for the observed moral preference for more predictable agents described in the Common Act condition¹⁰.

20. General Discussion

Across six studies we demonstrate that those signalling unpredictability with their actions, either by acting immorally without an intelligible motive (Studies 1-5) or by performing

¹⁰ Although note that judgments of intentionality did not explain the observed moral preference for more predictable immoral actors possessing more intelligible motives in Studies 1-3.

an immoral act in an unusual manner (Study 6), are viewed as possessing an especially poor moral character. Notably, in Studies 1-4, this moral preference for more predictable immoral actors was observed despite these actors being described as performing an additional immoral act (e.g., theft) compared to less predictable actors acting without a clear motive. Similarly, in Study 6, a moral preference for predictable actors was observed despite these actors being judged as causing *more* harm than less predictable actors performing immoral actions in an unusual way.

From traditional act-based perspectives (e.g., deontology and utilitarianism; Kant, 1785/1959; Mill, 1861/1998) this moral preference may appear puzzling, as participants judged actors causing *more* harm and violating *more* moral rules as *more* moral. Nevertheless, recent work suggests that people view actions not as the endpoint of moral evaluation, but as a source of information for assessing the moral character of those who perform them (Tannenbaum et al., 2011; Uhlmann et al., 2013). From this person-centered perspective (Pizarro & Tannenbaum, 2011; Uhlmann et al., 2015), a moral preference for more predictable immoral actors can be understood as participants judging the same immoral action (e.g., assault) as more indicative of negative character traits (e.g., a lack of empathy¹¹) when performed without an intelligible motive. That is, a person assaulting a stranger seemingly without reason or in an unusual manner (e.g., with a frozen fish) may be viewed as a more inherently unstable, violent, and immoral person compared to an individual performing an identical assault for a well-understood reason (e.g., to escape punishment for a crime in-progress). Such negative character assessments may lead unpredictable immoral actors to be considered a greater risk for causing future harms of uncertain severity to potentially random victims. Consistent with these claims, past work has

¹¹ For all studies in which participants made judgments of empathy (as part of a moral perceptions composite; Studies 1, 2, 3, and 6) actors described in our no reason item and unusual act condition were judged as being less predictable and as having less empathy compared to those described in our immoral reason items and common act condition.

shown that people judge those performing harmless-but-offensive acts (e.g., masturbating inside a dead chicken), as not only possessing more negative character traits compared to others performing more harmful acts (e.g., theft), but also as likely to engage in more harmful actions in the future (Chakroff et al., 2017; Uhlmann & Zhu, 2014).

Research on peoples' aversion to uncertainty (Chater & Loewenstein, 2016; FeldmanHall & Shenhav, 2019; Gneezy et al., 2006; Hogg, 2000) provides another plausible account of the observed moral preference for more predictable immoral actors. The future behavior of unpredictable individuals is likely to appear highly variable and rife with uncertainty, including a wide range of blameworthy and praiseworthy behaviors. In contrast, the future behavior of predictable individuals may be perceived as more or less certain, existing within a comparatively small range of common and easy-to-understand actions. Importantly, when thinking about the future behavior of those we deem predictable, highly immoral or severely harmful actions may not come to mind. Thus, the uncertainty and perceived variability of an unpredictable actor's possible future behavior may make such individuals appear frightening, avoidable, and immoral. This is likely especially true for unpredictable individuals known to have behaved immorally in the past.

20.1 Predictability and Cooperation

When evaluating the character of others, it is important to identify not only those likely to cause you harm, but also those who can be trusted as reliable cooperation partners. The results of Studies 4 and 5 suggest that people view predictability as relevant when assessing the cooperation potential of others. Specifically, we find that a majority of participants (85% in Study 4 and 88% in Study 5) view predictable people as preferable cooperation partners. Consistent with this finding, participants' judgments of predictability and trustworthiness

displayed a positive association within each item type in Studies 4 and 5. Moreover, participants judged actors they perceived as more predictable as more trustworthy, despite these actors engaging in the same (Studies 4 and 5) or an additional (Study 4) immoral act.

