UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

What Others Want and How We Evaluate Them

Permalink

https://escholarship.org/uc/item/4ct74133

Author

Jago, Carl

Publication Date

2020

Supplemental Material

https://escholarship.org/uc/item/4ct74133#supplemental

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA SAN DIEGO

What Others Want and How We Evaluate Them

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Experimental Psychology

by

Carl Pierre Jago

Committee in charge:

Professor Karen Dobkins, Chair
Professor Gail Heyman
Professor Dana Nelkin
Professor Christopher Oveis
Professor Piotr Winkielman

2020

Copyright

Carl Pierre Jago, 2020

All rights reserved.

The dissertation of Carl Pierre Jago is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California San Diego

2020

DEDICATION

To Christina and Cali.

TABLE OF CONTENTS

LIST OF SUPPLEMENTAL FILES

File 1: Chapter 4: Supplementary Materials, Chapteter4-SupplementaryMaterials.xlsx

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Karen Dobkins for her support as the chair of my committee. Her encouragement, positivity, and endless insight has proved to be invaluable. For me, our collaboration on Chapter 4 was a breath of fresh air.

I would like to extent my gratitude to the other members of my committee: Professor Gail Heyman, Professor Dana Nelkin, Professor Chris Oveis, and Professor Piotr Winkielman. Each one has provided valuable feedback and contributed uniquely to the development of my research.

I would also like to thank Nicholas Christenfeld for playing the key role in my early development as a graduate student. In addition to our collaboration on Chapter 3, he offered important insights and ideas when I began working on Chapters 1 and 2. Without him this dissertation would not be what it is.

Chapter 3, in full, is currently being prepared for submission for publication of the material. Jago, Carl P.; Ryazanov, Arseny A.; Christenfeld, Nicholas J. S. The dissertation author was the primary investigator and author of this chapter.

Chapter 4, in full, has been submitted for publication of the material as it may appear in Journal of Social and Personal Relationships, 2020, Jago, Carl P.; Dobkins, Karen R., Sage Publications, 2020. The dissertation author was the primary investigator and author of this paper.

VITA

2011-2013	Research Assistant, Columbia University
2013	Bachelor of Science, Columbia University
2014	Master of Arts, University of California San Diego
2013-2019	Research Assistant, University of California San Diego
2013-2019	Teaching Assistant, University of California San Diego
2020	Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Jago, C. P., Ryazanov, A. A., Christenfeld, N. J. S. (in prep). Intuitive Blaming of Peeping Toms Versus Cyber Peepers?

Jago, C. P., (in prep). Do Behavioral Base Rates Influence Associated Moral Judgments?

Jago, C. P., (in prep). Is the Side-Effect Effect a Function of Moral Judgment?

Jago, C. P., Dobkins, K. R., (under review). What women say they want versus what men imagine they do.

Stevens, S. M., Jago, C. P., Jasko, K., Heyman, G. D., (under review). Ideologically-driven selective empathy.

Cornwell, J. F., Jago, C. P., & Higgins, E. T. (2019). When group influence is more or less likely: the case of moral judgments. *Basic and Applied Social Psychology*, *41*(6), 386-395.

Jago, C. P (2012). Can science tell right from wrong? *Columbia Science Review*, *9*(1), 24-26.

ABSTRACT OF THE DISSERTATION

What Others Want and How We Evaluate Them

by

Carl Pierre Jago

Doctor of Philosophy in Experimental Psychology

University of California San Diego, 2020

Professor Karen R. Dobkins, Chair

Across four chapters, we answer four empirical questions and, in theorizing about each, we explore the relationships between moral or romantic evaluations of others and perceptions of others' desires and values. In Chapter 1, we ask whether the side-effect effect—an asymmetry in intention attribution—is a function of moral judgment, as leading hypotheses maintain. We test the predictions of an alternative moral-free hypothesis that

attributes the effect to the asymmetric involvement of costly and beneficial side-effects, and people's tendency to treat foreseen costs as sacrifices. We successfully produce the effect in cases that are morally neutral and successfully mute the effect in cases that are morally charged. In Chapter 2, we ask whether and to what extent the base rate of a behavior influences associated moral judgment. We find that base rates influence moral judgment with only very small effect sizes. Additional studies suggest that previous research purporting to find more substantial effects of base-rate information on moral judgment may have failed to properly isolate base rates from injunctive norms. We reconcile our findings with the more general but well-established impacts of social-influence by showing that the influence of others' behavior depends on whether we consider them part of a morally relevant reference group. In Chapter 3, we examine the factors underlying a blaming discrepancy between responses to a cyber peeper and to a peeping Tom; a discrepancy for which current theories of blame do not provide a straightforward explanation. We focus on major differences between cyber peeping and peeping Tom scenarios that may contribute to a blaming discrepancy, such as the effort involved on the part of the peeper and the risk-taking on the part of the victim. Four of eight such factors are found to account for most of the discrepancy. In Chapter 4, we use a new method to ask men and women to self-report the traits they prefer in a romantic partner and to indicate what they imagine the opposite gender prefers. The results reveal striking discrepancies between what people report wanting in a potential partner and what the opposite gender imagines they want.

INTRODUCTION

WHAT OTHERS WANT AND HOW WE EVALUATE THEM

Morality has been the subject of wonder for thousands of years, and philosophers have proposed numerous normative theories over the centuries. In recent decades, and with increasing attention, experimental psychologists have worked to develop descriptive theories of morality, aiming primarily to describe moral cognition and behavior as they are, rather than as they ought to be. Undoubtedly, well before recorded-history, thinkers were also entertained by the somewhat independent subjects of sexual attraction and mate selection. Particularly in the last three decades, researchers have made considerable progress developing scientific understandings of moral judgment and mate selection. However, because these fields are still in their infancy, the predictive power of current theory remains quite limited, and the answers to many rather basic questions remain unsettled.

Across four chapters, we answer four empirical questions and, in theorizing about each, we explore the relationships between moral or romantic evaluations of others and perceptions of others' desires and values. We suggest that to best understand others' behavior is to have considered their desires—among other things such as their behavioral options. This is not to suggest that any psychological mechanism will function in an ideal manner, nor that people's ability to infer others' desires is infallible (people's fallibility in inferring others' desires is a primary focus of Chapter 4). Rather, the idea is that the consideration of others' desires likely plays some role in any reasonably effective system of understanding and evaluating others' behavior (Reeder, 2009).

Indeed, to what ends others aim is understood to be of particular importance in describing moral judgments (Plaks, McNichols, & Fortune, 2009; Slote, 2003; Young & Tsoi, 2013). For example, whether the outcome of some behavior was intentional has been shown to be an important factor in ordinary blame attribution, character judgment, and punishment (e.g., Cushman, 2008; Young et al., 2010). However, as the scientific understanding of these phenomena develop, even foundational understandings upon which virtually everyone has agreed—such as the relationship between intention and moral judgment (Alicke & Rose, 2010)—have been called into question by researchers who may be underestimating the role that a consideration of others' desires plays in many social judgments (e.g., Egré, 2010; Knobe, 2003; Monroe, Dillon, Guglielmo, & Baumeister, 2018; Sauer, 2017; see Chapters 1 and 2). These desires include the goals and concerns that represent an actor's intentions and the moral values and preferences of others that collectively represent injunctive norms.

In Chapters 1 through 3, guided in part by the premise that considerations about others' desires play an important role in many social judgments, we test and ultimately find evidence to support alternative hypotheses for evaluative and attributional discrepancies that defy leading theories. In Chapter 4, we build on decades of research about the traits that are most important in mate selection. We examine the degree to which men and women may be mistaken about each other's romantic preferences, and explore potential reasons for any misunderstanding.

Chapter 1 asks whether the side-effect effect is a function of moral judgment. The side-effect effect is a phenomenon that has received a great deal of research attention over the past two decades (Laurent, Reich, & Skorinko, 2019). It refers to the asymmetric

intention attribution across responses to an actor creating either a morally good or morally bad side-effect, where a bad side-effect is more likely to be attributed with intention than a good side-effect (Knobe, 2003). The leading hypotheses maintain that the side-effect effect is a function of moral judgment and thus that intention attribution—upon which moral judgment is commonly thought to be based—is actually downstream from moral judgment. As a result of this interpretation, many believe the implication of the side-effect effect on our understanding of intention attribution is deeply revisionary (Sauer, 2017). Building primarily upon the insights of Machery (2008), we offer an alternative to this revisionary hypothesis. It proposes that the side-effect effect is a function of an asymmetry between costly and beneficial side-effects. Costly side-effects bear on one's decision in a way that beneficial side-effects do not, and their impact on one's decision depends on one's desires and concerns. In a series of experiments, our findings support this alternative hypothesis. As a result, we suggest that the side-effect effect is not a function of moral judgment, and therefore less revisionary than many suspect.

Chapter 2 asks whether, and to what extent, the base rate of a behavior influences moral judgments about that behavior and those who engage in it. Compared to the side-effect effect, this issue has only more recently been the focus of researchers' attention. The existing literature suggests that base rates are of critical importance to moral judgments (e.g., Monroe et al., 2018). However, across multiple experiments, we found that while base rates do indeed influence moral judgments, they do so with only very small effect sizes. We reason that previous research purporting to find more substantial effects of base-rate information on moral judgments may have conflated that which is descriptively typical (i.e., the base-rate) with that which is typically endorsed (i.e., the injunctive norm). That is, the researchers may

have ignored the influence of their participants' familiarity with other people's moral values. The results of our experiments support this reasoning. To reconcile our findings with well-established knowledge regarding the many powerful impacts of social-influence, we test the hypothesis that it is not simply what is done by others, even many others, that matters much to moral judgments, but instead what is done by those of a more morally relevant reference group. Our findings suggest that the extent to which others' behavior influences our moral judgments may depend heavily on what we think of those others.

Chapter 3 examines the large difference between the blaming of cyber peepers and the blaming of peeping Toms. To our knowledge, we are the first to investigate the factors contributing to this discrepancy. The phenomenon exposes the limitations of leading theories of blame, which offer no straightforward explanation for the discrepancy. In a series of studies, we focus in on major differences between cyber peeping and peeping Tom scenarios that may contribute to a blaming discrepancy, such as the effort involved on the part of the peeper and the risk-taking on the part of the victim. The perpetrator's degree of effort and his prurient interests, both of which can be understood in terms of his intentions and desires—and both of which may play a role in evaluations of his character (Uhlmann, Pizarro, Diermeier, 2015)—are among the four factors that in our results appear to contribute to the blaming discrepancy. The other two influential factors are whether the victim was a celebrity and whether a middleman was involved. The number of other transgressors engaging in the same behavior, which typically differs between cyber peeping and peeping Tom cases, is among the factors that do not appear to contribute to the blaming discrepancy. This finding is consistent with the findings of Chapter 2, in which we show that the base rate of a behavior has only a very small influence on moral judgments.

