THE IMMORALITY BIAS: WHY “JOHN FLURBED MARY” SEEMS WRONG

Neil Randal Hester

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirement for the degree of Master of Arts in the department of Psychology and Neuroscience (Social Psychology).

Chapel Hill
2017

Approved by:

Kurt Gray

B. Keith Payne

Peter C. Gordon

ABSTRACT

Neil Randal Hester: The Immorality Bias: Why “John Flurbed Mary” Seems Wrong”
(Under the direction of Kurt Gray)

Seven experiments reveal the *immorality bias*: in morally ambiguous situations, people automatically jump to conclusions of wrongdoing. In Experiment 1, ambiguous acts (e.g., “A woman leaves work early to meet a man who is not her husband”) were rated as more immoral when people reported initial interpretations rather than most likely explanations. In Experiments 2-5, neutral nonsense actions (e.g., “John flurbed”) were judged as immoral to the extent that their context matched the dyadic moral template through the presence of a patient (“John flurbed Mary”; Experiments 2 and 3), intentionality (“John intentionally flurbed Mary”; Experiment 4), and suffering (“John intentionally flurbed Mary, who cried”; Experiment 5). The immorality bias is stronger under time pressure (Experiment 6), and process-dissociation reveals its automaticity (Experiment 7). The immorality bias suggests that intuitive moral judgment can be understood as a heuristic—one that hinges upon the dyadic moral template.

ACKNOWLEDGEMENTS

Thank you to my advisor, Kurt Gray, for his support and mentorship during this project and many others. Thanks also to Keith Payne for his wisdom and input on this project. Finally, thanks to Lindsey Helms and Hanne Schrickx for helping with stimulus creation for the experiments.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Imagine you came home early from work and saw the front door ajar with an unfamiliar car parked in driveway. Your first thought is likely either of a cheating spouse or a daylight burglary, despite many other innocuous possibilities, such as a repairperson fixing the cable or your spouse's nephew driving through town. In this ambiguous example, your mind seems to jump to nefarious explanations, even if they are relatively unlikely. We hypothesize that people generally assume wrongdoing in ambiguous situations, a phenomenon called the *immorality bias*. The presence of this fundamental asymmetry in moral intuitions would be practically important: while the law states that people are innocent until proven guilty, our minds may assume the opposite. Revealing an immorality bias would also suggest that moral judgment is similar to nonmoral decision making—with its heuristics and biases—and thus speak to debates about the nature of moral cognition.

Adaptive Biases

Biases are systematic errors in information processing that are especially prevalent in ambiguous situations. They are typically adaptive, nudging people towards safety, efficiency, reproductive success, or emotional wellbeing. Biases can be attentional, privileging the salience of information in consciousness, and include the bias to detect snakes in a visual field (Öhman, Flykt, & Esteves, 2001), to focus on sexually attractive women when mating goals are salient (Maner, Gailliot, & DeWall, 2007), and to attend to angry faces over happy faces (Hahn, Carlson, Singer, & Gronlund, 2006).

Given the importance of morality to individual and group well-being (de Waal, 2008; DeScioli & Kurzban, 2009), moral cognition might be expected to show an attentional bias. Consistent with this idea, morality cues are more quickly identified than competence cues (Ybarra, Chan, & Park, 2001), people are quicker to judge moral issues than nonmoral issues (Van Bavel, Packer, Haas, & Cunningham, 2012), and to react to language expressing moral disagreement than moral agreement (Van Berkum, Holleman, Nieuwland, Otten, & Murre, 2009). An attentional moral bias may even extend to visual perception such that moral stimuli are perceived more quickly than nonmoral stimuli (Gantman & Van Bavel, 2014).[1]

Biases can also be judgmental, in which evaluations or assessments consistently lean in the more adaptive direction. In ambiguous cases of failure, people blame the situation rather than themselves, because high self-esteem is generally beneficial (e.g., Bradley, 1978; Mezulis, Abramson, Hyde, & Hankin, 2004). In ambiguous social interactions, people assume that others hold the same attitudes, because similarity facilitates beneficial outcomes such as interpersonal liking and closeness (Mullen et al., 1985; Ross, Greene, & House, 1977). In ambiguous situations of potential loss, people err on the side of caution and avoid risk (Tversky & Kahneman, 1981, 1991)

Given that morality is itself adaptive—it increases evolutionary and cultural fitness—we might also expect that moral judgment should show a judgmental bias, such that people assume either others' innocence or guilt. On one hand, people might give others the benefit of the doubt, judging something as innocuous when it is actually immoral; on the other hand, they might instead judge actions as automatically immoral.

---

[1] This claim is controversial. For the debate, see Firestone and Scholl, 2015, 2016; Gantman and Van Bavel, 2016.

Innocence or Guilt?

At first blush, there are reasons why assuming innocence could be adaptive. Accusations, especially false ones, can destroy relationships and create enemies (Coy, Lambert, & Miller, 2016; Hall & Hall, 2001)—many families, congregations and political parties have been torn apart by harsh allegations. People appear to have an intuitive understanding of the risk involved in condemning others (Bottoms, Goodman, Schwartz-Kenney, & Thomas, 2002; Whitman & Davis, 2007) and show a reluctance to accuse others of cheating (Hyland, 2001; Nora & Zhang, 2010; Trevino & Victor, 1992) or lying (Bottoms et al., 2002; Reuben & Stephenson, 2013). However, although people may be reluctant to publically condemn others, the mind may still show a bias for immorality. Disjunctions between initial intuitions and eventual actions and judgments are frequent (e.g., Bodenhausen, 1990; Stroop, 1935; Toplak, West, & Stanovich, 2011; see Chaiken & Trope, 1999), and it may be adaptive to first assume immorality in ambiguous situations.

Most canonical and common cases of immorality involve physical or emotional harm (Graham et al., 2011; Gray, Young, & Waytz, 2012; Hofmann, Wisneski, Brandt, & Skitka, 2014; Malle, 2006), and an immorality bias might protect against such harm. As a potential victim of wrongdoing, the more quickly you recognize immorality, the better able you are to prevent yourself from being killed, injured or otherwise harmed (Blanchette, 2006; LoBue, 2010; LoBue & DeLoache, 2010; Öhman et al., 2001). As a potential bystander, the immorality bias may help you quickly aid victims, preserving the well-being of family members, friends, and other members of your coalition (Latané & Darley, 1968; Schroeder, Penner, Dovidio, & Piliavin, 1995). Furthermore, quickly recognizing wrongdoers is a necessary step to quickly distancing yourself from them, thereby guarding against guilt-by-association (Fortune & Newby-

Clark, 2008; Walther, 2002). Finally, although moral judgments can sometimes tear communities apart, these judgments are also often instrumental in the formation and maintenance of social groups (DeScioli & Kurzban, 2013; Lewis, Gray, & Meierhenrich, 2014; Rai & Fiske, 2011)—few things bring people together more than shared perceptions of villainy (Bosson, Johnson, Niederhoffer, & Swann, 2006; Sherif, 1961).

Biases Arise from Heuristic Cognition

Judgmental biases are often associated with heuristics, which are rules of thumb that simplify and expedite decision-making (Gilovich, Griffin, & Kahneman, 2002). These heuristics make complex situations more manageable and often yield accurate judgments (Gigerenzer & Brighton, 2009)—and they have a set of similar features. First, heuristics function efficiently, demanding little time and few cognitive resources (Chaiken, 1980; Gigerenzer & Goldstein, 1996). Second, they are intuitive, relying on natural assessments that often occur without deliberation (Tversky & Kahneman, 1983). Third, they involve substitution, relying on simple judgments as proxies for complex judgments (Kahneman & Frederick, 2002).

By efficiently and intuitively using judgmental proxies, heuristics lead to systematically incorrect judgments in certain situations—i.e., bias. Because judgmental biases are dependent on the operation of heuristics, an immorality bias requires that moral judgments proceed heuristically. Sunstein (2005) argues for just such “moral heuristics” that simplify complex decisions such as assignments of blame and punishment. Many of these heuristics apply to certain situations, such as taboo tradeoffs (Tetlock, Kristel, Beth, Green, & Lerner, 2000), corporate neglect (Viscusi, 2000), and judgments of medicine and pollution (Bergström & Lynöe, 2008; Scheske & Schnall, 2012). However, there is evidence that moral judgments are globally heuristic—featuring efficiency, intuition, and substitution.

Efficiency. Heuristics are partly adaptive due to their “fast and frugal” nature, which allows them to function with limited knowledge, time, and cognitive resources (Gigerenzer & Goldstein, 1996). Moral decisions are certainly efficient, as people pass moral judgments in under a second—and on stimuli as spare as a single word (Schein & Gray, 2015; Wright & Baril, 2011). In fact, moral judgments are likely more efficient than even nonmoral judgments, as moral judgments are processed more quickly than other judgments (Van Bavel et al., 2012).

Intuition. Heuristics are efficient largely because of their intuitive nature, generating judgments without conscious deliberation or reasoning (Gilovich et al., 2002). For example, the availability heuristic is unimpaired by cognitive load (Menon & Raghubir, 2003), the affect heuristic influences judgments without deliberation (Schwarz & Clore, 1983; Slovic, Finucane, Peters, & MacGregor, 2007) and gaze heuristics for catching moving objects are unconsciously employed (McLeod, Reed, & Dienes, 2003). Notably, many biases arising from heuristics are not appreciably reduced by incentivizing rational thought (Camerer & Hogarth, 1999). Moral judgments are also intuitive, relying more upon affect-based gut reactions than upon reasoned calculation (Haidt, 2001). As with nonmoral heuristics, these initial affective reactions are difficult to dispel with conscious deliberation (Gray, Schein, & Ward, 2014; Jacobson, 2012; Royzman, Kim, & Leeman, 2015).

Substitution. When someone is faced with a complex question, such as “How likely is it that this job candidate could be tenured in our department?” they may instead think “How impressive was the talk?” and answer this simple question instead (Kahneman & Frederick, 2002, p. 53). Heuristics are efficient and intuitive because they substitute judgments of easy-to-understand attributes in place of more complex attributes. For example, the availability heuristic substitutes ease of recall (“How readily does this come to mind?”) for base rates (“How often does this happen, taking into account many factors?”).

People’s moral judgments also seem to rely on substitutions. Negative affect inductions often lead to harsher moral judgments (e.g., Helzer & Pizarro, 2011; Wheatley & Haidt, 2005), as do high arousal inductions (Cheng, Ottati, & Price, 2013), suggesting that people often refer to the question “How do I feel?” when judging immoral situations, rather than carefully weighing aspects of the situation. However, moral judgments are based upon more than just negative affect—otherwise everything that left us feeling bad would seem immoral (Schein, Ritter, & Gray, in press). What are the elements within acts that moral cognition uses as substitutes for moral wrongness? The Theory of Dyadic Morality (Gray et al., 2012) suggests that three key elements for heuristic moral judgments: a dyad (person x acts upon person y), intention (x acts intentionally) and suffering (y suffers).

Dyadic Morality and the Immorality Bias

The Theory of Dyadic Morality suggests that moral judgment is fundamentally rooted in a cognitive template of two perceived minds—an intentional agent causing suffering to a vulnerable patient (i.e., a perpetrator and victim)—and that moral judgment proceeds by comparing acts to this dyadic template (i.e., dyadic comparison), with closer matches resulting in stronger moral judgment (Gray et al., 2012; Schein, Goranson, & Gray, 2015). Thus, acts which feature an obvious intentional agent and suffering patient should be most robustly judged as immoral, consistent with the greater perceived wrongness of murder and rape versus pornography and masturbation (Schein & Gray, 2015). The process of dyadic comparison, in which greater perceived harm translates to greater judged immorality, has all the characteristics of a heuristic: it is efficient, intuitive, and features a small number of simple elements—a dyad, intention, and suffering (Schein & Gray, 2015).

Such a dyadic heuristic is adaptive because it affords quick judgments of the most canonical and dangerous cases of immorality (Wilson, 1997), such as murder and rape; however, it also allows for bias—the immorality bias. Given the dyadic elements, dyadic morality predicts that the immorality bias should emerge most robustly when acts seem to feature an intentional dyad with suffering. Revealing a systematic link between the "dyadicness" of situations and assumptions of immorality would not only reveal the cognitive mechanism of the immorality bias (i.e., dyadic comparison), but would also address an important question in moral psychology.

There is a debate in the field about whether moral judgments involve a domain-general cognitive template (espoused by dyadic morality) or special domain-specific modules (espoused by moral foundations; Graham et al., 2013). Revealing the immorality bias would lend support to the dyadic morality position for two reasons. First, if the immorality bias is amplified by each additional element of dyad, intention, and suffering, it would argue against "basic" and indivisible moral modules which—by definition—cannot be deconstructed into more fundamental components (Ekman, 1992; Haidt, 2012). Second, if the immorality bias hinges on the presence of dyadic elements, it would reveal that the dyad is *causally* involved in moral cognition (i.e., it determines moral judgment), rather than being a common, but non-essential component of moral situations.

Substantial past research is consistent with both of these dyadic claims (Cameron, Lindquist, & Gray, 2015; Gray & Schein, in press), but these studies have often used specific acts (e.g., murder, incest), which some have argued could involve specialized moral judgment

(Graham, 2015; but see Gray & Keeney, 2015a, 2015b). For example, murder seems very wrong, but is it wrong because it matches the dyad (i.e., intentionally caused harm), or because it's just "murder?" The current research will address this concern by—among other methods—manipulating the dyad with nonsense actions (e.g., "pelled"). If it intuitively seems wrong when "John intentionally pelled Mary, who cried," then such assumptions of immorality can only be explained by a dyadic context because "pelled" is not a real immoral act.

The Present Research

Seven experiments tested for the *immorality bias*: the tendency to automatically assume wrongdoing in morally ambiguous situations. In Experiment 1, participants read ambiguous vignettes (e.g., "While a high school student takes a shower, he thinks of his younger sister") and we predicted that their first thoughts would show more immorality than what they believed to be the most likely explanation. In Experiment 2, people read about nonsense actions taken from psycholinguistics research that were either dyadic (e.g., "John pelled Mary") or nondyadic ("John pelled"). We predicted that the immorality bias would be much stronger in a full dyadic context (with agent acting upon patient), rather than a partial dyadic content (with only agent acting). In Experiment 3, people provided ratings of either immorality or virtue in response to dyadic and nondyadic sentences. This experiment tested the alternate explanation of a general "morality bias" in which people assume both immorality and virtue. In Experiments 4 and 5, we manipulated the presence of intention and suffering with these nonsense actions, predicting that both would amplify the immorality bias.

In Experiment 6, participants judged nonsense actions under a shorter or longer time limit. As the immorality bias is hypothesized to be intuitive, we predicted that it would be stronger under time pressure. Finally, in Experiment 7, participants judged sentences under variable time pressure and process dissociation was used to reveal the extent of automatic versus controlled processes in the immorality bias. We predicted that participants would show a stable automatic assumption of immorality.

Revealing the immorality bias would be practically important: if people first assume guilt before later considering innocence, it would have implications for any cases of rapid moral decision-making, such as judgments of other drivers (i.e., leading to road rage), judgments of children (i.e., leading to child abuse), and police judgments of suspects (i.e., leading to decisions to shoot; Correll, Park, Judd, & Wittenbrink, 2002; Payne, 2001). Revealing the cognitive determinants of the immorality bias would also provide a key commentary on dyadic morality and suggest that moral judgment can be understood similarly to nonmoral judgments, with its heuristics and biases. This would imply that moral psychology need not reinvent the wheel when investigating the processes of moral judgment, but rather start from the voluminous literature on nonmoral decisions making.