According to theories of morality-as-cooperation (Curry, 2016; Curry et al., 2019; Greene, 2013; Haidt, 2012; Rai & Fiske, 2011; Tomasello & Vaish, 2013), peoples' sense for morality developed for the purpose of facilitating cooperation. It is necessary to be able to predict how an agent will behave with some reliability if one is to confidently enter a stable relationship with them. Therefore, despite performing an additional immoral act (e.g., stealing \$50,000), immoral agents acting with intelligible motives may nevertheless represent more favorable cooperation partners compared to those whose immoral actions cannot be easily understood (even if 'cooperation' in this instance simply amounts to being able to minimize the amount of harm caused to oneself). A person who is motivated by a well-understood albeit immoral goal (e.g., greed) can be expected to betray others only when doing so serves their goal. The same may not be said of unpredictable agents who may betray people at any moment without cause. Within this framework, it makes sense that peoples' perceptions of predictability would play a prominent role in judging the cooperation potential and moral character of others.

20.2 Predictability Influences Moral Judgments: Exploring Boundary Conditions

In Study 4 we compared peoples' moral judgments of an unpredictable actor assaulting a stranger for seemingly no reason to a sadistic actor performing the same assault in order to enjoy his victim's pain and suffering. We found that individuals described as assaulting a stranger for sadistic satisfaction are viewed both as more predictable and less moral compared to those assaulting a stranger for no reason. While we demonstrate cases in which being perceived as predictable leads to more favorable judgments of moral character, individuals who are

predictable in the sense that they can be relied upon to harm others on account that it is enjoyable to them present a natural boundary condition. Nevertheless, even when judging those with deeply immoral desires, it is an open question whether people possess some moral preference for actors whose immoral behavior appears more predictable, perhaps on account that it is carried out in a predictable as opposed to unpredictable manner. Another potential boundary condition, and direction for future study, is the case of moral actions. Here it is plausible that individuals carrying out praiseworthy actions unpredictably and seemingly without motive may be morally preferred to those carrying out the same actions for a well-understood reason. Overall, while factors other than the perceived predictability of an individual guide judgments of moral character in a variety of situations, we demonstrate cases within the domain of immoral acts in which individuals perceived as more predictable are also judged as more moral, despite performing an additional harmful and immoral action.

In Study 5, participants were given the opportunity to view the immoral actions (i.e., physical assault) of the individuals whose character they judged. While participants continued to display a moral preference for more predictable actors performing one of the presented assaults for an unstated reason, as opposed to an unintelligible reason, the same moral preference was not observed for more predictable immoral-reason actors. Thus, inconsistent with the findings of Studies 1-4 (which exclusively utilized text-based descriptions of each actor), actors performing two immoral acts (i.e., assault and theft) were judged to be less moral than less predictable actors performing one of these acts (i.e., assault) for an unintelligible reason.

While it is possible that the divergent findings of Studies 1-4 and Study 5 can simply be explained by the introduction of video footage in Study 5, additional methodological changes leave open the possibility of other explanations. For instance, in order to accommodate the

presentation of video footage depicting real-world assaults, we altered descriptions of no-reason and immoral-reason actors in Study 5. As such, immoral-reason actors were no longer described as assaulting a person to escape punishment (e.g., escaping from a bank they had just robbed), but rather as assaulting a person in order to facilitate stealing their belongings. Additionally, unlike Studies 1-4, no-reason actors were explicitly described as not having performed the additional immoral act (i.e., theft) done by immoral-reason actors. Lastly, video footage of each assault lacked a readily apparent motive, on account that the immoral actor depicted in each video was described as assaulting an individual for a variety of reasons across participants. Thus, while an accompanying text-based description described immoral-reason actors as performing the witnessed assault in order to steal a person's belongings, the seemingly random and unpredictable nature of the assault attributed to immoral-reason actors may have resulted in these actors being viewed as less predictable, and as a result less moral, than the immoral-reason actors described in Studies 1-4¹². Overall, it is possible that any of the above methodological changes (including the introduction of video footage) resulted in the observed moral preference for immoral-reason, compared to no-reason, actors in Studies 1-4 but not Study 5. Future studies should more clearly examine the role that witnessing an immoral action plays in peoples' judgments of immoral actors possessing more or less intelligible motives.

20.3 The role of Emotions and Deliberative Moral Reasoning

In light of the observed moral preference for more predictable immoral actors, an interesting question surrounds the extent to which this preference is driven by individuals' spontaneous emotions or more deliberative moral reasoning. While participants reported relying

¹² Consistent with this view, immoral-reason actors were judged as less predictable and less moral in Study 5 (Predictability: $M = 23.65$, 95% $CI = [21.28, 26.02]$; Morality: $M = 7.47$, 95% $CI = [6.49, 8.46]$) compared to Study 4 (Predictability: $M = 41.23$, 95% $CI = [38.60, 43.87]$; Morality: $M = 15.30$, 95% $CI = [13.60, 17.00]$).