Finally, Chapter 4 asks to what extent there are discrepancies between the traits that men and women say they prefer in a romantic partner and the traits the opposite gender imagines they prefer. Therefore, in Chapter 4 we turn our attention away from moral evaluations and toward romantic evaluations. And unlike previous chapters, in which we consider inferences about others' desires as part of various hypotheses to explain moral judgments, in Chapter 4, we measure inferences about others desires directly. Previous research has found that in their attempts to appeal to the opposite gender, men emphasize their wealth, status, and ambition, whereas women emphasize their physical attractiveness (e.g., Buss & Dedden 1990; Campos, Otta, de Oliviera Siqueira, 2002; Cicerello & Sheehan, 1995; Fletcher, Simpson, Campbell, & Overall, 2013). Such behavior seems surprising given other research in which men and women both report personally finding these traits to be less important than others such as trustworthiness, intelligence, and warmth (e.g. Buss & Barnes, 1986; Fletcher, Simpson, Thomas, & Giles, 1999; Lippa, 2007). This suggests, and our findings confirm, that there are striking differences between the traits men and women find most desirable and those that the opposite gender believes they find most desirable. We also find that in some, but not all ways, women appear to be better at imagining men's preferences than men are at imagining women's preferences. To our knowledge we are the first to investigate these comparisons. As part of this research, we develop a new method that, compared to existing survey methods, allows us to better characterize preferences along various dimensions and to better measure the importance of each trait. In future research, this method could be adapted to examine desires and preferences in other domains.

References

Alicke, M., & Rose, D. (2010). Culpable control or moral concepts?. Behavioral and Brain Sciences, 33(4), 330-331.

Buss, D. M., & Barnes, M. (1986). Preferences in human mate selection. *Journal of personality and social psychology*, *50*(3), 559.

Buss, D. M., & Dedden, L. A. (1990). Derogation of competitors. *Journal of Social and Personal Relationships*, *7*, 395–422.

Campos, L., Otta, E., & de Oliveira Siqueira, J. (2002). Sex differences in mate selection strategies: content analyses and response to personal advertisements in Brazil. *Evolution and Human Behavior*, *23*, 395–406.

Cicerello, A., & Sheehan, E. P. (1995). Personal advertisements: a content analysis. *Journal of Social Behavior and Personality*, *10*, 751–756.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. Cognition, 108(2), 353-380.

Egré, P. (2010). Qualitative judgments, quantitative judgments, and norm-sensitivity. Behavioral and Brain Sciences, 33(4), 335-336.

Fletcher, G. J., Simpson, J. A., Campbell, L., & Overall, N. C. (2013). *The science of intimate relationships* (Vol. 323). Malden, MA: Wiley-Blackwell.

Fletcher, G. J. O., Simpson, J. A., Thomas, G., & Giles, L. (1999). Ideals in intimate relationships. *Journal of Personality and Social Psychology*, *76*, 72–89.

Knobe, J. (2003). Intentional action and side effects in ordinary language. Analysis, 63(3), 190-194.

Laurent, S. M., Reich, B. J., & Skorinko, J. L. (2019). Reconstructing the side-effect effect: A new way of understanding how moral considerations drive intentionality asymmetries. Journal of Experimental Psychology: General.

Lippa, R. A. (2007). The preferred traits of mates in a cross-national study of hetero- sexual and homosexual men and women: an examination of biological and cultural influences. *Archives of Sexual Behavior*, *36*, 193–-208.

Machery, E. 2008: The folk concept of intentional action: philosophical and experimental issues. Mind & Language, 23, 165– 189.

Monroe, A. E., Dillon, K. D., Guglielmo, S., & Baumeister, R. F. (2018). It's not what you do, but what everyone else does: On the role of descriptive norms and subjectivism in moral judgment. *Journal of Experimental Social Psychology*, *77*, 1-10.

Plaks, J. E., McNichols, N. K., & Fortune, J. L. (2009). Thoughts versus deeds: Distal and proximal intent in lay judgments of moral responsibility. Personality and Social Psychology Bulletin, 35(12), 1687-1701.

Reeder, G. (2009). Mindreading: Judgments about intentionality and motives in dispositional inference. Psychological inquiry, 20(1), 1-18.

Sauer, H. (2017). Between Facts and Norms: Ethics and Empirical Moral Psychology. In Moral Psychology (pp. 5-27). Springer, Cham.

Slote, M. A. (2003). Morals from motives. Oxford University Press, USA.

Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science*, 10(1), 72-81.

Young, L., Bechara, A., Tranel, D., Damasio, H., Hauser, M., & Damasio, A. (2010). Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. Neuron, 65(6), 845-851.

Young, L., & Tsoi, L. (2013). When mental states matter, when they don't, and what that means for morality. Social and personality psychology compass, 7(8), 585-604.

CHAPTER 1

IS THE SIDE-EFFECT EFFECT A FUNCTION OF MORAL JUDGMENT?

Carl P. Jago

Abstract

In a series of studies, we ask whether the side-effect effect is a function of moral judgment. The effect is the asymmetric intention attribution observed across morally good and morally bad side-effects, wherein bad side-effects are more likely to be attributed with intention than good side-effects. The leading hypotheses maintain that the effect is a function of moral judgment. An important implication of this claim is that intention attribution—upon which moral judgment is commonly thought to be based—is actually downstream from moral judgment. We test the predictions of these hypotheses against those of an alternative that does not rely on morality in its explanation. This moral-free alternative attributes the effect to a tendency for observers to treat foreseen costs as sacrifices in the sense that those costs are considered intentional. Our studies' results support this hypothesis. First, we demonstrate the importance of an actor's decision-costs to an observer's intention attribution. Then, using costs, we both successfully produce the side-effect effect in cases that are morally neutral and successfully mute the effect in cases that are morally charged. We also show the side-effect effect should not be interpreted as an outright theory-of-mind failure.

When a company chairman, who admits he does not care about the environment, enacts a plan to increase profits despite knowing that the plan will also harm the environment, the vast majority (82% in one study) of people say that he harmed the environment intentionally. If, instead, the same chairman, who again admits he does not care about the environment, enacts a plan to increase profits with full knowledge that the plan will also help the environment, relatively few people (only 23% in that same study) say that he helped the environment intentionally (Knobe, 2003). This effect, known as the side-effect effect, has entertained moral psychologists for nearly two-decades (Laurent, Reich, & Skorinko, 2019). Evidently, the phenomenon has proven challenging to explain without conceding that the relationship between intention attribution and moral judgment does not align with conventional thinking. Along the lines of conventional thought, moral judgments depend on, and follow after, an assessment of whether somebody's behavior, or its consequence, was intentional. Folk psychology says that, morally speaking, it matters whether somebody meant for the consequences of an action to happen (Karniol, 1978; Malle & Knobe, 1997). Yet, according to many researchers, the side-effect effect suggests that the reverse is true, descriptively speaking (e.g., Egré, 2010; Knobe, 2010; Mallon, 2008; Menzies, 2010; Sauer, 2017). The basic interpretation is that people are more likely to consider bad outcomes intentional than they are to consider good outcomes intentional (Knobe, 2003). According to this interpretation, intentionality, a concept previously thought to be non-evaluative, is actually downstream from moral judgment, to the extent that one assesses morality before and intentionality after (Sauer, 2017). This phenomenon should be

of interest to anyone curious about intentional action, from philosophers and psychologists to legal professionals (Laurent et al., 2019). Leading researchers consider the implications of the side-effect effect to be deeply revisionary (Sauer, 2017), but whether this belief is correct depends on the veracity of the interpretation at hand. In the current project, we test the predictions of two alternatives to this revisionary interpretation—one that builds on it and one that challenges it—and in so doing, we ask whether the side-effect effect is a function of moral judgment.

To further clarify what it is that so many have found surprising, consider first that an actor's goals are the outcomes that incentivize their behavior. We take for granted that those outcomes are desired and, when they are obtained, they are obtained intentionally. In contrast to goals, foreseen side-effects are non-incentivizing outcomes in that they are undesired. Given that intention is largely about acting on desires (Malle & Knobe, 1997), we might expect that the attainment of undesired side-effects would not be attributed with intention. In the paradigmatic thought experiment about the company chairman who pursues profit, people's responses to the help case are consistent with this expectation, in that the majority of people do not attribute intention to the attainment of the undesired outcome. It is in response to the harm case, with its morally bad side-effect, that most people diverge from this expectation and do attribute intention to the undesired outcome. Therefore, we will place a disproportionately large emphasis on trying to understand the harm case, and especially how intention attribution relates to beliefs about the chairman's desires.

For researchers who see the moral quality of the side-effect as the only difference between the help and harm cases (e.g., Gintis, 2010; Knobe 2010), some version of the basic interpretation—according to which the side-effect effect is a function of that moral

difference—seems to be the only available explanation. But others (e.g., Machery, 2008**;** Nanay, 2010; Girotto, Surian, & Siegal, 2010) have pointed out that the variable moral quality of the side-effect is confounded by whether the chairman is confronted with an opportunity to pursue two benefits (help case), or an opportunity to pursue a benefit accompanied by a cost (harm case). This means that, in the harm case, the chairman must decide whether to make a trade-off. This is not so in the help case. To defend against this challenge, a researcher could argue that there can be no trade-off if the actor does not care about the side-effect, because the side-effect would neither be regarded as a cost nor as a benefit. However, this defense may not go far enough, because, even when an actor claims not to care about an outcome, it is unclear if the actor is truly indifferent or whether they mean only that they do not care enough about the outcome for it to impact their decision. Moreover, even truly indifferent actors may still have the ability to distinguish between costs and benefits about which they do not care. Therefore, before accepting that the side-effect effect is a function of the moral quality of the side-effect, we should continue to explore whether costliness—i.e., the confound—might influence intention attribution.

A priori, the suggestion that costliness could impact intention attribution is plausible because costly side-effects relate differently to an actor's desires than do beneficial side-effects. The company chairman's decision to proceed with the harmful plan depends specifically on him subjectively weighing the profit more heavily than the environmental harm. Therefore, in the harm case, but not in the help case, there is a counterfactual consideration whereby the actor would have chosen otherwise if he had cared more about the side-effect (Girotto, Surian., & Siegal, 2010). Provided foresight is unambiguous, that he did not care enough to choose otherwise is, given his deliberate choice to proceed, also

unambiguous. This is also true of costly side-effects in general, rather than only being true of morally bad side-effects. The costly side-effect (harm) bears on the chairman's decision in a way that the beneficial side-effect (help) does not; and its impact on his decision depends on his own desires and concerns. This leads to a confound that involves a systematic difference in the mental state upon which intentionality is based. However, even if costliness plays a role in increasing the likelihood of intention attribution due to its unique relationship with the actor's desires, it remains unclear whether the side-effect effect is a function of a larger moral assessment in which costliness plays a role or, whether the side-effect effect is simply a function of the side-effect being a cost. These two possibilities are represented by the following alternatives to the basic interpretation.

We refer to the first alternative as the Sufficient Commitment Hypothesis. Like the basic interpretation, this alternative maintains that the side-effect effect is a function of moral judgment. According to this view, intention attribution is downstream from moral judgment (Sauer, 2017), and suffused with moral judgment from the outset (Knobe 2010). It is believed that outcomes, whether good or bad, are more likely to be attributed with intention when they result from an attitude that is judged to be morally good or bad than when they result from an attitude that is judged to be merely acceptable. In other words, as observers, we judge the actor's decision in aggregate (balancing competing considerations) and attribute intention when we determine that the actor's concern for the negative outcome, or desire for the positive outcome, is deserving of moral criticism or praise. This occurs when the actor's attitude toward the outcome appears to either exceed or fail to meet the levels of concern or desire considered acceptable. In contrast to the basic interpretation, which focuses on a moral assessment of the outcome and attributes intention in a unipolar manner (where bad

outcomes are more likely attributed with intention than good outcomes), the Sufficient Commitment Hypothesis focuses on morally assessing the actor's commitment to achieving or avoiding the outcome. It also attributes intention in a bipolar manner (where both good and bad outcomes are likely to be attributed with intentional if they demonstrate a commitment sufficient to warrant moral praise, or so insufficient as to warrant moral criticism). This hypothesis aligns most closely with Knobe's (2010) view, and its foundational element—that the side-effect effect is a function of moral judgment—is a belief echoed by numerous others (e.g., Egré, 2010; Mallon, 2008; Menzies, 2010; Sauer, 2017) and has been promulgated as the distillation of this discourse (e.g., Knobe & Doris, 2010; Sauer, 2017).