Of course, there are some notable differences between moral judgment and nonmoral judgment: for example, moral judgments are more affective in nature (Haidt, 2001) and more motivating of action (Skitka & Bauman, 2008) compared to nonmoral judgments. Perhaps most importantly, there is no obvious objective standard for measuring the immorality bias; judgments of immorality lack the objective base rates or tradeoffs that are often present for heuristic judgments (and facilitate strong claims about accuracy). However, such a bias can be revealed by contrasting intuitive and more considered decisions (Experiments 1, 6, 7), and by examining nonsense actions which lack any intrinsic immorality (Experiments 2-6). An immorality bias would be revealed if rapid, more intuitive judgments displayed more immorality than slower, more reasoned judgments.

EXPERIMENT 1: FIRST THING VERSUS MOST LIKELY

In this first experiment, participants read five morally ambiguous scenarios (e.g., A woman leaves work early to meet a man who is not her husband) and wrote down either the “first thing that comes to mind” or the “most likely explanation.” The immorality bias predicts that first thoughts should be more immoral (for example, cheating on a partner) than later thoughts (meeting a brother or friend).

Method

Participants and Design

Participants for all experiments were recruited using Amazon Mechanical Turk (MTurk) and paid between $.20 and $.75. Previous research has established MTurk as a viable marketplace for recruiting diverse and high-quality participants (Buhrmester, Kwang, & Gosling, 2011; Goodman, Cryder, & Cheema, 2012). For this first experiment, we recruited one hundred participants via MTurk who completed a two-condition (Instructions: First Thing, Most Likely) between-subjects experiment. Thirteen participants failed the attention check, leaving 87 participants (52.9% female, M_{age} = 34 years).

Procedure and Materials

Scenarios. Participants read five scenarios (presented in random order) and then described either “the first thing that comes to mind” (First Thing condition) or “the most likely explanation” (Most Likely condition) for each scenario. See Table 1 for a list of the scenarios. We also recorded the amount of time taken to complete each scenario.

Ratings. After providing their responses, participants rated the extent to which these responses were related to "immoral thoughts or behaviors" using a 5-point scale from *Not at all* (1) to *Extremely* (5). No effect of specific scenarios emerged, ($\alpha = .83$), and so participants' ratings were collapsed into a single rating.

Manipulation Checks. As participants' first thoughts should reflect rapid intuitions, we predicted that thoughts in the First Thing condition should be listed more quickly, and perhaps also involve less perceived difficulty. For this reason, we assessed the speed of responses and had participants rate "how difficult" they found the task on a scale from *Not at all* (1) to *Extremely* (5), as well as how often they had to "subdue or ignore other thoughts about the passage to follow instructions" on a scale from *Never* (1) to *Very often* (5).

After excluding one extreme outlier (Cook's D = .79), a between-subjects *t*-test (Instructions: First Thing, Most Likely) confirmed that participants in the First Thing condition used fewer seconds to complete each scenario ($M = 16.38$, $SD = 8.58$) than participants in the Most Likely condition ($M = 25.90$, $SD = 11.71$), $t(84) = 4.34$, $p < .001$, $d = .94$.[2] Contrary to our predictions, the same *t*-test (Instructions: First Thing, Most Likely) revealed no significant differences between conditions for either the difficulty or the intrusion item, *p*s > .2. This result suggests that participants may not subjectively view the Most Likely task as more difficult, even though it takes longer to complete.

Results and Discussion

Consistent with the immorality bias, a between-subjects *t*-test (Instructions: First Thing, Most Likely) revealed that people perceived greater wrongdoing in the First Thing condition (*M*

---

[2] Due to a programming error, the time for one of the scenarios in the First Thing condition was not recorded in Experiment 2. The Cronbach's alpha for the four remaining times is .85, which suggests that the remaining values sufficiently capture participants' average response times.

$= 3.78$, $SD = 1.13$) than in the Most Likely condition ($M = 2.58$, $SD = 1.29$), $t(85) = -4.63$, $p < .001$, $d = 1.00$.

One limitation of this experiment is that these longer scenarios are evocative and may have led people "down the garden path." Specific language can subtly influence our perceptions of meaning, and people have an intuitive sense of how euphemisms and roundabout descriptions often describe unsavory events (Bohner, 2001; Frazer & Miller, 2008; Henley, Miller, & Beazley, 1995).

Furthermore, detailed scenarios do not provide the flexibility necessary to cleanly manipulate contextual factors, such as the presence of a patient, the intentionality of an act, or the suffering of a patient. For these reasons, we used a nonsense action paradigm for Experiments 2 through 6.

EXPERIMENT 2: THE IMMORALITY OF NONSENSE ACTIONS

The immorality bias predicts that people will see immorality even in minimal situations, given the right context. To create these minimal situations, we provided participants with nonsense actions, such as “John pelled Mary” and “Jennifer gished Lisa” and they categorized these sentences as either immoral or not immoral. Of course, some actions might seem more “intrinsically” immoral by sounding similar to actual moral words (e.g., “frangled” sound like “strangled”), and we could be (unconsciously) predisposed to choose such words in our experiments (even “flurbed” might seem intrinsically bad). For this reason, we used only nonsense actions from past research in linguistics and cognitive psychology—actions which were used to investigate hypotheses completely unrelated to morality.

As biases are triggered by the presence of heuristic elements, dyadic morality predicts that the immorality bias should occur largely in the dyadic context of agent and patient. Accordingly, this experiment manipulated whether nonsense actions targeted a patient (“John pelled Mary”) or not (“John pelled”). Although the very presence of an agent might trigger some immoral judgments, we predicted that the immorality bias would be much larger in the dyadic context.

Method

Participants and Design

We recruited 54 participants via MTurk (46.3% female, M_{age} = 37 years), who completed a two-condition (Patient: Absent, Present) within-subjects experiment. No participants’ data were excluded from the study.

Procedure and Materials

Nonsense actions. In this experiment, we used nonsense actions as ambiguous stimuli, an often-used approach in cognitive psychology and linguistics. To create our stimulus set, we compiled nonsense actions from 15 cognitive psychology and linguistics articles. We then excluded verbs with irregular conjugations (e.g., strink and strunk) and verbs longer than two syllables (e.g., dorfinize) to create a more uniform set of actions. This selection process left us with a word bank containing 76 nonsense actions. See Table 2 for all nonsense actions and citations.

Agent and patient names. In addition to nonsense actions, the sentences also involved specifying agents and patients. Because specific names can influence judgments (Erwin, 2006; Silver & McCann, 2014), we created two name banks to ensure that the immorality bias is not driven by certain names. The agent and patient word banks each contained 40 names—20 male and 20 female—drawn from a list of the 40 most popular male and female names in the United States in the last 100 years (Social Security Administration, 2016). In all experiments, male and female names were randomly chosen to ensure that effects were not driven by particular agent or patient genders. See Appendix A for a full list of male and female names.

Sentence presentation. Using Inquisit Lab (version 4.0.9.0), we designed a program that dynamically creates sentences for each participant by combining a random agent, a random nonsense action, and a random patient (e.g., Jose “stiped” Louis; Helen “blicked” Kenneth”). This approach ensures that our effect is not driven by the inclusion of specific stimuli (Wells & Windschitl, 1999).[3]

---

[3] Across the four studies that use this nonsense action paradigm, this dynamic approach to creating stimuli yielded the possibility of providing participants with over 17 million unique sentences.

Sentence categorization task. Each participant viewed 76 unique sentences— one for each nonsense action—and categorized each as either Immoral or Not Immoral. The use of a binary outcome variable is common when participants make judgments of ambiguous or quickly-presented stimuli (e.g., Correll et al., 2002; Greenwald, Mcghee, & Schwartz, 1998; Payne, Cheng, Govorun, & Stewart, 2005). Within-subjects, we manipulated the presence of a Patient (Absent/Present): half of the sentences did not feature a patient (e.g., "John pelled"), whereas the other half did ("John pelled Mary"). The first eight sentences were presented as practice trials and were not included for analysis. We asked participants to provide each of their responses within five seconds; this amount of time proved ample, with participants successfully categorizing sentences in 98.9% of trials with an average latency of 1.12 seconds.

Results and Discussion

To account for variance owing to specific effects of participants, actions, or names, we analyzed the data using a fully cross-classified linear model with a binary outcome variable (Baayen, Davidson, & Bates, 2008). This model provided a more accurate and more powerful test of our manipulations than a traditional repeated measures ANOVA and also allows for some missing data (i.e., missing trials; Krueger, 2004). In all of the models, the intercepts varied randomly across both participant and verb levels so that our effects generalize beyond the current sample.

As predicted, the analysis of fixed effects revealed a main effect of Patient, $F(1, 3630) = 321.40$, $p < .001$, such that participants were more likely to rate sentences as immoral when a patient was Present ($M_{pct} = 54.3$, 95% CI [52.0, 56.6]) than when a patient was Absent ($M_{pct} = 24.7$, 95% CI [22.7, 26.7]). These results suggest that the presence of a patient influences the extent to which people show the immorality bias, and also suggests that the immorality bias cannot simply be explained by the negativity bias—the greater power of negativity versus positive stimuli (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001)—because the presence of a patient is not intrinsically negative.

However, another alternate explanation is that people gravitate toward morally relevant assumptions, both immoral and virtuous, in a dyadic context. In other words, the “immorality bias” could just be a “morality bias” in which people assume both helpful and harmful actions in a dyadic context. To test this alternate explanation, we replicated Experiment 2 and had a second group of participants categorize the sentence by virtue, instead of immorality.

EXPERIMENT 3: IMMORALITY AND VIRTUE

The first two experiments provide evidence that that people jump to conclusions of wrongdoing in dyadic situations. However, one possible explanation for this pattern is that people simply jump to morally relevant conclusions when an agent and patient are present—that is, we may assume that people are more likely to both help and hurt in a dyadic context, compared to a nondyadic context. In Experiment 3, participants either gave ratings of immorality (Immoral–Not Immoral) or virtue (Virtuous–Not Virtuous) for both dyadic and nondyadic sentences. We predicted that people would show a bias toward immoral responses, but not virtuous responses, when sentences were dyadic.

Method

Participants and Design

We recruited 102 participants via MTurk (57.7% female, M_{age} = 35 years), who completed a 2 (Rating Type: Immoral, Virtuous) by 2 (Patient: Absent, Present) between-within experiment. One participant did not respond to over 50% of the trials and skipped the demographics; his or her data were excluded from the analyses.

Procedure and Materials

Sentence manipulations. Participants again read sentences with nonsense actions. As in Experiment 2, a patient was either absent or present in each sentence.

Sentence categorization task. Participants categorized each of the sentences as either Immoral–Not Immoral or Virtuous–Not Virtuous. The first eight sentences were presented as practice trials and were not included for analysis. We asked participants to provide each of their responses within five seconds; this amount of time proved ample, with participants successfully categorizing sentences in 98.4% of trials with an average latency of 1.08 seconds.

Results and Discussion

We predicted that participants would show the immorality bias, such that Immoral ratings were higher when the Patient was Present. We also predicted that no such effect would emerge for Virtuous ratings. We used a fully cross-classified linear model to analyze the effects of Rating Type (Immoral, Virtuous) and Patient (Absent, Present) on participants' ratings.

The analysis revealed no main effect of Rating Type, $F(1, 6891) = .39$, $p = .54$, as well as a main effect of Patient, $F(1, 6891) = 84.43$, $p < .001$. The main effect of Patient was qualified by a Rating Type x Patient interaction, $F(1, 6891) = 291.74$, $p < .001$. The pattern found in Experiment 2 replicated: Immoral ratings were higher when the patient was Present ($M_{pct} = 57.1$, 95% CI [54.7, 59.4]) than when the patient was Absent ($M_{pct} = 26.0$, 95% CI [23.9, 28.1]). This finding suggests that, in a dyadic context, people assume immorality. They do not, however, appear to assume virtue in the same way: Virtue ratings were actually higher when the patient was Absent ($M_{pct} = 44.7$, 95% CI [42.4, 47.0]) rather than Present ($M_{pct} = 35.1$, 95% CI [32.9, 37.4]). This reversal is likely explained by the opposition of immorality and virtue: if actions seem more immoral in a dyadic context, then they also seem less virtuous. Furthermore, a comparison of Immoral and Virtue ratings when the patient was Present shows that ratings of immorality ($M_{pct} = 57.1$, 95% CI [54.7, 59.4]) are higher than ratings of virtue ($M_{pct} = 35.1$, 95% CI [32.9, 37.4]) in a dyadic context, again consistent with the immorality bias. See Figure 1 for means and confidence intervals.

These results of Experiment 3 suggest that the immorality bias—a tendency to assume wrongdoing in ambiguous social situations—cannot be explained by a more general "morality bias” that compels people to assume both virtuous and immoral actions in a dyadic context. Ratings of immorality were higher than those of virtue when the patient was present; they were also higher when the patient was present, rather than absent, consistent with dyadic morality.

EXPERIMENT 4: INTENTIONAL AGENTS

Dyadic morality suggests that intention is a key element of the moral template, and so the immorality bias should be stronger when acts are dyadic and clearly intentional. In this study, we therefore manipulated both whether the patient was absent or present, and whether the agent acted intentionally ("John intentionally pelled Mary"), accidentally ("John accidentally pelled Mary"), or ambiguously ("John pelled Mary"). We predicted that ratings of immorality would be highest for intentional actions with a patient ("John intentionally pelled Mary"), and lowest for accidental actions without a patient ("John accidentally pelled").

The predictions regarding the other conditions are more complex. Although the most canonical immoral actions feature a complete dyad with intention and suffering, the mind often fills in other elements of the dyad when they are left unspecified. This is the phenomenon of "dyadic completion" in which immoral contexts prompt people to perceive either intentional agents or suffering patients in incomplete dyads of either random suffering or "victimless" crimes (DeScioli, Gilbert, & Kurzban, 2012; Gray, 2012; Gray et al., 2014; Gray & Wegner, 2010; Shweder, Much, Mahapatra, & Park, 1997). In other words, dyadic completion nudges acts with some elements of immorality (e.g., the dyad) to seem to include more elements of immorality (e.g., intention).

Therefore, in our study, we expected that when a patient was present but intention was ambiguous, participants' would perceive some intent—and therefore give immorality judgments closer to "intention present" than "intention absent." On the other hand, in the absence of a dyadic context (patient absent), we expected lesser assumptions of intention (and immorality).

Revealing this pattern of results would support the existence of an immorality bias that "assumes the worse" when morally-relevant—but ambiguous—elements are present.

Method

Participants and Design

We recruited 81 participants via MTurk (58.0% female, M_{age} = 37 years), who completed a 3 (Intention: Intentional, Accidental, Ambiguous-Intent) by 2 (Patient: Absent, Present) within-subjects experiment. No participants' data were excluded from the study.

Procedure and Materials

Sentence manipulations. Participants again read sentences with nonsense actions. As in Experiments 2 and 3, a patient was either absent or present in each sentence. In order to further reduce ambiguity about the absence of presence of a patient, we added "by himself/herself" when the patient was absent (e.g., "John pelled by himself").

We also manipulated agent intention. Participants read sentences with clear intentional action (e.g., "John intentionally/willfully/purposely pelled Mary"), clear accidents ("John accidentally/unintentionally/inadvertently pelled Mary"), and Ambiguous-Intent actions ("John pelled Mary"). Because there were 76 total trials, each-within subjects cell included 12 or 13 sentences (*M* = 12.67). The multilevel framework used to analyze the data allowed us to easily account for differences in trial numbers.