on emotions to some extent, they endorsed relying on reason to a greater extent when judging the immoral actors presented in Study 5. Furthermore, participants' reliance on reason was found to be particularly strong when judging actors assaulting a person for an immoral reason, perhaps on account of these actors being described as performing an additional immoral act. Nevertheless, despite these initial findings, the role that emotion and reason play in the observed moral preference for more predictable immoral actors remains an open and interesting question for future research. First, future work should aim to investigate the processes linking judgments of predictability and morality without relying exclusively on measures of self-report (as in Study 5). Second, it may be the case that participants' reliance on reason is restricted to cases in which one is evaluating actors signalling predictability with the intelligibility of their motives, as opposed to the typicality of their actions. Thus, future work should examine whether participants' greater reliance on reason, as opposed to emotion, persists when judging predictable and unpredictable actors performing immoral actions in common and unusual ways.

20.4 First Trial Comparisons

The present investigation used within-subjects designs to assess peoples' judgments of immoral actors described as possessing different motives for performing the same immoral act. Therefore, the observed moral preference for more predictable immoral actors assaulting a person for a well-understood immoral reason (as opposed to no reason) was demonstrated in a context which allowed for comparisons between actors to be made, potentially influencing participants' judgments. In order to examine whether this moral preference was observed absent any comparison between actors, we conducted exploratory between-subjects analyses assessing only participants' judgments of the first actor described. In Studies 1-3, while participants judged actors assaulting another person for a well-understood immoral reason as more predictable and

more moral than actors performing this assault for seemingly no reason, the observed moral preference was small (all d 's < 0.14) and not statistically reliable (all p 's > .331). However, in Study 4, immoral-reason actors were judged as significantly more predictable and more moral than no-reason actors. Additionally, the results of internal meta-analyses suggested that, across Studies 1-4, participants judged immoral-reason actors as more predictable, more moral, and as having caused less harm compared to no-reason actors when exclusively analyzing participants' first trial judgments. Therefore, participants' first trial judgments showed evidence of a moral preference for more predictable immoral actors across Studies 1-4.

We also examined between-subjects comparisons featuring only participants' first trial judgments in Study 5. Consistent with within-subjects analyses, actors assaulting a person for an unstated reason were judged as more predictable and more moral than actors performing the same assault for seemingly no reason¹³. Thus, within the video-based no reason and control items of Study 5, we find evidence of a moral preference for more predictable immoral actors in participants' first trial judgments. In summary, while the observed moral preference for more predictable immoral actors may have been strengthened by the use of within-subjects designs, the results of exploratory between-subjects comparisons suggest that this preference was also observed within participants' judgments of the first actor described.

20.5 Conclusion

Assessing the moral character of others is important for our social well-being. It is advantageous to be able to reliably identify who should be avoided and who can be trusted as a reliable cooperator. The future behavior of unpredictable actors is likely to be perceived as

¹³ Actors assaulting an individual for an unstated reason were also judged as more predictable and more moral than actors performing the same assault for seemingly no reason when exclusively analyzing first trial judgments in Studies 3 and 4 (see supplementary materials Part C).

uncertain. Understanding a person's motives allows us to confidently cooperate with them when our goals align and avoid conflict when our goals diverge. In contrast, the uncertainty surrounding unpredictable actors is unlikely to allow for such confidence, making them appear untrustworthy and dangerous. The current study demonstrates how individuals signalling unpredictability with their actions are perceived as possessing an especially poor moral character. We find that people judge unpredictable immoral actors as less moral and less trustworthy even when performing the same, or fewer, immoral acts. In revealing peoples' moral preference for predictable, as opposed to unpredictable, immoral actors, we propose that perceptions of predictability play an important, yet overlooked, role in judgments of moral character.

21. Open Practices

Data and materials from all studies can be accessed via the following link:

<https://osf.io/tc73f/>. Studies 2-6 were preregistered on the Open Science Framework. All preregistrations can be accessed via the following links: Study 2: <https://osf.io/bevtp/>; Study 3: <https://osf.io/avnug/>; Study 4: <https://osf.io/gjcwf/>; Study 5: <https://osf.io/em4tw/>; Study 6: <https://osf.io/zjen7/>.