We will refer to the second alternative as the Any Sacrifice Hypothesis. While the Sufficient Commitment Hypothesis expands and modifies the basic interpretation, the Any Sacrifice Hypothesis opposes such revisionary accounts by challenging the notion that the side-effect effect is a function of moral judgment. According to this view, a foreseen cost that an actor accepts in their intentional pursuit of some goal will tend to be treated as a sacrifice. Given that a prototypical sacrifice is something made (i.e., a means to an end) rather than something merely allowed to happen (i.e., a true side-effect), to treat a foreseen cost as a sacrifice is to treat it as intentional. The Any Sacrifice Hypothesis maintains that, relative to auxiliary benefits that people will more likely view as unintentional, people will likely see both goals and foreseen costs as intentional regardless of whether these outcomes are morally relevant and regardless of whether the actor's behavior demonstrates a commitment sufficient to warrant moral praise, or one so insufficient as to warrant moral criticism. This hypothesis builds most notably on the contribution of Machery (2008). More generally, it

represents to greater or lesser degrees many (e.g., Girotto, Surian, & Siegal, 2010; Mandelbaum, & Ripley, 2010; Nanay, 2010) who remain skeptical of the more revisionary interpretations discussed above.

The studies that follow test predictions that differentiate the Sufficient Commitment and Any Sacrifice hypotheses first, from the basic interpretation and second, from one another. Ultimately, the current project provides an answer to the question of whether—as maintained by the Sufficient Commitment Hypothesis—the side-effect effect is a function of moral judgment or—as maintained by the Any Sacrifice Hypothesis—it is a function of cost.

Studies 1 and 2 challenge the basic interpretation of the side-effect effect. Study 1 demonstrates the diagnosticity of costs that can guide inferences about an actor's priorities. Study 1 shows that cost is important to intention attribution and is therefore a potential confound in the original company chairman studies. Study 2 tests whether the pattern of intention attribution seen in the side-effect effect tracks the mental states and desires of the actor, or instead—as suggested by the basic interpretation—it tracks the goodness or badness of the side-effect per se (which is a subjective assessment on the part of the observer). Study 3 tests a unique prediction of the Any Sacrifice Hypothesis: in comparison to the attainment of foreseen and undesired benefits, people will more likely consider the attainment of foreseen and undesired costs as made rather than allowed. Studies 4 and 5 test the two most important predictions that differentiate the Sufficient Commitment and Any Sacrifice hypotheses from one another. Specifically, and in contrast to the Sufficient Commitment Hypothesis, the Any Sacrifice Hypothesis predicts that people will view foreseen costs as intentional, regardless of their moral relevance (Study 4) or their magnitude in relation to some sufficiency threshold (Study 5). Study 6 tests a prediction of a third alternative

hypothesis inspired by the results of Study 5. Finally, Study 7 tests the extent to which the results of Study 5 replicate.

Study 1

When the pursuit of some goal involves a cost, an actor must make a trade-off. Study 1 demonstrates how pairing a cost with a benefit, a cost with a cost, or a benefit with a benefit, influences an observer's intention attribution in a way that is consistent with the alternative hypotheses and inconsistent with the basic interpretation. Specifically, intention attribution for a secondary outcome, about which the actor's feelings are somewhat ambiguous, depends largely on the presence of a cost (or bad outcome). This prediction contrasts with that of the basic interpretation, whereby intention attribution for the secondary outcome instead depends on whether that outcome itself was morally bad.

Method

One hundred sixty-three American participants (53% female, M_{age} = 38.8 years, SD = 13.5) recruited on Mechanical Turk provided their age and gender, read one vignette picked randomly from four possible variations, and then supplied one response regarding whether the actor obtained a particular outcome intentionally. The original side-effect effect study (Knobe, 2003) involved approximately 80 participants and achieved high statistical power owing to its very large effect size. We aimed to double the sample size because we were using four groups rather than two, and two critical chi square tests rather than one.

The vignette involved two manipulated variables, namely a greater or lesser personal financial cost along with a helpful or harmful impact on another person. Crossing these two variables produced four variations.

The vignette read as follows (underlining highlights the manipulated variable; brackets contain the alternative level of the manipulated variable):

> A property owner was deciding whether to build his new beachfront home on the north or south side of his lot of land. At first, he had no preference; he didn't care on which side he built. But then he learned two things about his choice:
>
> *building on the south side would cost him less than building on the north side.
>
> *building on the south side would block [spare] the ocean-view from his brother-in-law's neighboring house, whereas building on the north side would spare [block] his brother-in-law's view.
>
> The property owner is known for being very competitive with his brother-in-law. However, he also doesn't like to waste money. Ultimately, he decided to build on the south [north] side where it would cost him less [more] and where his brother-in-law's view would be blocked [spared].

The property owner therefore either saved more money or spent more, and either spared his brother-in-law's view or blocked it. Each subject saw only one of the four vignette variations for a between-subjects design.

After reading the vignette, participants responded to the following question: "Did he block [spare] his brother-in-law's view intentionally?" The response options they could choose were, "Yes" or "No."

Results

When the property owner saved money and blocked his brother-in-law's view, 65.0% of participants said he blocked the view intentionally. In contrast, when he saved money and spared his brother-in-law's view, only 28.9% of participants said he spared the view intentionally ($X^2(1) = 10.40$, $p = .001$, $\varphi = .365$). Thus, consistent with the results of previous studies demonstrating the side-effect effect, when the actor seemed motivated by a personal benefit, intention attribution for the harmful outcome was higher than for the helpful outcome.

The pattern of results differed when it came to the additional pair of vignettes, where the property owner decided to take on the higher monetary cost. In both additional cases, the majority of people said the outcomes were intentional. Specifically, 92.7% of participants said he blocked the view intentionally and 88.5% said he spared the view intentionally. Thus, a similarly high majority attributed intention to both the harmful and helpful outcomes ($X^2(1) = 0.41$, $p = .521$, $\varphi = .070$).

Discussion

In the current study, the pair of vignette variations that best parallels the classic vignette (the company chairman pursuing a greater profit alongside a helpful or harmful impact on the environment) is the pair where the property owner saves money by building on the cheaper side of the lot, either blocking or sparing his brother-in-law's view. Consistent with responses to the chairman vignette, participants attributed the harmful outcome in this study with greater intention than the helpful outcome. However, in the second pair of vignettes where the property owner decides to take on the higher monetary cost, most participants indicate that both the helpful and harmful impacts were intentional. This suggests that it is not the character of the outcome itself that primarily impacts intention attribution, but instead the relationship between the outcomes. Across all four variations, most participants attributed intention if there was a cost involved.

Interestingly, when the property owner seems to take on two negative outcomes (i.e., a higher monetary cost and a harmful impact on his brother-in-law's view), the outcome toward which his feelings are more ambiguous (blocking the view) is the one that participants seem to assume must be his goal (to him a benefit). In this case, opting to take on a monetary cost is strong evidence of his spiteful commitment to blocking his brother-in-

law's view (e.g., Dik & Aarts, 2007). Similarly, in the variation where he accepts a monetary cost to spare the view, that monetary cost is strong evidence of his kind commitment to sparing the view. Therefore, intention attribution did not appear to depend on the helpfulness or harmfulness of the outcome. Instead, it was more likely in the cases involving costs.

Study 2

Any assessment of goodness or badness is largely subjective, and the basic interpretation of the side-effect effect is that people are more likely to attribute bad outcomes than good outcomes with intention. Given this, anybody who adopts the basic interpretation might imagine that the side-effect effect tracks the moral beliefs of the observer rather than the moral beliefs of the actor. If this is true, then the side-effect effect could demonstrate an outright failure of theory-of-mind. Common intention attribution, which purports to represent an actor's mental states, would instead represent the observer's own interests and beliefs.

Study 2 involves a vignette where an actor makes a trade-off involving a false belief, so that the harmful side-effect is the one that the actor believes to be helpful, and the helpful side-effect is the one he believes to be harmful. If intention attribution with regard to the side-effect depends on the observer's subjective assessment of that side-effect, one should predict that participants would be more likely to attribute intention in the case where participants know the side-effect is harmful. If intention attribution instead depends on the actor's subjective assessment, one should predict that participants would be more likely to attribute intention in the case where the actor believes the side-effect is harmful.

Method

One hundred and one American participants (55% female, M_{age} = 36.0 years, SD = 10.1) recruited on Mechanical Turk provided their age and gender, read a new vignette

randomly chosen from two variations, and then provided one response regarding the helpfulness or harmfulness of the actor's decision and whether it was intentional.

The new vignette involved a 19th century leech researcher with a belief about the utility of bloodletting that was common in his time but ultimately false. We experimentally manipulated whether pursuing his research goals would also mean beginning or putting an end to the bloodletting of his human subjects. Compared to the participants, who were made aware that beginning bloodletting would be harmful and ending it would be helpful, the researcher was said to have believed the reverse.

The vignette read as follows:

> Up until the 19th century, even the best medical professionals incorrectly believed that bloodletting was an effective way to improve or maintain health. Trusted methods for bloodletting included allowing leeches to suck blood from patients' bodies. Today we know these beliefs were incorrect, and that leeches are harmful rather than helpful.
>
> During this period (the 19th century) the assistant to a leech researcher approached the researcher with a plan to advance their understanding of the leech life cycle. The plan involved beginning to hold [no longer holding] research sessions where leeches would suck blood from human subjects. "If we go ahead with this plan," the assistant explained, "it will advance our research, and it will also help the human subjects by beginning [but it will also harm our research subjects by stopping] bloodletting sessions."
>
> "Look," the researcher responded, "I understand that this plan will also help [harm] the subjects, but I don't care at all about that, I just want to advance our leech research. Let's go ahead with the plan."
>
> They went ahead with their plan to advance their research and they believed that, as a side-effect, it would also help [harm] the subjects. However, today we know their belief was wrong and that the true side-effect of their plan was actually harmful [helpful].

After subjects read the vignette, we asked them about the actor's intentions. The standard approach is to ask, for example, whether the actor harmed some entity intentionally, but such an approach assumes agreement about what happened. To avoid guiding participants' responses toward one belief over another, we asked, "Which sentence best

describes the researcher's intention toward, and actual impact on, the human subjects?" We provided three groups of three sentences for a total of nine. The first three began with, "he intended to help", the following three began with "he intended to harm" and the last three began with, "he neither intended to help nor intended to harm". Within each group of three, the remainder of each sentence read either: "and he actually helped," "but he actually harmed," or, "but he actually neither helped nor harmed." Whether we used "and" or "but" to conjoin the clauses depended on whether the two clauses were in agreement or in contrast with one another.

Results

Among the participants who read the variation where the researcher thought the harmful side-effect was helpful, the most frequently chosen response, at 64.7%, was that "he neither intended to help nor intended to harm, but he actually harmed." The next most frequently chosen response, at 29.4%, was that "he intended to help but he actually harmed." The remaining responses were each chosen by less than 4% of the participants.

Among the participants who read the variation where the researcher thought the helpful side-effect was harmful, the most frequently chosen response, at 38.0%, was that "he intended to harm but he actually helped." The next most frequently chosen response, at 30.0%, was that "he neither intended to help nor intended to harm but he actually helped." The remaining responses were each chosen by less than 8% of the participants.