Sentence categorization task. Participants categorized each of the sentences as Immoral or Not Immoral. The first eight sentences were presented as practice trials and were not included for analysis. We asked participants to provide each of their responses within six and a half seconds; this amount of time proved ample, with participants successfully categorizing sentences in 99.5% of trials with an average latency of 1.48 seconds.

Results and Discussion

Intention Increases the Immorality Bias

Because intention is an important element of the moral template, we predicted that ratings of immorality would be highest for Intentional acts and lowest for Accidental acts. We again used a fully cross-classified linear model to analyze the effects of Patient (Absent, Present) and Intention (Intentional, Accidental, Ambiguous-Intent) on participants' ratings of immorality.

The analysis revealed the predicted main effect of Intention, $F(2, 5636) = 383.96$, $p < .001$. Participants rated Intentional actions ($M_{pct} = 60.3$, 95% CI [58.0, 62.6]) as more immoral than Ambiguous-Intent actions ($M_{pct} = 42.2$, 95% CI [39.8, 44.7]), which in turn were rated as more immoral than Accidental actions ($M_{pct} = 11.6$, 95% CI [10.2, 13.2]). These results support our basic prediction that intentional acts increases the immorality bias, whereas accidents mitigate it.

The analysis of fixed effects revealed a main effect of Patient, $F(1, 5636) = 304.02$, $p < .001$. Participants categorized sentences as immoral more often when the patient was Present ($M_{pct} = 48.8$, 95% CI [46.6, 51.1]) than when the patient was Absent ($M_{pct} = 22.6$, 95% CI [21.0, 24.3]), replicating the finding in Experiments 2 and 3. See Figure 1 for means and confidence intervals.

Intention Matters More for Dyads

We further predicted that the effect of Intention would be especially strong when the patient was Present. The main effects in the analysis were qualified by a Dyad x Intention interaction, $F(2, 5636) = 21.86$, $p < .001$. Although there was a significant difference between ratings of Intentional and Accidental actions when the patient was Absent, ($M_{diffpct} = 34.1$, 95% CI [30.5, 37.8]), this difference was even larger when the patient was Present ($M_{diffpct} = 60.6$, 95% CI [57.0, 64.2]). Intention impacted ratings of immorality more strongly in a dyadic context, as predicted by dyadic completion (Gray et al., 2014).

People Assume Immoral Intent

Although it is noteworthy that Intention is more immoral in a dyadic context, what is more important for the immorality bias is that people assume immoral intent when intention is ambiguous. We predicted that people would assumes immoral intent when a patient was Present, such that immoral ratings of Ambiguous-Intent actions more closely resemble ratings of Intentional actions than Accidental actions. In other words, even when people had no clear information about intent, we predicted that they would nevertheless assume its presence in dyadic contexts, rather than assuming innocent accidents.

To test whether ratings of Ambiguous-Intent actions more closely resemble ratings of Intentional actions when the patient is Present, we compared the confidence intervals of the difference scores for [Intentional – Ambiguous-Intent] and [Ambiguous-Intent – Accidental]. If the confidence intervals for two estimates do not overlap at 95%, then the values are significantly different.[4] When standard inferential tests are inaccessible due to the constraints of statistical programs, comparing confidence intervals has been established as an acceptable alternative for significance testing (MacGregor-Fors & Payton, 2013; Payton, Greenstone, & Schenker, 2003). These comparisons were conducted for when the patient was Absent and when the patient was Present.

---

[4] However, if the 95% confidence intervals for two estimates do overlap, they may still be significant. These instances can be clarified using 84% confidence intervals; checking whether or not these intervals overlap approximates a significance test with an alpha of .05 (MacGregor-Fors & Payton, 2013; Payton, Greenstone, & Schenker, 2003). Although the validity of this approach has been debated for within-subjects models, the intervals generated in a multilevel framework should allow for accurate comparison, given a balanced design (Baguley, 2012).

When the patient was Present, Ambiguous-Intent ratings more closely resembled Intentional ratings ($M_{diffpct}$ = 12.2, 95% CI [8.1, 16.4]) than Accidental ratings ($M_{diffpct}$ = 48.4, 95% CI [44.5, 52.2])—that is, ratings of Ambiguous-Intent actions were more similar to ratings of Intentional actions, consistent with the immorality bias (and with dyadic morality). When the patient was Absent, no bias emerged as Ambiguous-Intent ratings resembled both Intentional ratings ($M_{diffpct}$ = 19.4, 95% CI [15.2, 23.6]) and Accidental ratings ($M_{diffpct}$ = 14.7, 95% CI [11.4, 18.0])[5] by approximately the same amount. This finding supports the idea that participants assume immoral intent when a patient is Present, even when the situation is more ambiguous. This is consistent with the immorality bias in which people automatically jump to immoral conclusions unless the action is unambiguously described as accidental.

---

[5] The non-significance of this difference was clarified by finding that the 84% CIs still overlap, [16.6, 22.7] and [12.5, 17.3].

EXPERIMENT 5: SUFFERING PATIENTS

Although intention is an important component of immorality, many intentional actions are perfectly benign. It is intentional actions that seem to cause suffering that best fits the dyadic moral template. Thus, the immorality bias should be stronger when actions appear to cause suffering and substantially weaker when actions do not appear to cause suffering. Furthermore, dyadic completion—coupled with the immorality bias—could mean that the mere presence of suffering in a dyadic context may prompt robust judgments of immorality.

To test this idea, participants categorized nonsense actions as in the previous experiment. All actions were dyadic in nature (i.e., all had a patient present), and we manipulated both the intention of the agent and whether the patient suffered ("John pelled Mary, who cried"), benefited ("John pelled Mary, who laughed"), or gave no clear reaction ("John pelled Mary").

We predicted that ratings of immorality would be highest with a suffering patient and lowest with a benefiting patient, especially when the action was clearly intentional. We also predicted that participants' would assume suffering and intention when its presence is ambiguous, especially when other dyadic elements are present. In other words, when suffering is present, people should assume intention, and when intention is present, people should assume suffering. These findings would support an immorality bias: people assume the elements of immorality when they are ambiguous to the extent that the context is dyadic.

Method

Participants and Design

We recruited 64 participants via MTurk (51.6% female, M_{age} = 39 years), who completed a 3 (Suffering: Suffering, Benefiting, Ambiguous-Suffering) by 3 (Intention: Intentional, Accidental, Ambiguous-Intent) within-subjects experiment. No participants' data were excluded from the study.

Procedure and Materials

Sentence manipulations. Participants read sentences with nonsense actions. As in Experiment 3, participants read sentences with clear intentional action, clear accidents, and Ambiguous-Intent actions.

We also manipulated patient suffering. Participants read sentences with a clearly suffering patient (e.g., "John pelled Mary, who cried/shuddered/screamed/yelled/sobbed"), a clearly benefiting patient ("John pelled Mary, who laughed/smiled/grinned/beamed/nodded"), and Ambiguous-Suffering patients ("John pelled Mary"). Because there were 76 total trials, each-within subjects cell included either 8 or 9 sentences (*M* = 8.44). The multilevel framework used to analyze the data allowed us to easily account for differences in trial numbers.

Sentence categorization task. Participants categorized each of the sentences as Immoral or Not Immoral. The first eight sentences were presented as practice trials and were not included for analysis. We asked participants to provide each of their responses within eight seconds; this amount of time proved ample, with participants successfully categorizing sentences in 99.3% of trials with an average latency of 1.78 seconds.

Results and Discussion

Suffering Increases the Immorality Bias

Because suffering is a key element of morality, we predicted that participants' immoral ratings would be highest for suffering patients. We again used a fully cross-classified linear model to analyze the effects of Intention (Intentional, Accidental, Ambiguous-Intent) and Suffering (Suffering, Benefiting, Ambiguous-Suffering) on participants' ratings of immorality.

The analysis of fixed effects revealed the predicted main effect of Suffering, $F(2, 4311) = 295.27$, $p < .001$, such that participants rated sentences with Suffering patients ($M_{pct} = 67.9$, 95% CI [65.2, 70.6]) as more immoral than those with Ambiguous-Suffering patients ($M_{pct} = 38.7$, 95% CI [35.9, 41.6]), which in turn were rated as more immoral than those with Benefiting patients ($M_{pct} = 17.3$, 95% CI [15.4, 19.4]). These results suggest that a suffering patient increase the immorality bias, whereas a benefiting patient mitigates it.

The analysis of fixed effects revealed a main effect of Intention, $F(2, 4311) = 182.27$, $p < .001$, such that participants rated Intentional actions ($M_{pct} = 56.4$, 95% CI [53.4, 59.4]) as more immoral than Ambiguous-Intent actions ($M_{pct} = 48.7$, 95% CI [45.7, 51.8]), which in turn were rated as more immoral than Accidental actions ($M_{pct} = 18.6$, 95% CI [16.5, 20.8]). This result replicates the finding in Experiment 4 that intention amplifies the immorality bias. See Figure 2 for means and confidence intervals.

Suffering Matters More for Intentional Acts

We further predicted that the effect of Suffering would be especially strong when acts were Intentional (further completing the moral template). The main effects in the analysis were qualified by an Intention x Suffering interaction, $F(4, 4311) = 12.73$, $p < .001$. As predicted, the difference in participants' ratings of Suffering and Benefiting patients was greater in the

Intentional condition ($M_{diffpct}$ = 58.2, 95% CI [53.0, 63.4]) than in the Accidental condition ($M_{diffpct}$ = 23.5, 95% CI [18.3, 28.6]).

People Assume Suffering Victims

The immorality bias specifically suggests that people assume the presence of a suffering patient even when the patient's experience is ambiguous, and rate ambiguous sentences as immoral—especially when that action is Intentional. The data supported this prediction: when acts were Intentional, Ambiguous-Suffering ratings more closely resembled Suffering ratings ($M_{diffpct}$ = 19.9, 95% CI [14.3, 25.5]) than Benefiting ratings ($M_{diffpct}$ = 38.3, 95% CI [32.5, 44.1]). Unless the patient is clearly benefiting, people assume immorality.

Intention Matters More for Suffering Patients

Just as information about Intention influences how Suffering changes people's responses, we also predicted that information about Suffering would change the influence of Intention. Specifically, we predicted that the effect of Intention would be especially strong when the patient was Suffering. The data supported this prediction: there was a greater difference in participants' ratings of Intentional and Accidental actions in the Suffering condition ($M_{diffpct}$ = 46.8, 95% CI [41.2, 52.3]) than in the Benefiting condition ($M_{diffpct}$ = 12.0, 95% CI [7.3, 16.8]).

People Assume Immoral Intent

The immorality bias specifically suggests that that people assume immoral intent even when the agent's intentions are unclear, especially when the patient is Suffering. The data again supported our prediction: when the patient was Suffering, Ambiguous-Intent ratings more closely resembled Intentional ratings ($M_{diffpct}$ = 1.5, 95% CI [-3.5, 6.4]) than Accidental ratings ($M_{diffpct}$ = 45.3, 95% CI [39.7, 50.8]). In fact, as shown by the confidence intervals, we found no difference between Intentional and Ambiguous-Intent ratings, suggesting that the presence of suffering powerfully implies the presence of immoral intent.

The findings of Experiment 5 converge to support an immorality bias that is amplified by the presence of intention and suffering. When one of these elements is present, people not only weigh the other element more heavily when making judgments of immorality, but also simply assume that the other element is present. In this way, people “complete” the moral dyad and assume intention and suffering—and thus immorality—even when the action itself remains ambiguous.

EXPERIMENT 6: TIME PRESSURE AND REAL ACTIONS

Biases in judgment are typically stronger when people are placed under either cognitive load (Goldinger, Kleider, Azuma, & Beike, 2003; Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008) or time pressure (Finucane, Alhakami, Slovic, & Johnson, 2000; Rosset, 2008), especially for ambiguous stimuli. In this experiment, we varied the amount of time available to participants to categorize sentences as immoral or not immoral, with the prediction that participants under time pressure would express the immorality bias more strongly by categorizing nonsense actions as more immoral.

Additionally, we added two unambiguous action categories—harmful actions and helpful actions—for two reasons. First, we wanted to compare overall ratings of nonsense actions to those of harmful and helpful actions. Consistent with the immorality bias, we predicted that nonsense actions would elicit ratings that more closely resemble harmful actions than helpful actions. Second, we did not expect time pressure to influence participants' ratings of harmful actions, since these stimuli are already unambiguous immoral.

Method

Participants and Design

We recruited 110 participants via MTurk (56.5% female, M_{age} = 38 years), who completed a 3 (Action: Nonsense, Harmful, Helpful) by 2 (Speed: Fast, Slow) within-between experiment. No participants' data were excluded from the study.

Procedure and Materials

Sentence categorization task. Participants again read sentences and categorized each sentences as either Immoral or Not Immoral. Unlike previous experiments, some of the actions included in these sentences were real actions, both harmful (e.g., killed, slapped, threatened) and helpful (e.g., accepted, hugged, romanced). These actions were chosen to be very clearly immoral or very clearly not immoral, both to serve as objective comparison points for the nonsense actions and to test the effect of time pressure on unambiguous targets. In total, participants rated 30 nonsense actions, 30 harmful actions, and 30 helpful actions. The first twelve sentences were practice trials and were not included for analysis. See Appendix C for a full list of harmful and helpful actions.

Time pressure manipulation. Participants had either 1.5 seconds (Fast) or 5 seconds (Slow) to categorize each sentence. To check the effectiveness of the time manipulation, we used a linear model to test whether Speed influenced response latency. We found a significant effect of time pressure, $F(1, 8629) = 632.28$, $p < .001$, such that participants in the Fast condition responded to the sentences more quickly (771ms) than those in the Slow condition (1248ms). Overall, participants successfully responded 96.2% of trials, suggesting that participants had adequate time to respond.

Results and Discussion

Harmful and Helpful Actions Are Unambiguous

The harmful and helpful actions were chosen to serve as unambiguous stimuli for comparison. We used a hierarchical linear model to analyze the effects of Action Type (Negative, Nonsense, Positive) and Speed (Fast, Slow) on participants’ ratings of immorality. The analysis of fixed effects revealed a main effect of Verb Type, $F(2, 8300) = 1127.04$, $p <$

.001, such that participants rated sentences with Harmful actions (M_{pct} = 90.3, 95% CI [87.7, 92.4]) as more immoral than those with Nonsense actions (M_{pct} = 58.2, 95% CI [51.8, 64.3]), which in turn were rated as more immoral than those with Helpful actions (M_{pct} = 5.5, 95% CI [4.2, 7.1]). The extreme ratings and small confidence intervals for the Harmful and Helpful actions suggest that participants had little uncertainty categorizing these stimuli.

Nonsense Actions are More Immoral under Time Pressure

The current study primarily tested whether participants expressed the immorality bias more strongly under time pressure, as is typically the case with judgmental biases. In particular, we predicted that the immorality bias would influence judgments of nonsense actions more strongly under time pressure.