References

- Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, *126*(4), 556-574.
- Ames, D. L., & Fiske, S. T. (2013). Intentional harms are worse, even when they're not. *Psychological Science*, *24*(9), 1755-1762.
- Chakroff, A., Russell, P. S., Piazza, J., & Young, L. (2017). From impure to harmful: Asymmetric expectations about immoral agents. *Journal of Experimental Social Psychology*, *69*, 201-209.
- Chater, N., & Loewenstein, G. (2016). The under-appreciated drive for sense-making. *Journal of Economic Behavior & Organization*, *126*, 137-154.
- Coen, J., & Coen, E. (Directors). (2007). *No country for old men* [Film]. Paramount Vantage.
- Curry, O. S. (2016). Morality as cooperation: A problem-centred approach. In *The evolution of morality*. T. K. Shackelford and R. D. Hansen, eds. pp. 27-51. New York: Springer.
- Curry, O. S., Mullins, D. A., & Whitehouse, H. (2019). Is it good to cooperate? Testing the theory of Morality-as-Cooperation in 60 societies. *Current Anthropology*, *60*(1), 47-69.
- Cushman, F. (2015). Deconstructing intent to reconstruct morality. *Current Opinion in Psychology*, *6*, 97-103.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191.
- FeldmanHall, O., & Shenhav, A. (2019). Resolving uncertainty in a social world. *Nature Human Behavior*, *3*(5), 426-435.

- Gneezy, U., List, J. A., & Wu, G. (2006). The uncertainty effect: When a risky prospect is valued less than its worst possible outcome. *The Quarterly Journal of Economics*, *121*(4), 1283-1309.
- Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, *23*(2), 101-124.
- Greene, J. D. (2013). *Moral tribes: Emotion, reason, and the gap between us and them*. Penguin.
- Haidt, J. (2012). *The Righteous Mind: Why good people are divided by religion and politics*. New York: Pantheon.
- Hogg, M. A. (2000). Subjective uncertainty reduction through self-categorization: A motivational theory of social identity processes. *European Review of Social Psychology*, *11*(1), 223-255.
- Kant, I. (1959). *Foundation of the metaphysics of morals* (L. W. Beck, Trans.). Indianapolis, IN: Bobbs-Merrill. (Original work published 1785).
- Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry*, *25*(2), 147-186.
- Mill, J. S. (1998). *Utilitarianism* (R. Crisp, Ed.) New York, NY: Oxford University Press. (Original work published 1861).
- Paal, T., & Bereczkei, T. (2007). Adult theory of mind, cooperation, Machiavellianism: The effect of mindreading on social relations. *Personality and Individual Differences*, *43*(3), 541-551.
- Pizarro, D. A., & Tannenbaum, D. (2011). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In M. Mikulincer & P. R. Shaver

- (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 91-108). New York, NY: APA Books.
- Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review*, *118*(1), 57-75.
- Royzman, E. B., & Baron, J. (2002). The preference for indirect harm. *Social Justice Research*, *15*(2), 165-184.
- Schein, C., & Gray, K. (2018). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review*, *22*(1), 32-70.
- Tannenbaum, D., Uhlmann, E. L., & Diermeier, D. (2011). Moral signals, public outrage, and immaterial harms. *Journal of Experimental Social Psychology*, *47*(6), 1249-1254.
- Tomasello, M., & Vaish, A. (2013). Origins of human cooperation and morality. *Annual Review of Psychology*, *64*, 231-255.
- Turpin, M. H., Walker, A. C., Fugelsang, J. A., Sorokowski, P., Grossmann, I., & Białek, M. (2021). The search for predictable moral partners: Predictability and moral (character) preferences. *Journal of Experimental Social Psychology*, *97*, 104196.
- Tversky, A., & Shafir, E. (1992). The disjunction effect in choice under uncertainty. *Psychological Science*, *3*(5), 305-310.
- Uhlmann, E. L., & Zhu, L. (2014). Acts, persons, and intuitions: Person-centered cues and gut reactions to harmless transgressions. *Social Psychological and Personality Science*, *5*(3), 279-285.
- Uhlmann, E. L., Zhu, L., & Tannenbaum, D. (2013). When it takes a bad person to do the right thing. *Cognition*, *126*, 326-334.

- Uhlmann, E. L., Zhu, L., & Diermeier, D. (2014). When actions speak volumes: The role of inferences about moral character in outrage over racial bigotry. *European Journal of Social Psychology, 44*, 23-29.
- Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science, 10*(1), 72-81.
- Van Dijk, E., & Zeelenberg, M. (2003). The discounting of ambiguous information in economic decision making. *Journal of Behavioral Decision Making, 16*(5), 341-352.
- Waldmann, M. R., & Dieterich, J. H. (2007). Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. *Psychological Science, 18*(3), 247-253.
- 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.