For each group of participants, combining the responses that affirm intention (i.e., those beginning with "he intended") showed that 35.3% of participants attributed intention in the variation where the actor thought the harmful side-effect was helpful, whereas 62% of

participants attributed intention in the variation where the actor thought the helpful side-effect was harmful ($X^2(1) = 7.30$, $p = .007$, $\varphi = .269$).

Discussion

The results of Study 2 challenge any version of the basic interpretation in which the effect tracks the moral beliefs of the observer rather than those of the actor. Participants were more likely to attribute intention when the actor (rather than the observer) believed the side-effect was harmful, and they were less likely to do so when the actor (rather than the observer) believed the side-effect to be helpful. To the extent that the likelihood of intention attribution is positively related to the harmfulness of the side-effect, what matters here is the harmfulness perceived by the actor and not the harmfulness perceived by the observer. This result is more consistent with the alternative hypotheses than with the basic interpretation. The alternatives place a greater focus on the actor's subjective perspective as he weighs the competing considerations. The basic interpretation bypasses this altogether.

A researcher might take issue with the fact that, even while the actor correctly foresees the immediate side-effect (i.e., beginning or ending bloodletting), he does not correctly foresee its ultimate helpful or harmful impact. Indeed, it is upon the actor's false belief that this experiment is based. However, the very point is that the importance of the complete veracity of the actor's foresight seems to be a testament to the primary importance of his subjective mental states and his attitudinal relationships with the outcomes. It is not obvious why else the actor's imperfect foresight would reverse the pattern of the effect rather than simply attenuate it.

Although the side-effect effect does not appear to demonstrate a failure of theory-of-mind to the degree suggested by the basic interpretation (according to which matters of

mental state are inferred without a consideration for the actor's subjectivity), the Sufficient Commitment Hypothesis also maintains that intention attribution ultimately relies on the observer's subjective moral assessment. However, this Sufficient Commitment Hypothesis at least allows for a representation and consideration of the actor's beliefs and desires. The criteria for intentionality in the Any Sacrifice Hypothesis are even more objective in that it excludes the observer's subjective moral assessment of the behavior. However, we do not mean to suggest that the hypothesis with the more objective criteria is necessarily the one that more accurately describes the psychological mechanisms behind the side-effect effect. In addition, to whatever extent these alternative hypotheses describe psychological mechanisms that fail to correctly infer the actor's intention, they too represent a failure of theory of mind—though perhaps to a lesser extent than the basic interpretation.

Study 3

The Any Sacrifice Hypothesis maintains that the side-effect effect results from observers treating foreseen costs as sacrifices. A prototypical sacrifice is a means to an end. It is also an intermediate or secondary goal, and something that one actively pursues rather than passively allows. If the Any Sacrifice Hypothesis is correct that foreseen costs are treated as sacrifices, then we may expect people to respond differently to the premise of the question in the original chairman experiment according to whether they read the help or harm variation. The researchers asked participants whether the chairman harmed [helped] the environment intentionally. The premise of the question takes for granted that participants agreed that the chairman harmed [helped] the environment. This is not the same as saying he allowed the harm [help] to occur. It seems possible, as the Any Sacrifice Hypothesis leads us to imagine, that participants could be more inclined to agree that the chairman harmed the

environment in the harm case than they are inclined to agree that the chairman helped the environment in the help case. Indeed, it is possible that there is a discrepancy between responses to the help and harm cases before participants even begin to address intentionality as an issue. In Study 3, we test this possibility by replicating Knobe's (2010) original study, although we modify one element: instead of asking participants whether the chairman harmed [helped] the environment intentionally, we ask only whether he harmed [helped] the environment.

Method

Eighty-five American participants (60% female, M_{age} = 35.3 years, SD = 11.8) recruited on Mechanical Turk participated in a study identical to that of Study 2, but with two exceptions. We replaced the vignettes and the questions with those from Knobe's (2003) original experiment—the study which first described the side-effect effect.

The vignette read as follows:

> The vice-president of a company went to the CEO and said, "We are thinking of starting a new program. It will help us increase profits, but [and] it will also harm [help] the environment."
>
> The CEO answered, "I don't care at all about harming [helping] the environment. I just want to make as much profit as I can. Let's start the new program."
>
> They started the new program. Sure enough, the environment was harmed [helped].

We removed the word "intentionally" from Knobe's original question. We also added the words, "In your opinion," to the start of the sentence to ensure that participants knew they were free to choose either option. Therefore, the question read as follows: "In your opinion, did the CEO harm [help] the environment?" The response options were either "Yes", or

"No". The question used the word "harm" for those participants who read the harm variation of the vignette, and it used "help" for those who read the help variation.

Results

Among the participants who read the harm variation of the vignette, 97.6% responded with, "Yes", to the question of whether the chairman harmed the environment. In contrast, among those who read the help variation of the vignette, 58.1% responded with, "Yes", to the question of whether the company chairman helped the environment ($X^2(1) = 22.41$, $p < .001$, $\varphi = .513$).

Discussion

Study 3's results support the unique prediction of the Any Sacrifice Hypothesis, wherein foreseen costs are treated as sacrifices in that they are regarded as something more actively achieved than passively allowed.

While the Sufficient Commitment Hypothesis does not readily inspire this prediction, a researcher could argue that the finding is not necessarily at odds with it. They could say that the participants believe the chairman should be held responsible for harming the environment because doing so is morally bad. It could be for this reason that participants both say that he harmed the environment and say that he harmed it intentionally.

The results seem consistent with both hypotheses—although not equally supportive of each—because this study maintains the confound of the original experiment. Specifically, the harmful plan differs from the helpful plan in two ways. It involves a foreseen cost and it is morally bad.

Study 4

Following Machery (2008) and a critical prediction of the Any Sacrifice Hypothesis, Study 4 tests whether including a cost in one case but not the other, regardless of its moral relevance, is enough to produce a pattern of intention attribution that resembles the side-effect effect. If the side-effect effect is a function of moral judgment, as the Sufficient Commitment Hypothesis maintains, then we would not expect to see the classic asymmetric pattern of results across two variations of a vignette that are morally neutral. Instead, if the side-effect is produced by a cost asymmetry, then we would expect to see the classic pattern of results across two variations of a vignette where one involves a goal accompanied by a foreseen auxiliary benefit, and the other involves a goal accompanied by a foreseen cost.

Method

Eighty-seven American participants (50.6% female, M_{age} = 34.5 years, SD = 10.1) recruited on Mechanical Turk participated in a study identical to Study 3, with the exception that we replaced the vignette and the follow-up question.

The vignette read as follows:

> Tim is staying overnight in a small town. He tells his innkeeper that he'd like to get some tacos for dinner.
>
> The innkeeper responds, "Only a taco shop and a pizza shop are open. You can get tacos at the taco shop, <u>but you will miss out on the free beer they are serving with any pizza at the pizza shop</u> [and you will also get the free beer they are serving with any tacos]."
>
> Tim then says, "I don't care at all about the free beer, I just want to get some tacos. Could you please direct me to the taco shop?"
>
> Tim goes to the taco shop. Sure enough, <u>he does not get</u> [he gets] a free beer with his meal.

In making his decision to visit the taco shop, Tim receives a free beer in one case and not in the other. We asked participants, "Did Tim <u>get</u> [miss out on] the free beer intentionally?" As in Study 3, the available responses were "Yes," or "No.

Results

Among the participants who read the variation in which Tim does not receive a free beer, 83.3% responded with, "Yes", to the question of whether Tim missed out on the free beer intentionally. In contrast, among those who read the variation in which Tim receives the free beer, only 20.0% responded with, "Yes", to the question of whether he received the free beer intentionally ($X^2(1) = 37.71$, $p < .0001$, $\varphi = .658$).

Discussion

Study 4's results challenge the proposed role of moral judgment because they show the asymmetric pattern of intention attribution known as the side-effect effect. The pattern of results, including the proportions and the effect size, is nearly identical to that of the original study (Knobe, 2003). The latter's results relied on the profit-seeking company chairman vignette, which is an example that confounded morality and cost. The side-effect effect appears to result from a cost asymmetry in the absence of any moral difference. This directly supports the Any Sacrifice Hypothesis and challenges the Sufficient Commitment Hypothesis. The results suggest that, compared to undesired foreseen auxiliary benefits, people are likely to view foreseen costs as intentional, regardless of the costs' moral relevance.

Although a moral asymmetry does not seem necessary to produce the side-effect effect, it remains possible that moral judgments may still impact intention attribution in cases where the actor's decisions are morally relevant (i.e., morally good and bad rather than morally neutral).

Study 5

Study 5 tests another important prediction that differentiates the Any Sacrifice Hypothesis from the Sufficient Commitment Hypothesis. The Sufficient Commitment Hypothesis maintains that moral judgment drives the side-effect effect in the following manner: intention attribution increases when the actor's attitude toward the outcome appears to either exceed or to fail to meet the levels of concern or desire that people consider acceptable. In other words, it increases when people deem the actor deserving of moral praise or criticism. Although a moral asymmetry does not seem necessary to produce the side-effect effect in morally neutral contexts (see Study 4), it remains unclear whether a moral asymmetry may still play a role in morally charged contexts. To test this, Study 5 uses a vignette where both variations involve a harmful side-effect. In one and not in the other, the accompanying benefit morally justifies the cost of that side-effect, so people deem it morally acceptable.

According to the Sufficient Commitment Hypothesis, we should expect to see a pattern of results resembling the side-effect effect, wherein a majority of participants who read about the morally acceptable harm will say it was not intentional, and a majority of those who read about the morally unacceptable harm will say it was intentional. In contrast, according to the Any Sacrifice Hypothesis, we should expect both groups to attribute intention and, critically, the proportion who do so should not differ across the morally acceptable and unacceptable variations.

Method

Two hundred and three American participants (50.0% female, M_{age} = 40.4 years, SD = 13.9) recruited on Mechanical Turk participated in a study identical to Study 4, with two exceptions. We replaced the vignette and the question, and we included a manipulation

check. The effect, if any were to emerge, would be of a size yet to be determined because we modified the paradigm to involve a cost in both variations. An a priori power analysis using the proportions of 0.4 and 0.6—which simply represents a smaller difference in proportions than in Study 4—showed we would need at least 194 participants to achieve 80% power. We aimed for 200.

The vignette, which was based on the classic trolley problem (Foot, 1967), read as follows:

> Tim, a railway worker, sees a runaway trolley that is speeding down the tracks. The trolley is headed toward his five friends [one friend] who are [is] working on the tracks. His friends will all be [friend will be] killed if the trolley continues on its current course. Tim is standing next to a large switch that can divert the trolley onto a different track on which one other person is [five other people are] working. The only way to save the lives of his friends [life of his friend] is to divert the trolley onto the other track. Tim knows that if he diverts the trolley onto the other track, he will kill the one worker [five workers], but he will save his five friends [one friend].
>
> Tim pulls the switch and, as expected, his friends are [friend is] saved and the other worker is [workers are] killed.

In one case, Tim decides to save five friends at the cost of one other person's life. In the other case, Tim decides to save one friend at the cost of five other people's lives.

After reading the vignette, the participants responded to the following question: "Did Tim kill the other worker [workers] intentionally?" In addition, we included a manipulation check, asking, "Was Tim's behavior morally acceptable?" For both questions, the response options were, "Yes," or "No."