The analysis of fixed effects did not reveal a main effect of Speed, $F(1, 8300) = 1.79$, $p = .18$. However, the main effects were qualified by a significant interaction, $F(2, 8300) = 39.00$, $p < .001$. An analysis of simple effects showed that the effect of Speed was significant for Nonsense actions, $t(8300) = 2.87$, $p = .004$, such that participants in the Fast condition categorized more sentences as immoral (M_{pct} = 66.8, 95% CI [58.2, 74.4]) than participants in the Slow condition (M_{pct} = 49.0, 95% CI [40.2, 58.0]), suggesting that time pressure amplifies the immorality bias. See Figure 3 for means and confidence intervals.

Furthermore, the effect of Speed was significant in the opposite direction for Negative verbs, $t(8300) = -2.06$, $p = .04$, such that participants in the Fast condition categorized fewer sentences as immoral (M_{pct} = 87.4, 95% CI [82.6, 91.0]) than participants in the Slow condition (M_{pct} = 92.6, 95% CI [89.5, 94.8]). These results support our prediction that time pressure would amplify the immorality bias for Nonsense actions, but not for Harmful actions. In fact, Harmful actions showed the opposite pattern, suggesting that these actions are unambiguous and that time pressure simply caused a loss of accuracy.

A similar "loss of accuracy" effect occurred for Helpful actions: the effect of Speed was significant for Helpful actions, $t(8300) = 2.77$, $p = .006$, such that those in the Fast condition categorized more sentences as immoral ($M_{pct} = 8.2$, 95% CI [5.7, 11.7]) than those with more time to respond ($M_{pct} = 3.6$, 95% CI [2.4, 5.3]). Overall, these results suggest that the immorality bias influences judgments more strongly under time pressure, but only for ambiguous stimuli.

Nonsense Actions Resemble Harmful Actions

Finally, we predicted that ratings of Nonsense actions would more closely resemble ratings of Harmful actions than Helpful actions. A comparison of difference scores shows that Nonsense actions more closely resemble Harmful actions ($M_{diffpct} = 32.1$, 95% CI [27.7, 36.6]) than Helpful actions ($M_{diffpct} = 52.7$, 95% CI [47.6, 57.9]). Nonsense actions are perceived as more similar to harmful actions than helpful actions, further suggesting that people show an immorality bias in response to ambiguous dyadic situations.

These results suggest that people's moral judgments rely more strongly on the immorality bias when they have less time to think, suggesting that the immorality bias exhibits one of the common characteristics of biases. The next experiment attempts to clarify the processes underlying this effect of time pressure using a process dissociation procedure.

EXPERIMENT 7: PROCESS DISSOCIATION AND SHORT SENTENCES

Just as judgmental biases are often influenced by factors such as cognitive load and time pressure, so too are these effects often understood using dual process models that include both controlled and automatic pathways. Contextual factors such as cognitive load and time pressure typically inhibit controlled responding, rather than increasing automatic assumptions. Though the previous experiment showed that time pressure can increase participants' expression of the immorality bias, it did not pinpoint the mechanism of this effect.

To test whether the effect of time pressure is explained by a shift in controlled responding, we designed an experiment that allowed us to use a process dissociation procedure to differentiate automatic and controlled processes (see Table 4; Jacoby, 1991; Payne, 2001). In order to use a process dissociation procedure, we created two sets of short sentences: *Probably Immoral* or *Possibly Immoral*. This procedure allowed us to separately test the influence of time pressure on automatic and controlled processes (see Table 4; Jacoby, 1991; Payne, 2001). Automatic processes require little cognitive effort and operate regardless of conscious intent. Controlled processes, on the other hand, are consciously executed and require greater cognitive effort; these processes can be disturbed by time pressure, cognitive load, and depleted cognitive resources. People attempt to respond using controlled processes, but are often unable to.

We predicted that participants who are unable to use controlled processes will instead rely on an automatic assumption of immorality. In particular, this bias will be influential when controlled and automatic processes are expected to yield opposite outcomes. That is, for the Possibly Immoral sentences, an automatic assumption of immorality would lead subjects to respond “immoral,” but a thoughtfully controlled response would lead them to respond that it is not immoral. These results would suggest that people tend to automatically assume wrongdoing—the immorality bias— and that this assumption is more likely to lead to errors when it is difficult to exert effortful control over responses.

Method

Participants and Design

We recruited 104 participants through MTurk, who completed a 2 (Sentence Type: Probably Immoral, Possibly Immoral) by 2 (Speed: Fast, Slow) within-between experiment. Six of these participants either failed to respond to three or more of the sentences in a given category or encountered technical difficulties; these participants were excluded, leaving 98 participants, 4 of whom did not provide demographic information but completed all other elements of the experiment (52.1% female, M_{age} = 33 years).

Procedure and Materials

Piloting the short sentence sets. To create a set of ambiguous immoral sentences for this experiment, two 60-participant groups recruited through MTurk categorized short sentences as either “Not Immoral” or “Immoral.”[6] These participants were not placed under any time pressure. We created two sets of stimuli, “Probably Immoral” and “Possibly Immoral,” by sorting the scenarios by the percentage of immoral responses and creating two sets with an average of approximately 75% immoral responses (Probably Immoral) and 25% immoral responses (Possibly Immoral). The final sets included 14 sentences in each category.

Short sentence categorization task. Participants received instructions to categorize each sentence that flashed on the screen as either “Not Immoral” or “Immoral.” Participants received

---

[6] See the supplemental materials for the full set of short sentences included in the pilot study.

either 1500ms or 5000ms to complete each trial. Participants completed 14 practice trials using a set of practice items, then completed 28 main trials in which the sentences from the Probably Immoral and Possibly Immoral sets were randomly presented.

Time pressure manipulation check. As in Experiment 6, we used a linear model to check the effectiveness of our time manipulation. We found that Speed (Fast, Slow) significantly influenced response latency, $F(1, 2908) = 104.34$, $p < .001$, such that participants in the Fast condition responded to the sentences more quickly (897ms) than those in the Slow condition (1819ms). Participants provided responses for 92.3% of trials, suggesting that they had adequate time to respond.

Results and Discussion

Immoral Responses

We used a fully cross-classified hierarchical linear model to analyze the effects of Sentence Type (Probably Immoral, Possibly Immoral) and Speed (Fast, Slow) and on participants' ratings of immorality. Unsurprisingly, the analysis of fixed effects showed a main effect of Sentence Type, $F(1, 2684) = 370.82$, $p < .001$, such that participants rated Probably Immoral sentences as more immoral ($M_{pct} = 74.9$, 95% CI [72.4, 77.2]) than Possibly Immoral sentences ($M_{pct} = 36.2$, 95% CI [33.5, 38.9]). The analysis of fixed effects also revealed a main effect of Speed, $F(1, 2684) = 22.25$, $p < .001$, such that participants with less time to respond categorized more sentences as immoral ($M_{pct} = 61.7$, 95% CI [58.5, 64.7]) than those with more time to respond ($M_{pct} = 51.3$, 95% CI [48.3, 54.2]).

These main effects were qualified by a significant interaction, $F(1, 2684) = 16.50$, $p < .001$. An analysis of simple effects showed that the effect of Speed was significant for Possibly Immoral verbs, $t(2684) = 6.56$, $p < .001$, such that participants with less time to respond categorized more sentences as immoral (M_{pct} = 45.5, 95% CI [41.4, 49.7]) than those with less time to respond (M_{pct} = 27.8, 95% CI [24.6, 31.2]). However, the effect of Speed was not significant for Probably Immoral, $t(2684) = 0.58$, $p = .56$.[7] See Figure 4 for means and confidence intervals.

These results fit a control-impairment explanation: participants who have an automatic tendency to assume immorality will do so for both kinds of sentences, but this assumption will be opposed by more controlled thinking when the statement is likely to be nonmoral. Controlled processes are more likely to fail under fast responding, leading subjects to incorrectly “guess” that a sentence is immoral more often for the Possibly Immoral sentences than for the Probably Immoral sentences. We more directly addressed this possibility by analyzing the data using a process dissociation procedure.

Process Dissociation

In order to more directly test the mechanisms that account for the effect of Speed, we calculated two dependent variables—controlled processing and automatic assumption—using the guidelines provided in Payne (2001). These estimates can be dissociated because the experiment includes both congruent trials, in which controlled and automatic processes lead to the same answer, and incongruent trials, in which controlled and automatic processes lead to different answers. When a trial is congruent, the probability of responding that a sentence is “Immoral” is the probability of Control, C, plus the probability of assuming immorality when control fails, A(1 – C):

$$\text{Congruent} = C + A(1 - C). \tag{1}$$

---

[7] The dichotimized categories of Possibly Immoral and Probably Immoral were necessary for process dissociation. However, dichotomizing variables can raise concerns due to lost information or variability about the specific stimuli. To address this concern, additional analyses using the pilot test’s Percent values instead of dichotomized Categories is available in the Supplemental Materials.

In this experiment, Probably Immoral trials are congruent, since both controlled processing and automatic assumptions lead to answering “Immoral.” Possibly Immoral trials, on the other hand, are incongruent, since controlled processing leads to answering “Not Immoral,” but automatic assumptions lead to answering “Immoral.” The probability of answering “Immoral” for an incongruent trial is the probability of assuming immorality, A, whenever control fails, (1 – C):

$$\text{Incongruent} = A(1 - C). \tag{2}$$

These equations for congruent and incongruent trials allow for the separation of controlled and automatic processing. Estimates of controlled processing represent a person’s ability to intentionally provide a certain response (i.e., “Immoral”) when they intend to, and not provide that response when they do not intend to. A higher estimate indicates greater controlled processing across all trials. The control estimate is the difference between answering “Immoral” in congruent and incongruent trials:

$$C = \text{Congruent} - \text{Incongruent}. \tag{3}$$

On the other hand, estimates of automatic assumption represent a person’s tendency to provide a certain response (i.e., Immoral”) regardless of whether or not that response aligns with controlled processing. A higher automatic estimate indicates a stronger bias toward immorality. Solving for an estimate of control allows the automatic estimate to be solved:

$$A = \text{Incongruent}/(1 - C). \tag{4}$$

If the immorality bias is driven by a stable automatic assumptions of wrongdoing, then the Speed condition should influence people’s ability to engage in controlled processing (i.e., their ability to accurately categorize the sentence based on their content and counteract their automatic assumptions), but not their automatic assumptions (i.e., their stable tendency to categorize sentences as immoral)..

Excluding one outlier (Cook's D = .16), a one-way ANOVA analyzing controlled processing revealed the expected effect of Speed, $F(1, 95) = 16.62$, $p < .001$, $\eta_p^2 = .15$, such that participants with less time to respond showed lower levels of controlled processing ($M = .31$, $SD = .24$) than those with more time to respond ($M = .47$, $SD = .16$).[8] A one-way ANOVA examining automatic assumption showed a marginal effect of Speed, $F(1, 96) = 3.22$, $p = .08$, $\eta_p^2 = .03$, such that participants with less time to respond showed greater automatic assumptions of immorality ($M = .65$, $SD = .27$) than those with more time to respond ($M = .54$, $SD = .32$). This result suggests that when subjects had little time to respond, they exerted less control and also relied more heavily on their automatic intuitions. Overall, the process dissociation analysis suggests that the immorality bias is a stable tendency to automatically assume wrongdoing, and that controlled processing can be used to override this initial assumption when cognitive resources are available.[9]

---

[8] Including the outlier still yielded a main effect of Speed, $F(1, 96) = 11.34$, $p = .001$, $\eta_p^2 = .11$.

[9] Two replications of the results of Experiment 6 are available in the supplemental materials. The first replication only includes Clearly Nonmoral and Possibly Immoral sentences to address concerns about semantic priming. The second replication includes Clearly Nonmoral, Possibly Immoral, Probably Immoral, and Clearly Immoral items to examine the immorality bias across different sentence types.

GENERAL DISCUSSION

In seven experiments, we demonstrated the immorality bias: people assume wrongdoing in ambiguous social situations. We observed the bias in response to vignettes (Experiment 1) as well as nonsense action sentences, which also showed that the bias emerges primarily when both an agent and a patient are present (Experiment 2) and emerges for ratings of immorality, but not ratings of virtue (Experiment 3). Furthermore, we found that information about intention (Experiment 4) and suffering (Experiment 5) can amplify the immorality bias, especially within a dyadic context. Finally, we found that participants express the immorality bias more strongly under time pressure (Experiment 6), and that this effect is best understood by conceptualizing the bias as a stable tendency to automatically assume wrongdoing that can be counteracted using controlled processes (Experiment 7). As a whole, these studies reveal people's tendency to assume wrongdoing in a variety of ambiguous contexts. Moreover, these assumptions build off each other, such that the presence of one ambiguous factor (e.g., suffering) leads to more immoral assumptions of another ambiguous factor (e.g., intention). See Figure 6 for a summary of findings.

These studies highlight an important phenomenon that has practical consequences in domains such as law enforcement, team management, education, and parenting. From a theoretical perspective, the existence of the immorality bias suggests that moral judgment can be understood as heuristic judgments that use a dyadic template. That the bias responds so consistently with the dyadic template adds to evidence in support of the Theory of Dyadic

Morality, which suggests that the elements of perceived harm—causal dyad, intention, suffering—all causally contribute to intuitive and automatic moral judgment.

There are, of course, other moral considerations when making deliberative and effortful decisions, such as philosophical beliefs systems such as utilitarianism or deontology, perceived base rates of a given immoral action, and abstract concepts such as “purity” or “social order” (which research has nevertheless rooted in a dyadic template; Schein & Gray, 2015). However, most of our moral judgment are intuitive (Haidt, 2001)—and heuristic, which leads them to be biased toward guilt, rather than innocence in the right context. Given that context is pervasive—the presence of two people—the immorality bias is also likely to be pervasive. Unless an action is clearly accidental or clearly benefits someone, people may well assume that action is immoral, especially when they have little time or motivation for deliberate thought.

Caveats

These findings are not without limitations. Although our study benefits from the greater diversity of race, gender, and age afforded by recruiting participants through MTurk (Buhrmester et al., 2011; Goodman et al., 2012), we nevertheless acknowledge the use of a relatively WEIRD (White, educated, industrialized, rich, democratic; Henrich, Heine, & Norenzayan, 2010) sample of participants. For this reason, cross-cultural examinations of the immorality bias would provide useful insight into the generalizability of our findings. We also acknowledge that the landscape of moral wrongs is remarkably diverse, as shown by theories that highlight extensive variety in moral rules and judgments (Haidt, 2013; Shweder, Mahapatra, & Miller, 1987; Shweder et al., 1997). The present research does not directly test whether the immorality bias occurs across all types of moral transgressions. Nevertheless, Experiments 1 and 7 included situations suggestive of infidelity, incest, rape, pedophilia, theft, assault, housebreaking, lying, and kidnapping.

Furthermore, the substantial ambiguity of the nonsense actions used in Experiments 2-6 reveals that the immorality bias is not limited to any specific class of moral actions.

Actions and attributions are not the only elements of the scenarios that might raise questions about the generalizability of the effect. For one, the scenarios we used always focused on two human individuals: one human agent acting on one human patient. Although this structure represents the most common instances of wrongdoing, humans can also act on other humans as unified groups (Cohen, Montoya, & Insko, 2006; Waytz & Young, 2012), and nonhuman entities such as animals (Bastian, Loughnan, Haslam, & Radke, 2012; Epley, Waytz, Akalis, & Cacioppo, 2008) machines (Melson et al., 2009; Waytz, Heafner, & Epley, 2014) can also assume a role in moral judgments. The current research does not address these nonhuman entities. Additionally, the present research focused on presumably adult individuals acting on other adults (or sometimes children). This structure is congruent with people's prototypical understanding of morality and allows our research to generalize to many common scenarios. However, whether the immorality bias generalizes to atypical scenarios, such as a small child ambiguously acting on an adult, is an open (and interesting) theoretical question.