Results

Among the participants who read the variation where Tim saves five at the cost of one (the morally acceptable variation), 64.0% responded with, "Yes", when questioned as to whether Tim killed the other worker intentionally. Among those who read the variation in

which Tim saved one at the cost of five (the morally unacceptable variation), 78.6% responded with, "Yes", when questioned as to whether Tim killed the other five workers intentionally. These proportions were different ($X^2(1) = 5.36$, $p = .021$, $\varphi = .162$), albeit with a much smaller effect size compared to those which one finds in previous studies demonstrating the side-effect effect. Therefore, a greater majority said Tim killed intentionally in the morally unacceptable variation than in the morally acceptable variation. Even so, majority responded with, "Yes," in both the acceptable variation ($X^2(1) = 7.95$, $p = .005$, $\varphi = .198$) and the unacceptable variation ($X^2(1) = 35.94$, $p < .0001$, $\varphi = .421$).

The responses to the manipulation check validated the independent variable. Among participants who read the morally acceptable variation, 77.0% responded with, "Yes", when asked whether Tim's behavior was morally acceptable. Among those who read the morally unacceptable variation, just 22.3% responded, "Yes" ($X^2(1) = 64.12$, $p < .0001$, $\varphi = .562$).

Discussion

Consistent with the Any Sacrifice Hypothesis, participants were more likely than not to attribute intention in both variations, arguably due to the involvement of a cost. However, the finding that a greater majority attributed intention to the actor's behavior in response to the morally unacceptable variation than in response to the morally acceptable variation is inconsistent with the Any Sacrifice Hypothesis. This suggests that, although moral judgment may not drive the side-effect effect, it may still magnify it or attenuate it. However, the ability to draw this conclusion is limited in two important ways.

The first limitation is due to the fact that the degree of the cost differs across the variations (one worker's death versus five workers' deaths). This means that the difference in intention attribution, though reduced, may yet still be due to a difference in costs rather than

due to a difference in moral judgment. This is a difference of degree rather than of kind, and it is not something predicted by the Any Sacrifice Hypothesis. Nevertheless, we feel it is an interesting hypothesis in its own right. Study 6 tests this graded sacrifice hypothesis.

The second limitation to concluding that moral judgments influence intention attribution is that Study 5 was underpowered for an effect of this small size between majority proportions. A post-hoc analysis showed that the study achieved 63.5% power. A replication with increased power would increase our confidence in this effect. Study 7 accordingly addresses this reservation.

Study 6

Study 6 tests a modification of the Any Sacrifice Hypothesis. This graded sacrifice hypothesis maintains that the likelihood that a side-effect is attributed with intention depends on the degree to which it is costly.

Study 6 uses a vignette that, as with Study 4, aims to minimize morality's role, but, as with Study 5, involves a costly side-effect at both lesser and greater magnitudes. If the difference in intention attribution in Study 5 resulted from the difference in the degree of cost alone, then we should expect Study 6 to show a difference in intention attribution across two variations owing to the different degree of cost involved in each. At the same time, if that difference in Study 5 resulted from a difference in moral judgment, or if it was due to chance alone, then we should expect the likelihood of intention attribution in Study 6 to be similar across the two variations.

Method

Four hundred and four American participants (55.7% female, M_{age} = 37.6 years, SD = 13.0) recruited on Mechanical Turk participated in a study identical to Study 4, with the

exception that we replaced the vignette and question. We increased the sample size based on the possibility that there is a smaller impact on intention attribution that results from greater and lesser degrees of cost. We based our power analysis on the proportions of intention attribution that we observed earlier in Study 5, rounding each in the direction of a smaller effect. To detect a difference between 0.65 and 0.78 with 80% power, we required a sample size of at least 378 participants. We aimed for 400.

The vignette read as follows:

> Tim's doctor explained to him, "While you recover, you have the option to take this medication. It will treat your current symptoms, but it will also give you a mild [severe] headache for a day or two."
>
> Tim responded, "I don't care at all about having a mild [severe] headache, I just want to treat my current symptoms as well as I can. Let's go ahead with this medication."
>
> Tim took the medication and sure enough, he had a mild [severe] headache.

Tim therefore accepted the treatment of his symptoms, along with either a lower or a higher cost.

After reading the vignette, participants responded to the following question: "Did Tim give himself a headache intentionally?"

Results

Among the participants who read the variation wherein Tim's medication resulted in a mild headache, 58.0% responded with, "Yes", to the question of whether Tim gave himself a headache intentionally. Among those who read the variation in which Tim had a severe headache, a similar proportion, 61.3%, responded with, "Yes", to the same question ($X^2(1) = 0.45$, $p = .502$, $\varphi = .033$). Also, a majority responded with, "Yes," in both the mild-headache variation ($X^2(1) = 5.14$, $p = .023$, $\varphi = .113$) and the severe-headache variation($X^2(1) = 10.46$, $p = .001$, $\varphi = .161$).

Discussion

Study 6's results do not support the graded sacrifice hypothesis, which maintains that the likelihood of intention attribution with regard to some costly outcome is impacted by the degree of that cost. Considering this lack of support, the difference in intention attribution observed in Study 5 is better explained either as resulting from moral judgment or from chance.

Study 7

Although the results of Study 5 provided stronger support for the Any Sacrifice Hypothesis than for the Sufficient Commitment Hypothesis, there also appeared to be a small potentially additive effect of moral judgment on intention attribution. However, Study 5 was underpowered for an effect of that size. Study 7 attempts to replicate that smaller effect with sufficient statistical power.

Method

Four hundred and one American participants (57.1% female, M_{age} = 37.3 years, SD = 13.1) recruited on Mechanical Turk participated in a study identical to Study 5. As with Study 6, we based a power analysis on the proportions of intention attribution observed in Study 5's results, and rounded in the direction of a smaller effect. To detect a difference between 0.65 and 0.78 with 80% power, we required at least 378 participants. We aimed for 400.

Results

Among the participants who read the morally acceptable variation, a majority, 69.8% ($X^2(1) = 32.24$, $p < .0001$, $\varphi = .284$), responded with, "Yes", to the question of intentionality. Similarly, among those who read the morally unacceptable variation, a majority, 70.3%

($X^2(1) = 34.27$, $p < .0001$, $\varphi = .292$), responded with, "Yes." The proportions of participants who attributed intention in the morally acceptable and morally unacceptable variations were substantially similar ($X^2(1) = 0.01$, $p < .922$, $\varphi = .005$).

The manipulation check again (as in Study 5) validated the independent variable. Among participants who read the morally acceptable variation, 74.9% responded with, "Yes", to the question of whether Tim's behavior was morally acceptable. In contrast, only 32.2% of those who read the morally unacceptable variation responded with, "Yes" ($X^2(1) = 75.93$, $p < .0001$, $\varphi = .435$).

Discussion

Study 7 strongly supports the Any Sacrifice Hypothesis. The moral difference between the variations of the vignette appeared to have no impact on the likelihood of intention attribution. In addition, and consistent with Study 6's results, there was no apparent impact of the magnitude of the cost. Consistent with the Any Sacrifice Hypothesis, Study 7 suggests that people are likely to regard any sacrifice as intentional, regardless of its magnitude or how one may morally judge it.

General discussion

The basic interpretation, and a related more recent set of ideas that we call the Sufficient Commitment Hypothesis, both maintain that the side-effect effect is a function of moral judgment. This means that intention attribution—upon which most consider moral judgment to be based—is actually downstream from moral judgment. In a series of studies, we tested the predictions of these explanatory schemes against those of what we call the Any Sacrifice Hypothesis. The Any Sacrifice Hypothesis maintains that the side-effect effect is a function of an asymmetry between costly and beneficial side-effects, and not a function of

moral judgment. Among the alternatives studied here, not only is this hypothesis the one that our data best support, but it also seems most consistent with conventional thought. Specifically, the hypothesis does not imply that the ordinary understanding of the relationship between intention and moral judgment is backward. While this implication of the leading hypotheses is a fascinating prospect, it alone justifies skepticism. As a result of our findings, we suggest that the side-effect effect is not a function of moral judgment, and therefore, although illuminating, not quite as revisionary as is believed by prominent contributors to this discourse.

Although our data support the Any Sacrifice Hypothesis—according to which any foreseen cost will likely be treated as a sacrifice in the sense that it is attributed with intention—it is not clear, morally speaking, that any cost ought to be treated as a sacrifice, or in turn, treated as intentional. It is also unclear whether doing so yields accurate inferences about actors' intentions as the actors themselves understand them.

Readers may believe there is an inconsistency between the results of Study 1 and those of Study 6. In Study 1, compared to a lesser cost, a greater cost appears to increase the likelihood of intention attribution as it serves as greater proof of the actor's commitment to their goal (e.g., Dik & Aarts, 2007). Alternatively, in Study 6, we show that the magnitude of the cost does not impact the likelihood of intention attribution. The difference is that, where cost increases intention attribution in Study 1, it regards intention in relation to the goal, where a greater cost is better proof of commitment to that goal. In Study 6, however, the intention in question relates to the cost itself.

Even if the Any Sacrifice Hypothesis is correct—or if, more generally, it is true that the side-effect effect is not a function of moral judgment in the sense proposed and examined

here—some moral judgments may still precede intention attribution. For example, to classify harm to the environment as a cost is itself a moral judgment of sorts. However, this is not an interpersonal moral judgment, so it is not the kind most consider to be based on intention.

Future research could further explore the implications of the Any Sacrifice Hypothesis and test its predictions against the responses of other populations, and to the extent possible, leverage a different experimental paradigm.

References

Dik, G., & Aarts, H. (2007). Behavioral cues to others' motivation and goal pursuits: The perception of effort facilitates goal inference and contagion. Journal of Experimental Social Psychology, 43(5), 727-737.

Egré, P. (2010). Qualitative judgments, quantitative judgments, and norm-sensitivity. Behavioral and Brain Sciences, 33(4), 335-336.

Foot, P. (1967). The problem of abortion and the doctrine of double effect.

Gintis, H. (2010). Modalities of word usage in intentionality and causality. Behavioral and Brain Sciences, 33(4), 336-337.

Girotto, V., Surian, L., & Siegal, M. (2010). Morals, beliefs, and counterfactuals. Behavioral and Brain Sciences, 33(4), 337-338.

Knobe, J. (2003). Intentional action and side effects in ordinary language. Analysis, 63(3), 190-194.

Knobe, J. (2010). Person as scientist, person as moralist. Behavioral and Brain Sciences, 33(4), 315-329.

Knobe, J., & Doris, J. M. (2010). Responsibility.

Karniol, R. (1978). Children's use of intention cues in evaluating behavior. Psychological Bulletin, 85(1), 76.

Laurent, S. M., Reich, B. J., & Skorinko, J. L. (2019). Reconstructing the side-effect effect: A new way of understanding how moral considerations drive intentionality asymmetries. Journal of Experimental Psychology: General.

Machery, E. 2008: The folk concept of intentional action: philosophical and experimental issues. Mind & Language, 23, 165– 189.

Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology, 33*, 101-121.

Menzies, P. (2010). Norms, causes, and alternative possibilities. Behavioral and Brain Sciences, 33(4), 346-347.

Mallon, R. (2008). Knobe versus Machery: testing the trade- off hypothesis. Mind & Language, 23(2), 247-255.

Nanay, B. (2010). Morality or modality? What does the attribution of intentionality depend on?. Canadian Journal of Philosophy, 40(1), 25-39.

Sauer, H. (2017). Between Facts and Norms: Ethics and Empirical Moral Psychology. In Moral Psychology (pp. 5-27). Springer, Cham.

CHAPTER 2

DO BEHAVIORAL BASE RATES INFLUENCE ASSOCIATED MORAL JUDGMENTS?

Carl P. Jago

Abstract

In a series of studies, we ask whether and to what extent the base rate of a behavior influences associated moral judgment. Existing moral cognition and social influence theory is inconclusive. On the basis of recent experiments, some researchers have argued that base rates are of critical importance to moral judgment. However, in multiple studies, we find that, when properly isolated, base rates do indeed influence moral judgment, but they do so with only very small effect sizes. We reason that previous research purporting to find more substantial effects of base-rate information on moral judgment may have conflated that which is descriptively typical (i.e., the base-rate) with that which is typically endorsed (i.e., the injunctive norm). An additional experiment supports this reasoning. In another experiment, we test the possibility that the limited influence of base rates on moral judgment could be a result of the fundamental attribution error, which is not specific to moral judgment. The results suggest that moral judgment may be uniquely resilient to the influence of base rates. Finally, we reconcile our findings with well-established knowledge regarding the many powerful impacts of social-influence. We test the hypothesis that it is not simply what is done by others, even many others, that matters much to moral judgment, but instead what is done by those of a more morally relevant reference group. The results suggest that the extent to which the behavior of others influences our moral judgments may depend on our perception of those others.

DO BEHAVIORAL BASE RATES INFLUENCE ASSOCIATED MORAL JUDGMENTS?

Might each new #metoo revelation simultaneously raise awareness about the frequency of sexual misconduct and, in doing so, somehow also dilute the perceived severity of each instance? Do we think the badly behaving teenager is making a morally relevant point when she informs us that 'everybody is doing it'? Descriptively, it is not obvious whether we make moral judgments based on comparisons to statistical typicality, or whether we can think everybody is bad if everybody behaves badly. It seems self-evident that there is some relationship between what most people do and what most people find acceptable, but the details of this relationship are unclear. In this project, we ask whether and to what extent the base rate of a behavior influences moral judgments about that behavior and those who engage in it.

Theory predicting that the base rate of a potentially objectionable behavior influences moral judgments can be divided into two groups. The first stems from theories of moral cognition, while the second stems from social psychology. Moral cognition theory suggests that causality, control, and harmfulness are key variables affecting moral judgment. Actor causality plays a key role in moral judgments (e.g., Lagnado & Channon, 2008, Scanlon, 2008;) and prominent models of blame and punishment include it as a primary factor (e.g., Malle, Gugliemo, & Monroe, 2014; Sloman, Fernbach, & Ewing, 2009). Kelley's (1973) covariation model of causal attribution maintains that when an actor's behavior in response to the same stimulus is similar to that of other people, causality is less likely to be attributed to the actor than if the actor's behavior differed. As a result, we might expect moral judgments to be less harsh against an actor engaging in a behavior that has a high base rate.

Actor control refers to the freedom to behave as desired (Berofsky, 1966), and figures prominently in descriptive models of moral culpability (e.g., Alicke, 2000; Weiner, 1995). Greater perceived control typically leads to harsher moral judgments, whereas constraints on freedom, including external forces and limited behavioral options, tend to mitigate blame (Alicke, 2000; Nichols & Knobe, 2007). Evidence (e.g., Nahmias, Morris, Nadelhoffer, & Turner, 2005) further suggests that, more generally, a belief in free will makes an observer more punitive in their judgments (Clarke et al. 2014). Therefore, to the extent that everybody behaves similarly in some situation (a high base rate of the behavior) an observer might infer a lower degree of control and, accordingly, limit the severity of their judgment. Harmfulness is a third key variable in moral cognition theory understood to affect moral judgment: degrees of blame and punishment are positively correlated with the harmfulness of actions and their consequences (Cushman, 2008; Schein & Grey, 2018). This may occur even when the harm is unforeseen and unintended (Young, Nichols, & Saxe, 2010). An observer might infer, at least in unfamiliar or ambiguous contexts, that some act is less harmful if a greater number of people engage in it, either because they assume that few people will perform a more harmful action, or because they perceive a declining marginal harm (e.g., the perception that one additional piece of litter makes little difference to an already littered field).

The second group of theory supporting the hypothesis that behavioral base rates will influence moral judgments pertains to social influence. As social animals, human beings have a remarkable tendency to imitate one another and to adopt each other's opinions. Social psychology is replete with demonstrations of various kinds of apparent herd mentality, where individuals seem to assume that if many are doing something, it must be reasonable

(Milgram, Bickman, & Berkowitz, 1969). Moreover, the idea that social norms play an important role in morality is an old one (McArthur, 1997). It is used in discussions from understanding cultural shifts over time, to concerns that media depictions of undesirable behaviors could corrupt an audience. The Bobo doll experiments (Bandura, Ross, & Ross, 1961) famously showed children imitating the aggressive behaviors of adult models they had merely observed. In another study, children were found to protest actions differing from those they had previously observed (Schmidt, Rakoczy, & Tomasello, 2011), seemingly inferring that what was previously done is also what ought to be done. In adults, socially-desirable behaviors have been effectively encouraged by suggesting that a high proportion of others behave that way (Goldstein, Griskevicius, & Cialdini, 2007). Adults have also been shown to conform to the majority opinion of peer groups, even when those groups provide otherwise counter-normative moral judgments (Cornwell, Jago, & Higgins, 2019). Finally, evidence suggests that people morally disapprove of behaviors that they deem "weird," even when no specific harmful consequences are evident (Gray, & Keeney, 2015). To the extent that atypical behaviors are viewed as abnormal, and in turn undesirable, we might expect low base rate behaviors to be judged more critically.

Recently, more direct empirical evidence for the role of behavioral base rates in moral judgment was published in a paper entitled, "It's not what you do, but what everyone else does," in which the authors argued that, "moral actions are evaluated in relation to what most other people do" (Monroe, Dillon, Guglielmo, & Baumeister, 2018). In their experiments, participants provided more positive evaluations of actors whose restaurant tips exceeded the tipping norm (described as the amount most people give), and less positive evaluations of actors whose tips fell below the norm. Most pertinently, when they recruited participants in

both the U.S. and the U.K., where tipping norms differ (U.S. norm: 15%; U.K. norm: 10%), judgments of various tips were independent of the absolute magnitude of the tip, and instead relative to the local norm. The authors take this as evidence that participants base appropriate action on what, in each country, is typically done.

Without doubt, at least for extremely bad behaviors, moral badness tends to be correlated with low base rates. Less clear, is the causal relationship. While we can hypothesize that some behavior is particularly bad because particularly few people do it, we can just as readily hypothesize that particularly few people perform the behavior because it is particularly bad. In the latter hypothesis, the moral status and the base rate are still associated, but with the reversed causal direction. Thus, the badness of the act is not a result of the base rate, but instead comes from some other factor. Yet the theory discussed above presumes causality in the first direction (the behavior is bad because few people do it). Following, we call into question the above reasons to hypothesize a moral influence of base-rates and highlight other variables potentially related to or conflated with behavioral base rates.

Although indirect effects of base rates on moral judgments may be mediated by effects on key variables of moral cognition (e.g., cause, control, harm), whether base rates have such intermediate effects is an empirical question without a clear answer. To begin with, tests of Kelley's (1973) theory of causal attribution suggest that, of the three variables in the model, base rate had the least predictive power (McArthur, 1972), perhaps because the base rate itself depends on numerous other factors (Kassin, 1979). In addition, the fundamental attribution error—the tendency in causal attribution to underestimate situational factors and emphasize dispositional ones (Ross, 1977)—may undermine base rate effects,

especially if base rate is not perceived as a salient situational factor. Finally, effects mediated by harmfulness rely first on effects of base rates on harmfulness, which to our knowledge, has no existing empirical support. Thus, this line of inquiry leads to another reasonable but equally unanswered question: do base-rates impact perceptions of harmfulness?

Norms are widely believed to influence both moral decisions and judgments (e.g., Alicke, Rose, & Bloom, 2011; Malle et al., 2014; Nichols, 2002), however the nature of norms is a controversial topic (McAdams, 1997). Logical and theoretical reasons call in doubt whether a behavior's base rate, per se, establishes the norm by which it is judged. For example, while some atypical behaviors are indeed regarded as "weird" and are often met with criticism, atypicality itself is a feature not only of many bad behaviors, but also of many good and praiseworthy behaviors, such as a particularly large personal sacrifice made in the service of others. Thus, for any atypical behavior, there remains a question of valence. If the base rate itself cannot determine whether a behavior is good or bad, then base rate alone is not enough to establish a moral norm.

In addition, the concept of a "norm" can be understood in multiple ways (Schaffer, 1983; Cialdini, Kallgren, & Reno, 1991). Cialdini and colleagues (1991) drew attention to the distinction between social norms of the descriptive kind (i.e., the perception of how most others would behave) and social norms of the injunctive kind (i.e., the perception of the kind of behavior of which most people would approve/disapprove). Importantly, in this distinction, base rates relate directly to descriptive norms rather than injunctive norms. This distinction does more than to merely allow us to restate the original question from: Do base rates influence moral judgments? to: Do descriptive norms influence injunctive norms? It also highlights potential confounds present in previous research. For example, Monroe and

colleagues (2018) seem to have conflated descriptive norms with injunctive norms. Their study used tipping standards as descriptive norms, yet such norms are actually injunctive; they are associated with well-understood expectations and are considered part of a social contract to which people feel a sense of duty (Lynn, 2008; Lynn & Starbuck, 2015). Tipping norms are not mere descriptions of what people do but are regarded as prescribing what ought to be done. Therefore, while Monroe and colleagues (2018) conclude that "moral actions are evaluated in relation to what most other people do," they may instead have found that moral actions are evaluated in relation to what most other people endorse. Similarly, in experiments where participants are found to conform to a group's moral judgments (Cornwell et al., 2019), the opinion of the group is a matter of endorsement. Thus, to the extent that it establishes a norm, that norm is injunctive rather than descriptive. In addition, the extent to which descriptive norms influence injunctive norms remains unknown; perhaps a high base rate (a descriptive norm) effectively implies an injunctive norm. On the one hand, by comparison to other psychological phenomena, which are predicted on people's ability to hold an opinion that they believe differs from that of the majority, one can imagine that descriptive norms and injunctive norms are largely independent. These phenomena include pluralistic ignorance (Miller & McFarland, 1987) and group think (Esser, 1998). On the other hand, evidence suggests that that which is considered good appears to influence that which is considered typical (Bear & Knobe, 2017). Missing is evidence regarding an influence of what is typical on what is good.

Finally, if the behavior of others does influence judgments of right and wrong, the degree of influence likely varies depending on who the 'others' are. For example, children are differentially selective in who and what they imitate (Grusec & Abramovitch, 1982).

More generally, social comparison theory (Goethals & Darley, 1987) purports that people typically estimate their competencies and judge themselves, not in relation to everybody, but in relation to particular groups. Merton, who coined the term, role model, hypothesized that, "individuals compare themselves with reference groups of people who occupy the social role to which the individuals aspire" (Holton, 2004). The imitation of others' morality, therefore, may be limited to those others who are perceived as a relevant reference. Consider again the efforts to censor depictions of undesirable behaviors in the media; most stories involve villains, even children's stories. Do good characters have to outnumber bad characters to tip the normative balance? With this in mind, it seems more likely that, when it comes to influencing what is thought to be acceptable, it may not be so much what is done by anybody, or everybody, that matters, but instead what is done by those who are perceived as good. Under this hypothesis, it is still the case that what others do matters, but the extent to which it matters depends on the perception of those people.

In the current paper, we use a series of studies to investigate moral judgments made by individuals while manipulating behavioral base rate (descriptive norm), approval/disapproval rate (injunctive norm), and reference group. We describe novel scenarios involving unfamiliar norms and behaviors for which the degrees of harmfulness and other details about outcomes are uncertain. We then collect participants' judgments of blame, wrongness, moral character, and harmfulness of the morally questionable behavior described in the scenario.