Theoretical Implications

Adaptiveness. In the context of evolution, the immorality bias may be an adaptive heuristic. Evolutionary arguments for moral processes often focus on altruism (e.g., Bowles, 2006; de Waal, 2008), which includes specific mechanisms such as kin selection (Hamilton, 1963) and reciprocity (Trivers, 1971). These arguments attempt to explain why people aid and cooperate with others. Recent work has also addressed possible evolutionary explanations for moral condemnation, which instead concerns why people condemn and punish actions that violate moral rules (DeScioli & Kurzban, 2013). Moral condemnation may facilitate dynamic coordination, a process in which a person's actions serve as a public signal to bystanders that determines which side they choose in a conflict. Both altruism and dynamic coordination concern group fitness—the idea that some characteristics are evolutionarily adaptive because they help a group survive to reproduce.

Engaging in either altruism or dynamic coordination requires the identification of a person who needs helps (a victim) or a person to side against (a perpetrator). To the extent that the effectiveness of these actions is time-sensitive (i.e., helping and condemning should happen sooner, rather than later, for the best outcomes), the immorality bias allows for the quick identification of victims and perpetrators, allowing people to effectively help and condemn others. Though these assumptions are not always accurate, false positives may be less costly than false negatives, making the immorality bias a useful heuristic. In this way, the immorality bias enhances group fitness in the same way that threat detection and agency detection enhance individual fitness: detecting wrongdoing or agency when it is absent is not very costly, but failing to detect wrongdoing or agency when it is present can be extremely costly (Barrett & Behne, 2005; Öhman et al., 2001).

Research also shows that people are highly motivated to evaluate the moral character of other individuals, as well as other groups (Brambilla, Rusconi, Sacchi, & Cherubini, 2011; Goodwin, 2015; Goodwin, Piazza, & Rozin, 2014; Pizarro & Tannenbaum, 2011). On a group level, the immorality bias may preserve group cohesion by facilitating the quick condemnation of harmful members and preservation of the valuable members who are targets of harm. Additionally, condemning harmful members may serve an impression management function: people's impressions of groups rely more heavily on judgments of morality than on judgments of warmth and competence (Brambilla, Sacchi, Rusconi, Cherubini, & Yzerbyt, 2012), and people actually prioritize moral judgments of ingroup members because of concerns about group image (Brambilla, Sacchi, Pagliaro, & Ellemers, 2013). Group fitness is likely enhanced by maintaining a group image that includes honesty and trustworthiness: other groups are then more likely to trade, maximizing resource utility for both groups, and are also less likely to attack, minimizing the chance of physical conflict and death.

Moral cognition. Much of the previous research on immorality has focused on what it “means” for something to be immoral: what characteristics unify or distinguish immoral acts (Graham et al., 2011; Gray et al., 2012), what makes these acts more or less blameworthy (Cushman, Young, & Hauser, 2006; Pizarro, Uhlmann, & Salovey, 2003), what emotions are associated with these acts (Cameron et al., 2015; Gray & Wegner, 2011; Rozin, Lowery, Imada, & Haidt, 1999), and how these acts impact judgments of character (Critcher, Inbar, & Pizarro, 2012; Goodwin et al., 2014). However, the immorality bias addresses the more basic question of when and how actions enter the moral domain. It suggests that the mere presence of both an agent and patient predisposes people to perceive immorality, providing further evidence for a dyadic template that drives moral judgment (Gray et al., 2014; Schein & Gray, 2014).

Links between dyadic contexts and perceived immorality suggest that although the immorality bias applies in obvious situations—suspicious spouses doing overtime with an attractive coworker or shifty neighbors inviting kids over to their house—it might also apply more generally, even when the dyadic situation lacks such clear moral suggestiveness. Just seeing two people talking in a car or standing on street corner could prompt considerations of immorality. This suggests that solitude might be a good strategy for avoiding blame. When people maintain solitude during meditation, prayer, or pilgrimage, they may not just be avoiding sinful thoughts and actions; they may also be engaging in a form of impression management.

Because we are biased to perceive immorality in social situations, the best way to show clear innocence may be isolation.

The present research also reveals an interesting contrast between how we perceive other people and how we perceive their behaviors. Recent work shows that people believe that others' "true selves" are inherently virtuous (Newman, De Freitas, & Knobe, 2015), and people show a general "person-positivity bias" when evaluating human beings (Sears, 1983). Thus, people seem to generally evaluate others in a positive light. However, people also show the immorality bias, jumping to conclusions and evaluating others' behavior in a negative light. These contrasting phenomena suggest that judgments' of others actions and their character (in particular, their underlying essence; Newman et al., 2015) may not be as related as one might intuitively expect. Immoral actions may simply not change people's general perceptions of others' essential character (as considered in the Anne Frank quote "In spite of everything, I still believe that people are really good at heart"). People may truly see the two as unrelated, or they may simply be motivated to see others as being essentially good, even as they do terrible things (as considered in the less-quoted passage that directly follows: "I simply can't build my hopes on a foundation consisting of . . . misery and death").

The present research also suggests that the immorality bias is best understood as a heuristic process. The bias is efficient, intuitive, and likely serves an adaptive purpose, providing ample common ground with many basic judgmental biases in social psychology. More broadly, a heuristics and biases approach to moral judgment helps to bridge the gap between nonmoral and moral judgments and suggests that the same basic processes underlie both. Still, some meaningful differences between moral heuristics and other judgmental heuristics do persist. In particular, for many judgmental heuristics, accuracy is an important factor for evaluating effectiveness. Moral heuristics, however, cannot be evaluated on objective accuracy, because morality is a matter of perception (Schein, Hester, & Gray, 2016): what is right and wrong relies heavily on culture and individual differences (Shweder et al., 1987), and fact checks and mathematical proofs are unlikely to resolve major differences in perception (Goodwin & Darley, 2010). Because of this, defining “moral error” is no easy task and makes it difficult to evaluate moral heuristics on the basis of accuracy—for example, what is an error to a utilitarian may be an obvious truth to a deontologist (Sunstein, 2005).

However, even if moral heuristics do not lead to objective moral errors, they may nevertheless produce judgments that have negative practical implications (Sunstein, 2005). We suggest that the judgments produced by the immorality bias can have troubling outcomes for those who face blame, even after they are proved innocent.

Reducing the Immorality Bias

The immorality bias is likely adaptive in many cases, allowing people to quickly defend themselves and lend aid to others around them. However, when judgments are inaccurate, they may lead to well-intentioned actions that nevertheless harm innocent people. This possibility is particularly disconcerting when the immorality bias may disproportionately lead to inaccurate judgments for certain groups of people.

Ample evidence shows that Black people face more police abuse and false accusations than White people (Allen, 2013; Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006; Lowenstein, 2007). These differences are partly accounted for by people’s tendency to assume that Black people are committing crimes, and then act based on those assumptions (Correll et al., 2002; Goff, Jackson, Di Leone, Culotta, & DiTomasso, 2014). Although these findings are typically attributed to individual differences in perceived threat, differences in immorality bias might also help explain these findings. Moral stereotypes about Black people also include traits such as promiscuity and sexual perversion, which likely lack elements of threat but still possess elements of immorality (Collins, 2002; Devine & Elliot, 1995; Richeson, 2009). Focusing on a general immorality bias, rather than threat, also allows for broader insight into the mechanisms underlying stereotyping.

Once someone makes an immorality bias, the outcomes of that assumption may not be completely reversible. Consideration of the relationship between “knowing” and “believing” something suggests that knowing and believing occur simultaneously, and people have to effortfully “unbelieve” anything that they have learned (Gilbert, 1991). This process of unbelieving is seldom perfect, especially for judgments that we are already biased toward, such as assumptions of wrongdoing. Since immorality is a powerful indicator of character, learning that an accused individual is actually innocent may not fully restore their moral standing. When people sue others for libel or defamation, these lawsuits may be warranted: false accusations, even they are unquestionably false, may permanently damage someone’s perceived moral character. For cases in which the ostensible victim of the false accusation is a clear moral patient (e.g., a child), the negative effects of false accusations might be even worse. Accusations of child abuse, molestation, or neglect might severely tarnish someone’s reputation, even after they are deemed innocent by judge and jury.

Conclusion

One of the most valued principles of the American justice system is the treatment of the defendant as “innocent until proven guilty.” The current research suggests that meeting this ideal is often no easy task, due to a powerful tendency to automatically assume that others’ actions are guilty. These assumptions will sometimes be correct: the car parked outside of the house may well signal a robbery or an affair, and immediate action in response to these possibilities could lead to the best outcomes.

On the other hand, incorrect assumptions can have grave consequences. A man in Texas, responding to shuffling sounds he heard downstairs, assumed that a robber had broken in and opened fire on the suspect—his own wife who was getting something from the kitchen (she survived; Rasta, 2015). In other cases, the mistrust spawned by the immorality bias can fester over time, leading us to whisper lies to ourselves and each other, just like Othello’s Iago. Consider the case of Geraisimov Metaxas, who suspected his wife of an affair on the basis of a single Christmas card from a coworker. Although his suspicions were unfounded—his wife was faithful—he couldn’t escape the bias, and killed his wife’s coworker nine months later (Herszenhorn, 1998).

For better or for worse, humans are strongly attuned to the potential for immoral actions. The mere suggestion of immorality leads to the assumption of its presence, and only by exertion of will are we able to replace suspicion with trust.

Table 1. **Scenarios used in Experiment 1**

While he pulls down his pants, a man thinks about his boys’ soccer team.

A young woman is walking down the street at night. A man glances at her and reaches into his pocket.

A man gives a woman a drug so she is not aware of what he is doing to her body.

While a high school student takes a shower, he thinks of his younger sister.

A woman leaves work early to meet a man who is not her husband.

Table 2. **List of nonsense actions by source**

Verbs	Source
plurded, zorked, ruped, plaked, zoshed, blofed, rooged, yoded, hooled, sorned, weked, leamed, glotted	(Oetting, 1999)
stoffed, cugged, trabbed, crogged, vasked, bropped, satched, grushed, plammed, scurred, spuffed, dotched	(Thomas et al., 2001)
mooked, tived, kalled, geeped, voozed, mipped, zecked, dassed, fimed, bozed	(Van der Lely, 1994)
biffed, ziked, blicked, dacked, moked	(Fisher, Hall, Rakowitz, & Gleitman, 1994)
doaked, gumped, floosed, gomped, japed	(Pinker, Lebeaux, & Frost, 1987)
keefed, pudded, chammed, mibbed, koobed	(Olguin & Tomasello, 1993)
karded, semmed, larped, wugged, toped	(Waxman, Lidz, Braun, & Lavin, 2009)
splinged, prassed, crived, prussed, lecked	(Van der Lely & Ullman, 1996)
pelled, norped, mooped, keated	(Gropen, Pinker, Hollander, Goldberg, & Wilson, 1989)
stiped, braffed, pilked, gished	(Fisher, 1996)
tammed, gorped, goped	(Tomasello, 2000)
glorped, freped	(Roseberry, Hirsh-Pasek, Parish-Morris, & Golinkoff, 2009)
baffed	(Abbot-Smith, Lieven, & Tomasello, 2001)
daxed	(Tomasello & Barton, 1994)
hirshed	(Maguire, Hirsh-Pasek, Golinkoff, & Brandone, 2008)

Table 3. **Sentences categorized by participants in Experiment 6**

Probably Immoral	Possibly Immoral
He lied to his father. (71)	He stared at his daughter. (10)
He picked the car’s lock. (56)	He threw the axe. (19)
He grabbed her neck. (58)	He slapped her butt. (25)
She broke his leg. (64)	He picked up the child and ran. (24)
He knocked over the man. (61)	He snuck into the house. (36)
He broke into the house. (89)	She logged on to his Facebook. (27)
She didn’t pay for her meal. (76)	He took the child to the bathroom. (5)
He gave the drug to the child. (57)	He swung the baseball bat. (5)
He punched the man. (55)	He grabbed the knife. (21)
She slipped the jewelry in her purse. (74)	He kicked down the door. (38)
She lied to her brother. (88)	He thought about his sister. (19)
She bit his neck. (40)	She undressed the child. (16)
He sedated the woman. (48)	He fired a gun. (25)
She kicked him in the shin. (73)	He picked up the money. (10)

Note. Numbers in parentheses indicate the percentage of people who categorized the sentence as “Immoral” during pilot tests.

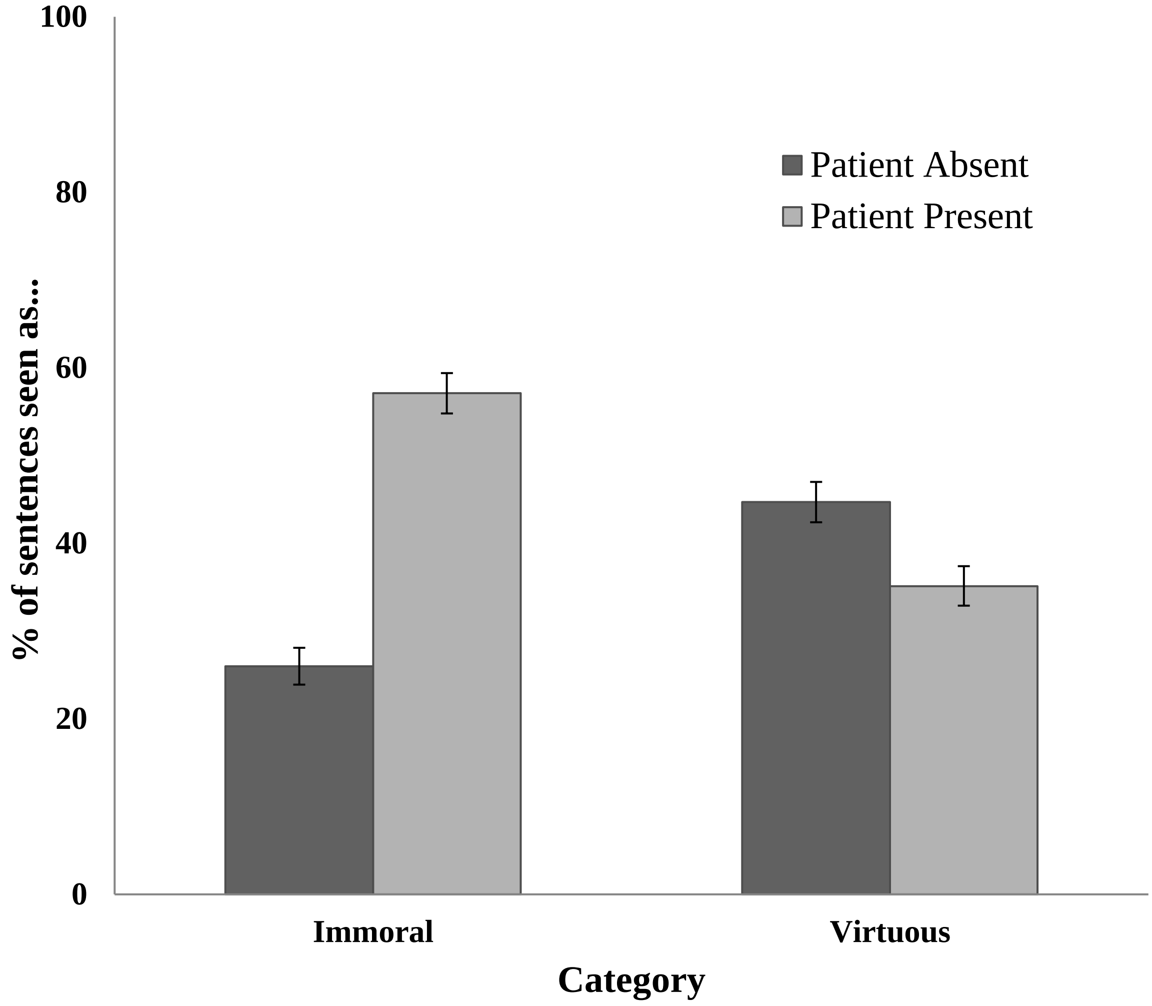

Figure 1. **Percentage of "Immoral" or "Virtuous" Responses**

Percentage of "Immoral" or "Virtuous" responses when the Patient is Absent or Present from the sentence. Bars represent 95% confidence intervals.