General Methods

For all studies, Americans were recruited on Amazon's Mechanical Turk platform and provided their age and gender. Each participant read a vignette (specific to the study); each study had two or more versions of the vignette that varied according to the level of the effect being tested (most often, the base rate of a behavior) and participants were assigned at random to effect level.

Following each vignette, participants answered three questions using an 11-point slider scale ranging from 0 ("not at all") to 10 ("extremely") that measured blame attribution ("How blameworthy is Tim?"), rating of act wrongness ("How wrong was his behavior?"), and rating of moral character ("How bad a person is Tim?"). All participants read only one vignette and questions were presented one at a time in random order. The difference in scores for each question between high and low base rate scenarios was tested with a two-tailed t-test.

Study 1a

Methods

Participants (N = 260) were 54% female with a mean age = 36.7 ± 11.4 years. Each participant read one of two vignettes, picked at random, and then provided three moral judgments, presented one at a time, in a randomized order.

The vignette described an unfamiliar behavior whose base rate was manipulated with two levels (low and high) and read as follows (underlining highlights the manipulated variable; brackets contain the alternative level of the manipulated variable):

> Because he is feeling tired, a worker at a water management facility chooses to skip doing the 12-point safety inspection of his post before taking his lunch break. No

anomalies have ever been detected, and there are other fail-safes in place to catch any problems. Less than 10% [More than 90%] of other water management workers have skipped this part of their duties.

The questions given in the general methods above were modified appropriately, replacing "Tim" with "the worker."

Results

All results are presented as (mean ± standard error) point score for each response, followed by *t*-statistic and effect size. Blame attributions for the low base rate (8.021 ± 0.187) and high base rate (7.947 ± 0.212) conditions were similar ($t(258) = 0.259$, $p = .796$, $d = 0.032$). Likewise, wrongness ratings for the low base rate (7.589 ± 0.197) and high base rate (7.202 ± 0.223) conditions were similar ($t(258) = 1.303$, $p = .194$, $d = 0.171$). Finally, character judgments for the low base rate (5.753 ± 0.221) and high base rate (5.140 ± 0.250) conditions differed marginally ($t(258) = 1.840$, $p = .067$, $d = 0.271$).

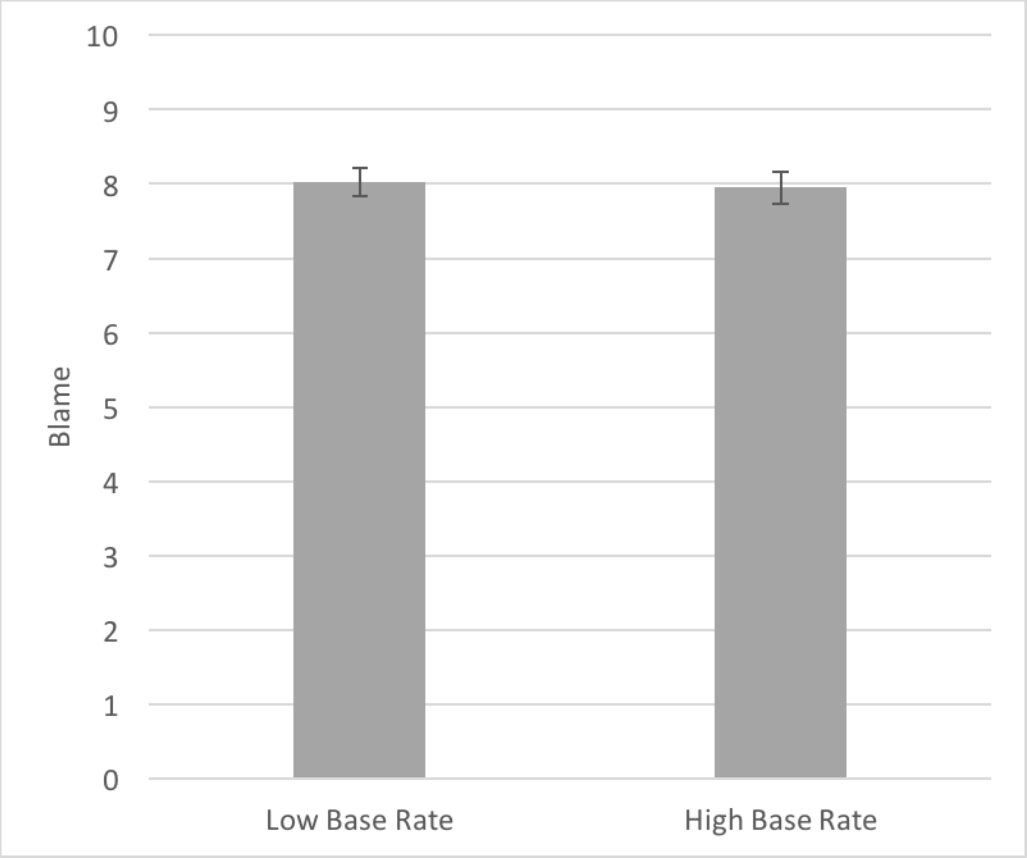

Figure 2.1a.a. Mean blame attributed for low versus high base rate conditions in Study 1a. Error bars show standard error of the mean.

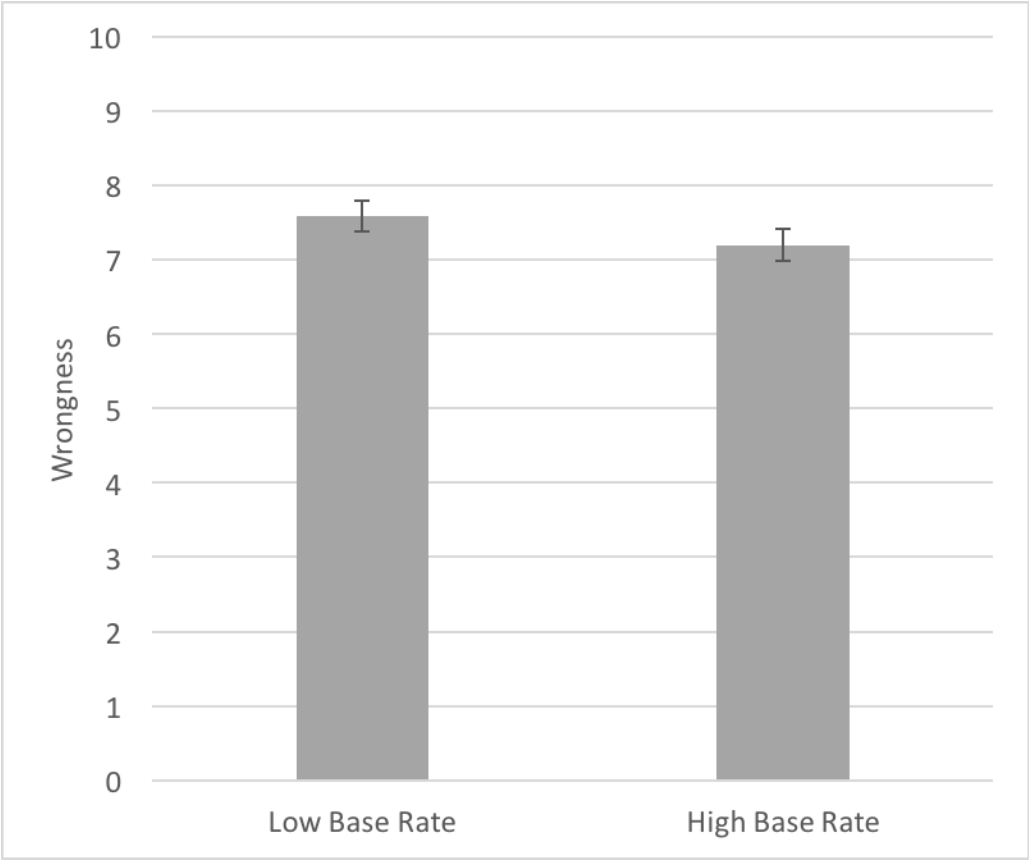

Figure 2.1a.b. Mean wrongness ratings for low versus high base rate conditions in Study 1a. Error bars show standard error of the mean.

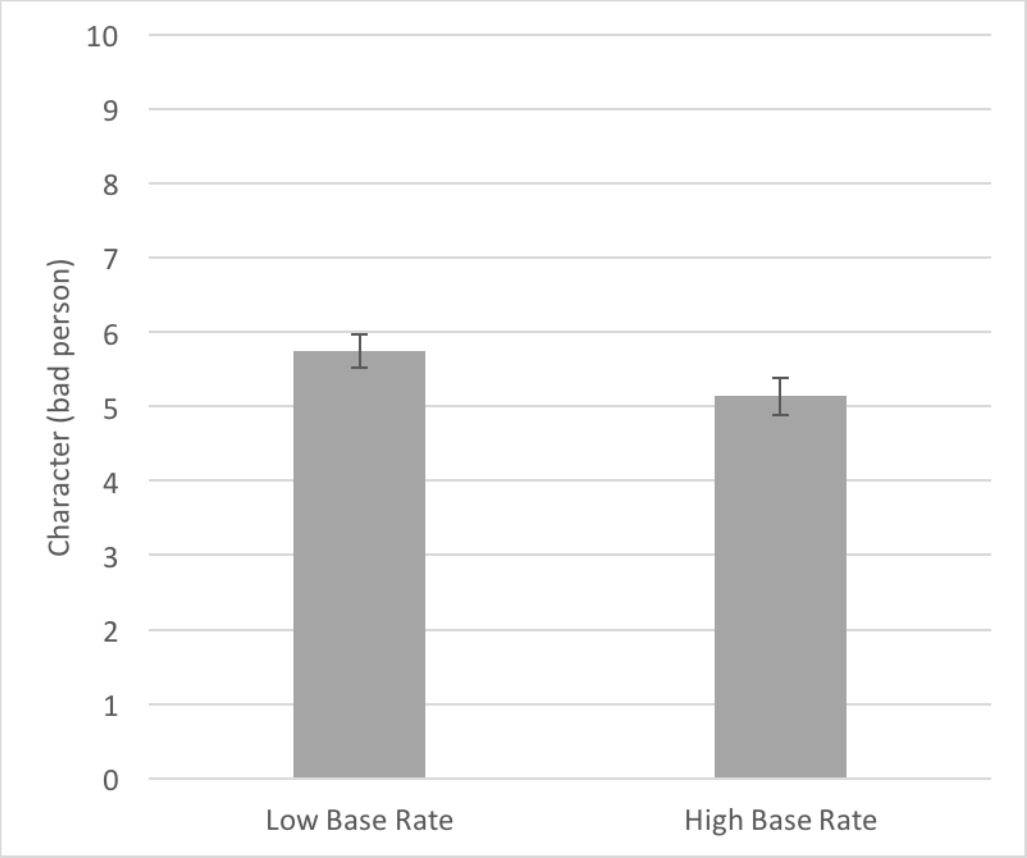

Figure 2.1a.c. Mean character ratings for low versus high base rate conditions in Study 1a. Error bars show standard error of the mean.

Study 1b

Given the arguably surprising lack of a base rate effect in Study 1a, we next tested a different vignette that aimed to involve a more general population of actors along with a less high-stakes scenario and fewer details. In addition, a different pair of base rate levels were adopted in case there was something unique about the pair used in Study 1a.

Methods

Participants (N = 263) were 59% female with a mean age of 37.8 ± 12.2 years.

The new vignette included a base rate manipulation and read as follows:

Every year, thousands of people visit the Rainbow Stream at Yosemite National Park.