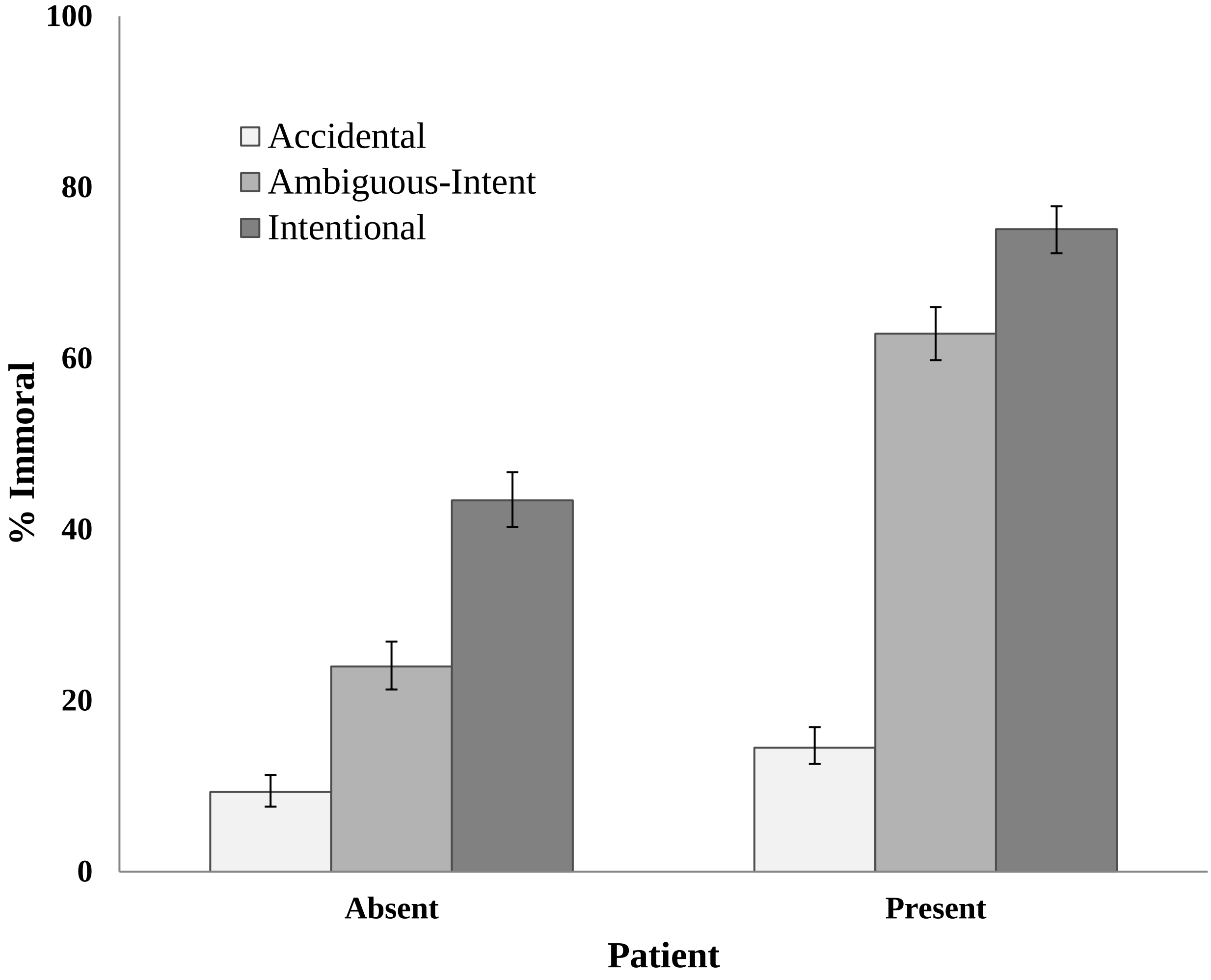

Figure 2. **Percentage of "Immoral" Responses by Intent and Patient.**

Percentage of "Immoral" responses by Intent and Patient. Bars represent 95% confidence intervals.

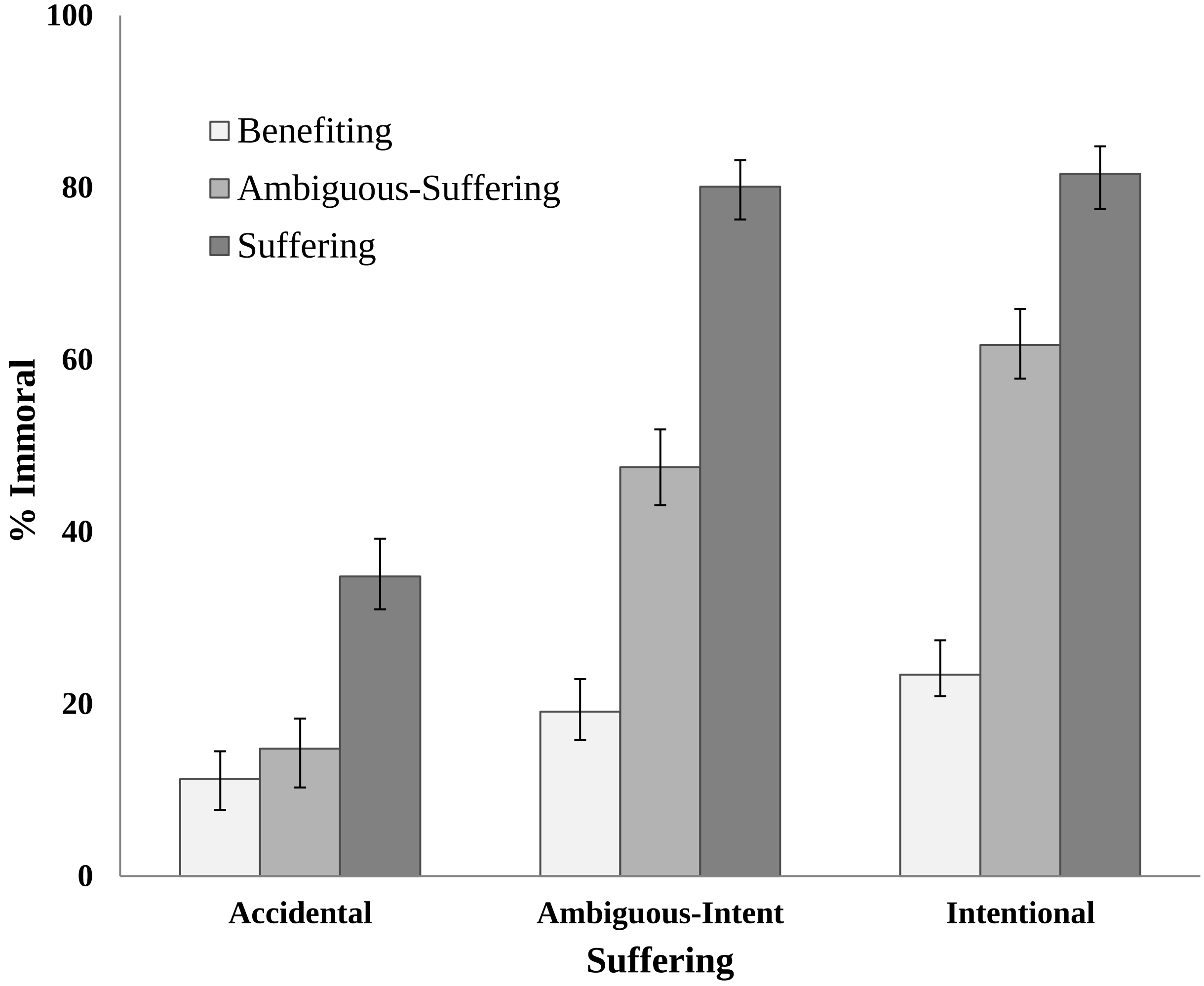

Figure 3. **Percentage of "Immoral" Responses by Intent and Suffering**

Percentage of "Immoral" Responses by Intent and Suffering. Bars represent 95% confidence intervals.

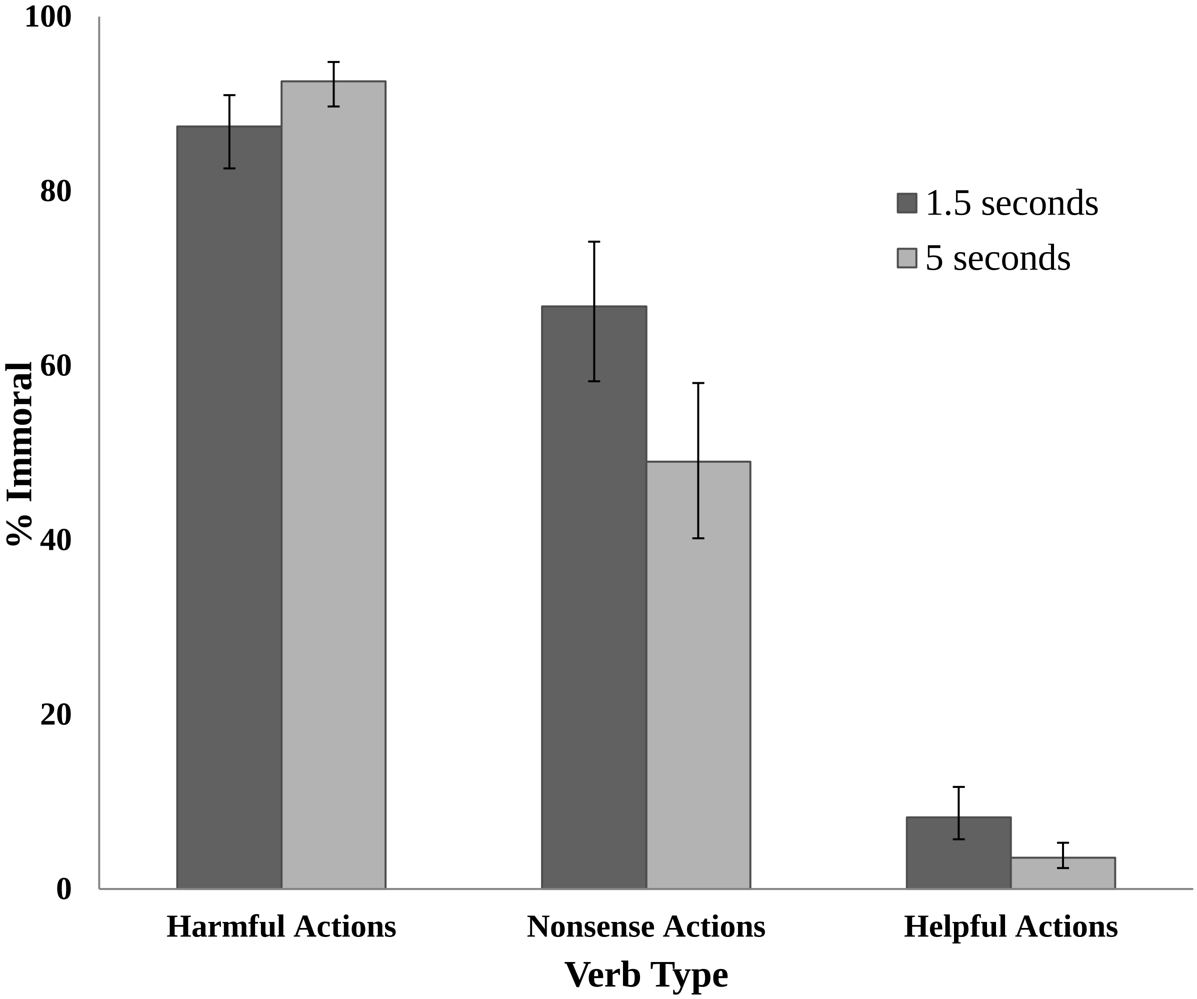

Figure 4. **Percentage of “Immoral” Responses by Time and Verb Type**

Percentage of “Immoral” responses by Time and Verb Type. Bars represent 95% confidence intervals.

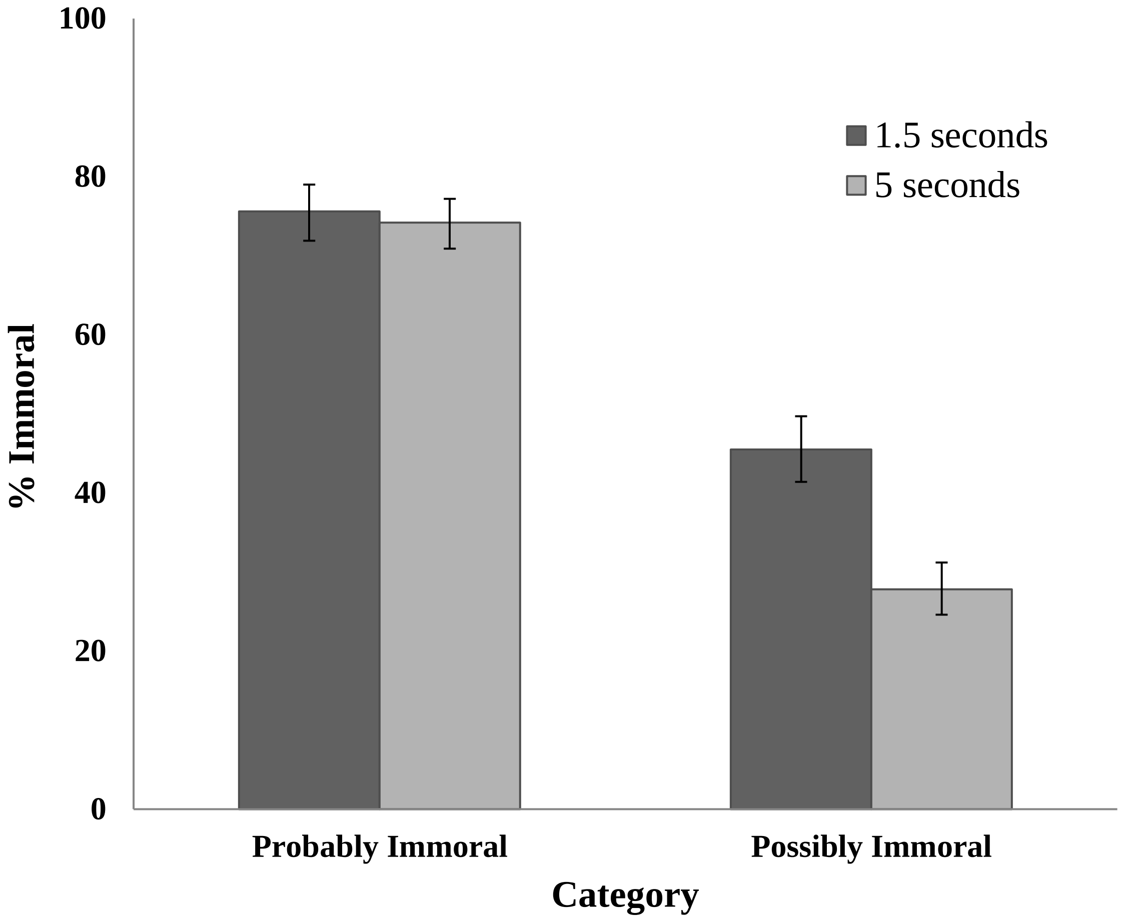

Figure 5. **Percentage of "Immoral" Responses by Time and Category**

Percentage of "Immoral" responses by Time and Category. Bars represent 95% confidence intervals.

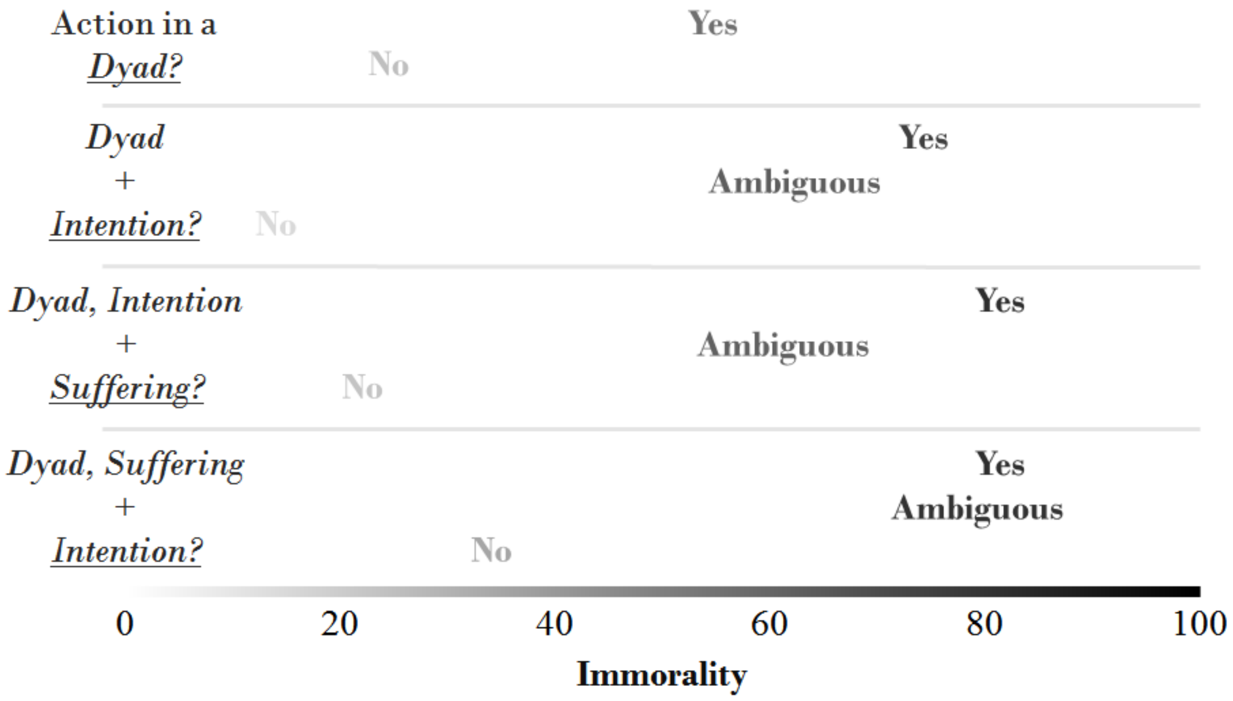

Figure 6. **Summary of Results for Experiments 2 through 5**

Summary of participants' percentage of "Immoral" responses across Experiments 2 through 5 as elements of the dyad, intention, and suffering were systematically added and manipulated. Labels in each row correspond with manipulations of the underlined element. Centered position and shading correspond with the percentage of Immoral responses.

REFERENCES

Abbot-Smith, K., Lieven, E., & Tomasello, M. (2001). What preschool children do and do not do with ungrammatical word orders. *Cognitive Development*, *16*(2), 679–692.

Allen, F. (2013, May 14). Blacks are still majority of the wrongfully convicted. *Black Voice News*. Retrieved from http://www.blackvoicenews.com/news/news-wire/48803-blacks-are-still-majority-of-the-wrongfully-convicted.html

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. http://doi.org/10.1016/j.jml.2007.12.005

Baguley, T. (2012). Calculating and graphing within-subject confidence intervals for ANOVA. *Behavior Research Methods*, *44*(1), 158–175. http://doi.org/10.3758/s13428-011-0123-7

Barrett, H. C., & Behne, T. (2005). Children's understanding of death as the cessation of agency: a test using sleep versus death. *Cognition*, *96*(2), 93–108. http://doi.org/16/j.cognition.2004.05.004

Bastian, B., Loughnan, S., Haslam, N., & Radke, H. R. M. (2012). Don't mind meat? The denial of mind to animals used for human consumption. *Personality and Social Psychology Bulletin*, *38*(2), 247–256. http://doi.org/10.1177/0146167211424291

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. http://doi.org/10.1037//1089-2680.5.4.323

Bergström, L. S., & Lynöe, N. (2008). Enhancing concentration, mood and memory in healthy individuals: An empirical study of attitudes among general practitioners and the general population. *Scandinavian Journal of Public Health*, *36*(5), 532–537. http://doi.org/10.1177/1403494807087558

Blanchette, I. (2006). Snakes, spiders, guns, and syringes: How specific are evolutionary constraints on the detection of threatening stimuli? *The Quarterly Journal of Experimental Psychology*, *59*(8), 1484–1504. http://doi.org/10.1080/02724980543000204

Bodenhausen, G. V. (1990). Stereotypes as judgmental heuristics: Evidence of circadian variations in discrimination. *Psychological Science*, *1*(5), 319–322.

Bohner, G. (2001). Writing about rape: Use of the passive voice and other distancing text features as an expression of perceived responsibility of the victim. *British Journal of Social Psychology*, *40*(4), 515–529.

Bosson, J. K., Johnson, A. B., Niederhoffer, K., & Swann, W. B. (2006). Interpersonal chemistry through negativity: Bonding by sharing negative attitudes about others. *Personal Relationships*, *13*(2), 135–150.

Bottoms, B. L., Goodman, G. S., Schwartz-Kenney, B. M., & Thomas, S. N. (2002). Understanding children's use of secrecy in the context of eyewitness reports. *Law and Human Behavior*, *26*(3), 285–313.

Bowles, S. (2006). Group competition, reproductive leveling, and the evolution of human altruism. *Science*, *314*(5805), 1569–1572. http://doi.org/10.1126/science.1134829

Bradley, G. W. (1978). Self-serving biases in the attribution process: A reexamination of the fact or fiction question. *Journal of Personality and Social Psychology*, *36*(1), 56–71.

Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology*, *41*(2), 135–143. http://doi.org/10.1002/ejsp.744

Brambilla, M., Sacchi, S., Pagliaro, S., & Ellemers, N. (2013). Morality and intergroup relations: Threats to safety and group image predict the desire to interact with outgroup and ingroup members. *Journal of Experimental Social Psychology*, *49*(5), 811–821. http://doi.org/10.1016/j.jesp.2013.04.005

Brambilla, M., Sacchi, S., Rusconi, P., Cherubini, P., & Yzerbyt, V. Y. (2012). You want to give a good impression? Be honest! Moral traits dominate group impression formation. *British Journal of Social Psychology*, *51*(1), 149–166.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3–5. http://doi.org/10.1177/1745691610393980

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, *19*(1–3), 7–42.

Cameron, C. D., Lindquist, K. A., & Gray, K. (2015). A constructionist review of morality and emotions: No evidence for specific links between moral content and discrete emotions. *Personality and Social Psychology Review*, 371–394. http://doi.org/10.1177/1088868314566683

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, *39*(5), 752–766.

Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. Guilford Press.

Cheng, J. S., Ottati, V. C., & Price, E. D. (2013). The arousal model of moral condemnation. *Journal of Experimental Social Psychology*, *49*(6), 1012–1018. http://doi.org/10.1016/j.jesp.2013.06.006

Cohen, T. R., Montoya, R. M., & Insko, C. A. (2006). Group morality and intergroup relations: Cross-cultural and experimental evidence. *Personality and Social Psychology Bulletin, 32*(11), 1559–1572. http://doi.org/10.1177/0146167206291673

Collins, P. H. (2002). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment*. Routledge.

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology, 83*(6), 1314–1329. http://doi.org/10.1037//0022-3514.83.6.1314

Coy, J. S., Lambert, J. E., & Miller, M. M. (2016). Stories of the accused: A phenomenological inquiry of MFTs and accusations of unprofessional conduct. *Journal of Marital and Family Therapy, 42*(1), 139–152. http://doi.org/10.1111/jmft.12109

Critcher, C. R., Inbar, Y., & Pizarro, D. (2012). How quick decisions illuminate moral character. *Social Psychological and Personality Science, 4*(3), 308–315. http://doi.org/10.1177/1948550612457688

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science, 17*(12), 1082–1089. http://doi.org/10.1111/j.1467-9280.2006.01834.x

de Waal, F. B. M. (2008). Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology, 59*(1), 279–300. http://doi.org/10.1146/annurev.psych.59.103006.093625

DeScioli, P., Gilbert, S., & Kurzban, R. (2012). Indelible victims and persistent punishers in moral cognition. *Psychological Inquiry, 23*(2), 143–149.

DeScioli, P., & Kurzban, R. (2009). Mysteries of morality. *Cognition*, *112*(2), 281–299. http://doi.org/10.1016/j.cognition.2009.05.008

DeScioli, P., & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological Bulletin*, *139*(2), 477–496. http://doi.org/10.1037/a0029065

Devine, P. G., & Elliot, A. J. (1995). Are racial stereotypes really fading? The Princeton trilogy revisited. *Personality and Social Psychology Bulletin*, *21*, 1139–1150.

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of black defendants predicts capital-sentencing outcomes. *Psychological Science*, *17*(5), 383–386. http://doi.org/10.1111/j.1467-9280.2006.01716.x

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3–4), 169–200. http://doi.org/10.1080/02699939208411068

Epley, N., Waytz, A., Akalis, S., & Cacioppo, J. T. (2008). When we need a human: Motivational determinants of anthropomorphism. *Social Cognition*, *26*(2), 143–155. http://doi.org/10.1521/soco.2008.26.2.143

Erwin, P. G. (2006). Children's evaluative stereotypes of masculine, feminine, and androgynous first names. *The Psychological Record*, *56*(4), 513–519.

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making*, *13*(1), 1.

Firestone, C., & Scholl, B. J. (2015). Enhanced visual awareness for morality and pajamas? Perception vs. memory in "top-down" effects. *Cognition*, *136*, 409–416. http://doi.org/10.1016/j.cognition.2014.10.014

Firestone, C., & Scholl, B. J. (2016). "Moral perception" reflects neither morality nor perception. *Trends in Cognitive Sciences*, *20*(2), 75–76. http://doi.org/10.1016/j.tics.2015.10.006

Fisher, C. (1996). Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive Psychology*, *31*(1), 41–81.

Fisher, C., Hall, D. G., Rakowitz, S., & Gleitman, L. (1994). When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua*, *92*, 333–375.

Fortune, J. L., & Newby-Clark, I. R. (2008). My friend is embarrassing me: Exploring the guilty by association effect. *Journal of Personality and Social Psychology*, *95*(6), 1440–1449. http://doi.org/10.1037/a0012627

Frazer, A. K., & Miller, M. D. (2008). Double standards in sentence structure: Passive voice in narratives describing domestic violence. *Journal of Language and Social Psychology*, *28*(1), 62–71. http://doi.org/10.1177/0261927X08325883

Gantman, A. P., & Van Bavel, J. J. (2014). The moral pop-out effect: Enhanced perceptual awareness of morally relevant stimuli. *Cognition*, *132*(1), 22–29. http://doi.org/10.1016/j.cognition.2014.02.007

Gantman, A. P., & Van Bavel, J. J. (2016). See for yourself: Perception is attuned to morality. *Trends in Cognitive Sciences*, *20*(2), 76–77.

Gigerenzer, G., & Brighton, H. (2009). Homo Heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, *1*(1), 107–143. http://doi.org/10.1111/j.1756-8765.2008.01006.x

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650–669.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, *46*(2), 107–119. http://doi.org/10.1037/0003-066X.46.2.107

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment* (1st ed.). Cambridge, UK: Cambridge University Press.

Goff, P. A., Jackson, M. C., Di Leone, B. A. L., Culotta, C. M., & DiTomasso, N. A. (2014). The essence of innocence: Consequences of dehumanizing Black children. *Journal of Personality and Social Psychology*, *106*(4), 526–545. http://doi.org/10.1037/a0035663

Goldinger, S. D., Kleider, H. M., Azuma, T., & Beike, D. R. (2003). "Blaming the victim" under memory load. *Psychological Science, 14*(1), 81–85. http://doi.org/10.1111/1467-9280.01423

Goodman, J. K., Cryder, C. E., & Cheema, A. (2012). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/bdm.1753/full

Goodwin, G. P. (2015). Moral character in person perception. *Current Directions in Psychological Science*, *24*(1), 38–44.

Goodwin, G. P., & Darley, J. M. (2010). The perceived objectivity of ethical beliefs: Psychological findings and implications for public policy. *Review of Philosophy and Psychology, 1*(2), 161–188.

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*(1), 148–168. http://doi.org/10.1037/a0034726

Graham, J. (2015). Explaining away differences in moral judgment: Comment on Gray & Keeney (2015). *Social Psychological and Personality Science*.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S., & Ditto, P. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, *47*, 55–130. http://doi.org/10.1016/B978-0-12-407236-7.00002-4

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*(2), 366–385. http://doi.org/10.1037/a0021847

Gray, K. (2012). The power of good intentions: Perceived benevolence soothes pain, increases pleasure, and improves taste. *Social Psychological and Personality Science, 3*, 639–645.

Gray, K., & Keeney, J. E. (2015a). Disconfirming moral foundations theory on its own terms: Reply to Graham (2015). *Social Psychological and Personality Science*, 1–4.

Gray, K., & Keeney, J. E. (2015b). Impure, or just weird? Scenario sampling bias raises questions about the foundation of moral cognition. *Social Psychological and Personality Science*, 1–10.

Gray, K., & Schein, C. (in press). No absolutism here: Harm predicts moral judgment 30x better than disgust—Commentary on Scott, Inbar & Rozin (2015). *Perspectives on Psychological Science*.

Gray, K., Schein, C., & Ward, A. F. (2014). The myth of harmless wrongs in moral cognition: Automatic dyadic completion from sin to suffering. *Journal of Experimental Psychology: General*, *143*(4), 1600–1615. http://doi.org/10.1037/a0036149

Gray, K., & Wegner, D. M. (2010). Blaming God for our pain: Human suffering and the divine mind. *Personality and Social Psychology Review*, *14*(1), 7–16. http://doi.org/10.1177/1088868309350299

Gray, K., & Wegner, D. M. (2011). Dimensions of moral emotions. *Emotion Review*, *3*(3), 227–229.

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, *23*, 101–124. http://doi.org/10.1080/1047840x.2012.651387

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*(3), 1144–1154.

Greenwald, A. G., Mcghee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.

Gropen, J., Pinker, S., Hollander, M., Goldberg, R., & Wilson, R. (1989). The learnability and acquisition of the dative alternation in English. *Language*, *65*(2), 203–257. http://doi.org/10.2307/415332

Hahn, S., Carlson, C., Singer, S., & Gronlund, S. D. (2006). Aging and visual search: Automatic and controlled attentional bias to threat faces. *Acta Psychologica*, *123*(3), 312–336. http://doi.org/10.1016/j.actpsy.2006.01.008

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814–834.

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York, NY: Pantheon Books.

Haidt, J. (2013). *The righteous mind: Why good people are divided by politics and religion*. Random House LLC.

Hall, R. C., & Hall, R. C. (2001). False allegations: the role of the forensic psychiatrist. *Journal of Psychiatric Practice®*, *7*(5), 343–346.

Hamilton, W. D. (1963). The evolution of altruistic behavior. *American Naturalist*, 354–356.

Hamlin, J. K., & Baron, A. S. (2014). Agency attribution in infancy: Evidence for a negativity bias. *PLOS ONE, 9*(5), e96112. http://doi.org/10.1371/journal.pone.0096112

Helzer, E. G., & Pizarro, D. (2011). Dirty liberals! Reminders of physical cleanliness influence moral and political attitudes. *Psychological Science*, *22*(4), 517–522.

Henley, N. M., Miller, M., & Beazley, J. A. (1995). Syntax, semantics, and sexual violence: Agency and the passive voice. *Journal of Language and Social Psychology, 14*(1–2), 60–84. http://doi.org/10.1177/0261927X95141004

Herszenhorn, D. (1998, February 14). Teacher's killer suspected an affair that never happened. *New York Times*. Retrieved from http://www.nytimes.com/1998/02/14/nyregion/teacher-s-killer-suspected-an-affair-that-never-happened.html

Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science*, *345*(6202), 1340–1343.

Hyland, F. (2001). Dealing with plagiarism when giving feedback. *ELT Journal*, *55*(4), 375–381.

Jacobson, D. (2012). Moral dumbfounding and moral stupefaction. *Oxford Studies in Normative Ethics*, *2*.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and Biases: The Psychology of Intuitive Judgment*, *49*. Retrieved from https://www.researchgate.net/profile/Shane_Frederick/publication/229071271_Representativeness_revisited_Attribute_substitution_in_intuitive_judgment/links/54087a8c0cf2c48563bd6c75.pdf

Krueger, C. (2004). A comparison of the general linear mixed model and repeated measures ANOVA using a dataset with multiple missing data points. *Biological Research For Nursing*, *6*(2), 151–157. http://doi.org/10.1177/1099800404267682

Latané, B., & Darley, J. M. (1968). Group inhibition of bystander intervention in emergencies. *Journal of Personality and Social Psychology*, *10*(3), 215–221.

Lewis, K., Gray, K., & Meierhenrich, J. (2014). The structure of online activism. *Sociological Science*, 1–9. http://doi.org/10.15195/v1.a1

LoBue, V. (2010). What's so scary about needles and knives? Examining the role of experience in threat detection. *Cognition & Emotion*, *24*(1), 180–187. http://doi.org/10.1080/02699930802542308

LoBue, V., & DeLoache, J. S. (2010). Superior detection of threat-relevant stimuli in infancy: Threat detection in infancy. *Developmental Science*, *13*(1), 221–228. http://doi.org/10.1111/j.1467-7687.2009.00872.x

Lowenstein, J. (2007, November 4). Killed by the cops. *ColorLines*. Retrieved from http://www.colorlines.com/articles/killed-cops

MacGregor-Fors, I., & Payton, M. E. (2013). Contrasting diversity values: Statistical inferences based on overlapping confidence intervals. *PLoS ONE*, *8*(2), e56794. http://doi.org/10.1371/journal.pone.0056794

Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., & Brandone, A. C. (2008). Focusing on the relation: fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science*, *11*(4), 628–634. http://doi.org/10.1111/j.1467-7687.2008.00707.x

Malle, B. F. (2006). Intentionality, morality, and their relationship in human judgment. *Journal of Cognition and Culture*, *6*(1–2), 87–112.

Maner, J. K., Gailliot, M. T., & DeWall, C. N. (2007). Adaptive attentional attunement: evidence for mating-related perceptual bias. *Evolution and Human Behavior*, *28*(1), 28–36. http://doi.org/10.1016/j.evolhumbehav.2006.05.006

McLeod, P., Reed, N., & Dienes, Z. (2003). Psychophysics: How fileders arrive in time to catch the ball. *Nature*, *426*, 224–245. http://doi.org/10.1038/426244a

Melson, G. F., Kahn, P. H., Beck, A., Friedman, B., Roberts, T., Garrett, E., & Gill, B. T. (2009). Children's behavior toward and understanding of robotic and living dogs. *Journal of Applied Developmental Psychology*, *30*(2), 92–102. http://doi.org/10.1016/j.appdev.2008.10.011

Menon, G., & Raghubir, P. (2003). Ease-of-retrieval as an automatic input in judgments: a mere-accessibility framework? *Journal of Consumer Research*, *30*(2), 230–243.

Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological Bulletin*, *130*(5), 711–747. http://doi.org/10.1037/0033-2909.130.5.711

Morewedge, C. K. (2009). Negativity bias in attribution of external agency. *Journal of Experimental Psychology: General*, *138*(4), 535–545. http://doi.org/10.1037/a0016796

Mullen, B., Atkins, J. L., Champion, D. S., Edwards, C., Hardy, D., Story, J. E., & Vanderklok, M. (1985). The false consensus effect: A meta-analysis of 115 hypothesis tests. *Journal of Experimental Social Psychology*, *21*(3), 262–283.

Newman, G. E., De Freitas, J., & Knobe, J. (2015). Beliefs about the true self explain asymmetries based on moral judgment. *Cognitive Science*, *39*(1), 96–125.

Nora, W. L. Y., & Zhang, K. C. (2010). Motives of cheating among secondary students: The role of self-efficacy and peer influence. *Asia Pacific Education Review, 11*(4), 573–584.

Oetting, J. B. (1999). Children with SLI use argument structure cues to learn verbs. *Journal of Speech, Language, and Hearing Research*, *42*(5), 1261–1274.

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General*, *130*(3), 466–478.

Olguin, R., & Tomasello, M. (1993). Twenty-five-month-old children do not have a grammatical category of verb. *Cognitive Development*, *8*(3), 245–272.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*(2), 181–192.

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*(3), 277–293.

Payton, M. E., Greenstone, M. H., & Schenker, N. (2003). Overlapping confidence intervals or standard error intervals: what do they mean in terms of statistical significance? *Journal of Insect Science*, *3*(1), 1–6.

Pinker, S., Lebeaux, D. S., & Frost, L. A. (1987). Productivity and constraints in the acquisition of the passive. *Cognition*, *26*(3), 195–267.

Pizarro, D., & 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). Washington, DC: APA Press.

Pizarro, D., Uhlmann, E., & Salovey, P. (2003). Asymmetry in judgments of moral blame and praise: The role of perceived metadesires. *Psychological Science*, *14*, 267–272. http://doi.org/10.1111/1467-9280.03433

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. http://doi.org/10.1037/a0021867

Rasta, A. (2015, April 24). Husband accidentally shoots wife after believing intruder was inside home. *Click2Houston*. Retrieved from http://www.click2houston.com/news/hcso-husband-accidentally-shoots-wife-after-thinking-intruder-was-inside-home/32545450

Reuben, E., & Stephenson, M. (2013). Nobody likes a rat: On the willingness to report lies and the consequences thereof. *Journal of Economic Behavior & Organization*, *93*, 384–391.

Richeson, M. P. (2009). Sex, drugs, and... race-to-castrate: A Black box warning of chemical castration's potential racial side effects. *Harv. BlackLetter LJ*, *25*, 95–131.

Roseberry, S., Hirsh-Pasek, K., Parish-Morris, J., & Golinkoff, R. M. (2009). Live action: Can young children learn verbs from video? *Child Development*, *80*(5), 1360–1375.

Ross, L., Greene, D., & House, P. (1977). The “false consensus effect”: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, *13*(3), 279–301.

Rosset, E. (2008). It’s no accident: Our bias for intentional explanations. *Cognition*, *108*(3), 771–780. http://doi.org/10.1016/j.cognition.2008.07.001

Royzman, E., Kim, K., & Leeman, R. F. (2015). The curious tale of Julie and Mark: Unraveling the moral dumbfounding effect. *Judgment and Decision Making*, *10*(4), 296–313.

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology*, *76*(4), 574–586. http://doi.org/10.1037/0022-3514.76.4.574

Schein, C., Goranson, A., & Gray, K. (2015). The uncensored truth about morality. *The Psychologist*, *28*(12), 982–985.

Schein, C., & Gray, K. (2014). The prototype model of blame: Freeing moral cognition from linearity and little boxes. *Psychological Inquiry*, *25*(2), 236–240. http://doi.org/10.1080/1047840X.2014.901903

Schein, C., & Gray, K. (2015). The unifying moral dyad: Liberals and conservatives share the same harm-based moral template. *Personality and Social Psychology Bulletin*, *41*(8), 1147–1163. http://doi.org/10.1177/0146167215591501

Schein, C., Hester, N., & Gray, K. (2016). The visual guide to morality: Vision as an integrative analogy for moral experience, variability and mechanism. *Social and Personality Psychology Compass*, *10*(4), 231–251. http://doi.org/10.1111/spc3.12247

Schein, C., Ritter, R., & Gray, K. (in press). Harm mediates the disgust-immorality link. *Emotion*.

Scheske, C., & Schnall, S. (2012). The ethics of “smart drugs”: Moral judgments about healthy people’s use of cognitive-enhancing drugs. *Basic and Applied Social Psychology*, *34*(6), 508–515.

Schroeder, D. A., Penner, L. A., Dovidio, J. F., & Piliavin, J. A. (1995). *The psychology of helping and altruism: Problems and puzzles.* McGraw-Hill.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, *45*(3), 513–523. http://doi.org/10.1037/0022-3514.45.3.513

Sears, D. O. (1983). The person-positivity bias. *Journal of Personality and Social Psychology, 44*(2), 233–250.

Sherif, M. (1961). *Intergroup conflict and cooperation: the robbers cave experiment*. Norman, OK: University Book Exchange.

Shweder, R. A., Mahapatra, M., & Miller, J. (1987). Culture and moral development. In J. Kagan & S. Lamb (Eds.), *The Emergence of Morality in Young Children* (pp. 1–83). Chicago, IL: University of Chicago Press.

Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The "big three" of morality (autonomy, community, and divinity), and the "big three" explanations of suffering. In *Morality and Health* (pp. 119–169). New York, NY: Routledge.

Silver, N., & McCann, A. (2014). How to tell someone's age when all you know is her name. Retrieved from http://fivethirtyeight.com/features/how-to-tell-someones-age-when-all-you-know-is-her-name/

Skitka, L. J., & Bauman, C. W. (2008). Moral conviction and political engagement. *Political Psychology*, *29*(1), 29–54. http://doi.org/10.1111/j.1467-9221.2007.00611.x

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European Journal of Operational Research*, *177*(3), 1333–1352. http://doi.org/10.1016/j.ejor.2005.04.006

Social Security Administration. (2016). *Top Names over the Last 100 Years*. Retrieved from https://www.ssa.gov/oact/babynames/decades/century.html

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662.

Sunstein, C. R. (2005). Moral heuristics. *The Behavioral and Brain Sciences*, *28*(4), 531–542. http://doi.org/10.1017/S0140525X05000099

Tetlock, P. E., Kristel, O. V., Beth, S., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, *78*(5), 853–870. http://doi.org/10.1037/0022-3514.78.5.853

Thomas, M. S. C., Grant, J., Barham, Z., Gsödl, M., Laing, E., Lakusta, L., … Karmiloff-Smith, A. (2001). Past tense formation in Williams syndrome. *Language and Cognitive Processes*, *16*(2–3), 143–176. http://doi.org/10.1080/01690960042000021

Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences*, *4*(4), 156–163.

Tomasello, M., & Barton, M. E. (1994). Learning words in nonostensive contexts. *Developmental Psychology*, *30*(5), 639–650.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*(7), 1275–1289. http://doi.org/10.3758/s13421-011-0104-1

Trevino, L. K., & Victor, B. (1992). Peer reporting of unethical behavior: A social context perspective. *Academy of Management Journal*, *35*(1), 38–64. http://doi.org/10.2307/256472

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, *46*(1), 35–57.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458. http://doi.org/10.1126/science.7455683

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293–315.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, *106*(4), 1039–1061. http://doi.org/10.2307/2937956

Van Bavel, J. J., Packer, D. J., Haas, I. J., & Cunningham, W. A. (2012). The importance of moral construal: Moral versus non-moral construal elicits faster, more extreme, universal evaluations of the same actions. *PLoS ONE*, *7*(11), e48693. http://doi.org/10.1371/journal.pone.0048693

Van Berkum, J. J. A., Holleman, B., Nieuwland, M., Otten, M., & Murre, J. (2009). Right or wrong? The brain's fast response to morally objectionable statements. *Psychological Science*, *20*(9), 1092–1099. http://doi.org/10.1111/j.1467-9280.2009.02411.x

Van der Lely, H. K. (1994). Canonical linking rules: Forward versus reverse linking in normally developing and specifically language-impaired children. *Cognition*, *51*(1), 29–72.

Van der Lely, H. K., & Ullman, M. (1996). The computation and representation of past-tense morphology in specifically language impaired and normally developing children. In *Proceedings of the 20th annual Boston University Conference on language development* (pp. 804–815).

Viscusi, W. K. (2000). Corporate risk analysis: A reckless act? *Stanford Law Review*, *52*(3), 547–597. http://doi.org/10.2307/1229473

Walther, E. (2002). Guilty by mere association: Evaluative conditioning and the spreading attitude effect. *Journal of Personality and Social Psychology*, *82*(6), 919–934. http://doi.org/10.1037//0022-3514.82.6.919

Waxman, S. R., Lidz, J. L., Braun, I. E., & Lavin, T. (2009). Twenty four-month-old infants' interpretations of novel verbs and nouns in dynamic scenes. *Cognitive Psychology*, *59*(1), 67–95. http://doi.org/10.1016/j.cogpsych.2009.02.001

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, 113–117.

Waytz, A., & Young, L. (2012). The group-member mind tradeoff: Attributing mind to groups versus group members. *Psychological Science*, *23*, 77–85.

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, *25*(9), 1115–1125.

Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science*, *16*(10), 780–784. http://doi.org/10.1111/j.1467-9280.2005.01614.x

Whitman, J. L., & Davis, R. C. (2007). *Snitches get stitches: Youth, gangs, and witness intimidation in Massachusetts* (pp. 1–81). Retrieved from http://masslib-dspace.longsight.com/handle/2452/38544

Wilson, J. Q. (1997). *The Moral Sense*. Simon and Schuster.

Wright, J. C., & Baril, G. (2011). The role of cognitive resources in determining our moral intuitions: Are we all liberals at heart? *Journal of Experimental Social Psychology*, *47*, 1007–1012. http://doi.org/10.1016/j.jesp.2011.03.014

Ybarra, O., Chan, E., & Park, D. (2001). Young and old adults' concerns about morality and competence. *Motivation and Emotion*, *25*(2), 85–100.