One of the stream's most notable features is its shiny pebbles, and visitors are asked to not remove any pebbles from the site. Less than 5% [More than 80%] of visitors take a pebble from the site.

When visiting, Tim takes a pebble for himself.

Results

As in Study 1a, blame attributions for the low base rate (7.591 ± 0.242) and high base rate (7.122 ± 0.243) conditions were similar ($t(261) = 1.365$, $p = .173$, $d = 0.168$). Wrongness ratings for the low base rate (6.121 ± 0.251) and high base rate (6.260 ± 0.252) conditions were also similar ($t(261) = 0.388$, $p = .698$, $d = 0.050$). Character judgments for the low base rate (4.447 ± 0.218) and high base rate (4.176 ± 0.219) conditions were also similar ($t(261) = 0.877$, $p = .381$, $d = 0.097$).

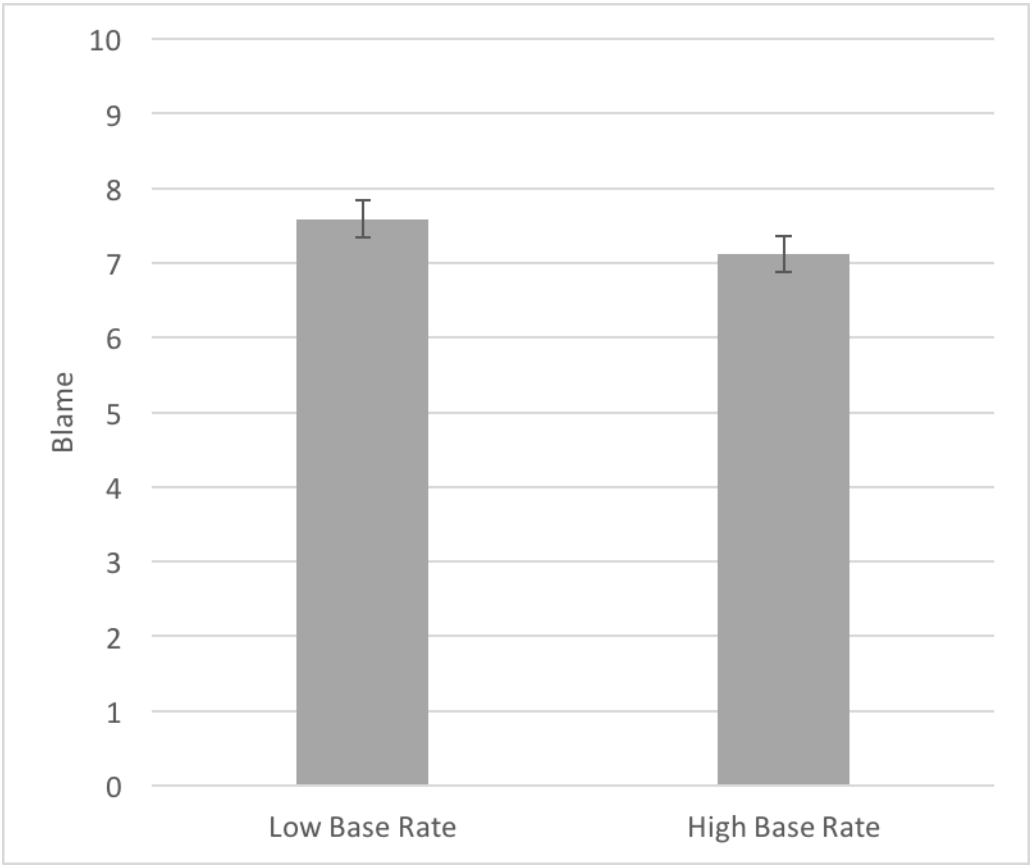

Figure 2.1b.a. Mean blame attributed for low versus high base rate conditions in Study 1b. Error bars show standard error of the mean.

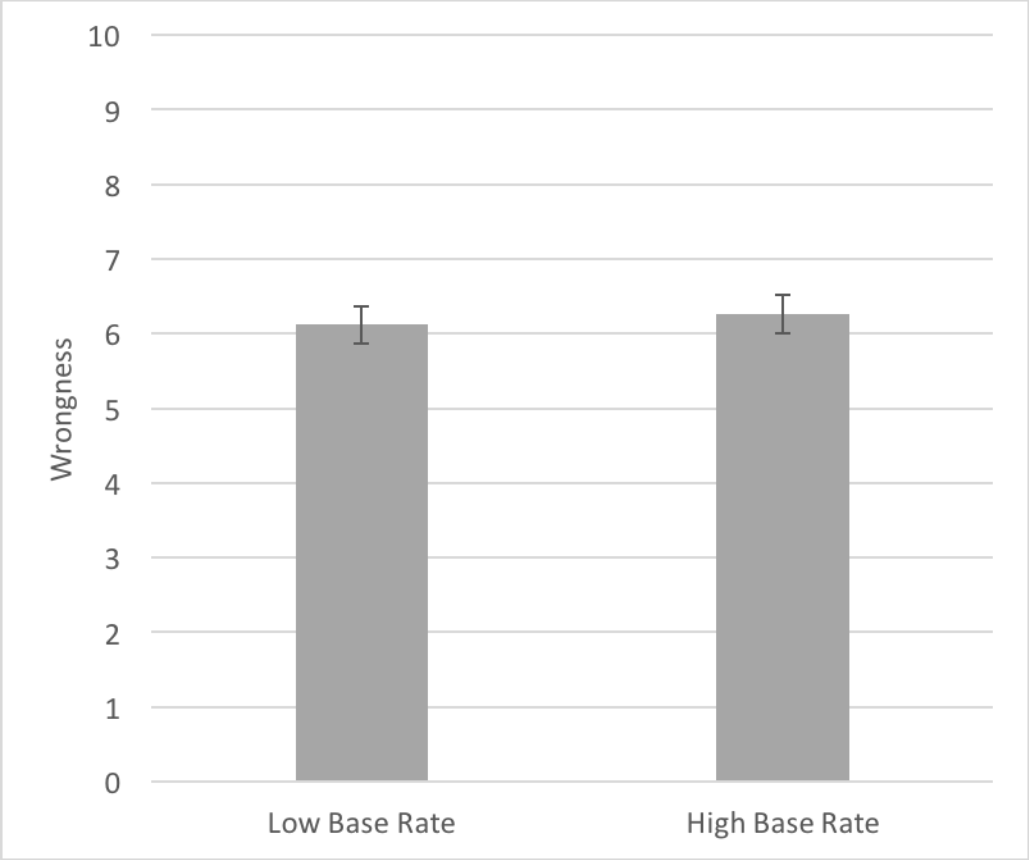

Figure 2.1b.b. Mean wrongness ratings for low versus high base rate conditions in Study 1b. Error bars show standard error of the mean.

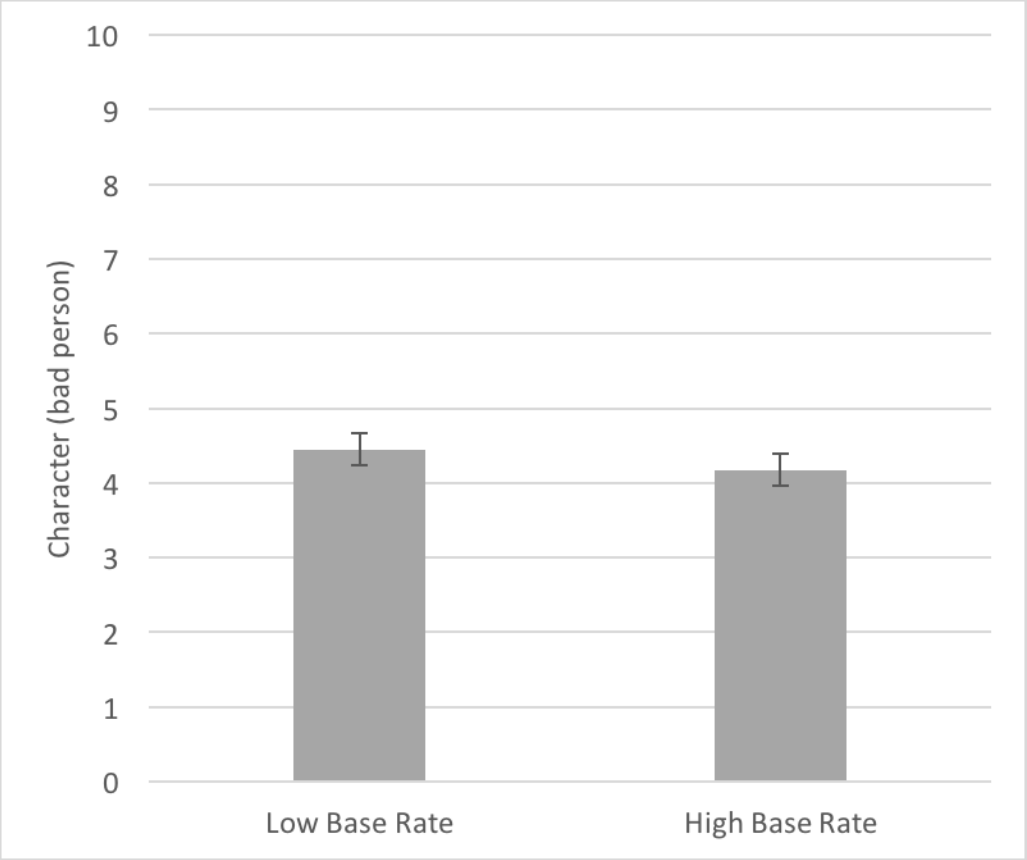

Figure 2.1b.c. Mean character ratings for low versus high base rate conditions in Study 1b. Error bars show standard error of the mean.

Study 1c

Post-hoc reasoning suggested that the low base rate used in study 1b may have indicated low interest in taking a pebble. In an attempt to preclude such a possibility, we modified the vignette from Study 1b to indicate that most visitors would like a pebble. The base rate manipulation was also modified to state the base rate numerically (as before) and non-numerically, in case the former was too abstract for some participants. Finally, we added a measure of perceived harmfulness: "How harmful was Tim's behavior?"

Methods

Participants (N = 262) were 58% female with a mean age of 38.0 ± 12.8 years).

The modified vignette read as follows (critical modifications in italics):

> Every year, park visitors flock to the Rainbow Stream at Yosemite National Park.
>
> One of the stream's most notable features is its shiny pebbles. *Almost all visitors would like to take a pebble when they leave*, but all visitors are asked to not remove any pebbles from the site.
>
> Despite the park's request, *a small number*—less than 5%— [*most visitors*—more than 80%—] do take a pebble with them.
>
> Consider the case of Tim. When visiting, Tim takes a pebble for himself.

The 11-point measure of harmfulness matched the existing measures and asked: "How harmful was Tim's behavior?"

Results

Blame attribution for low base rate (7.046 ± 0.251) and high base rate (6.710 ± 0.241) conditions were similar ($t(260) = 0.965$, $p = .335$, $d = 0.119$). Wrongness ratings for low base rate (5.947 ± 0.238) and high base rate (5.824 ± 0.239) conditions were similar ($t(260) = 0.362$, $p = .718$, $d = 0.043$). Character judgments for low base rate (3.962 ± 0.230) and high base rate (3.969 ± 0.219) conditions were similar ($t(260) = 0.003$, $p = .981$, $d = 0.003$). Harmfulness ratings for the low base rate (3.542 ± 0.243) and high base rate (3.954 ± 0.247) conditions were also similar ($t(260) = 1.192$, $p = .235$, $d = 0.147$).

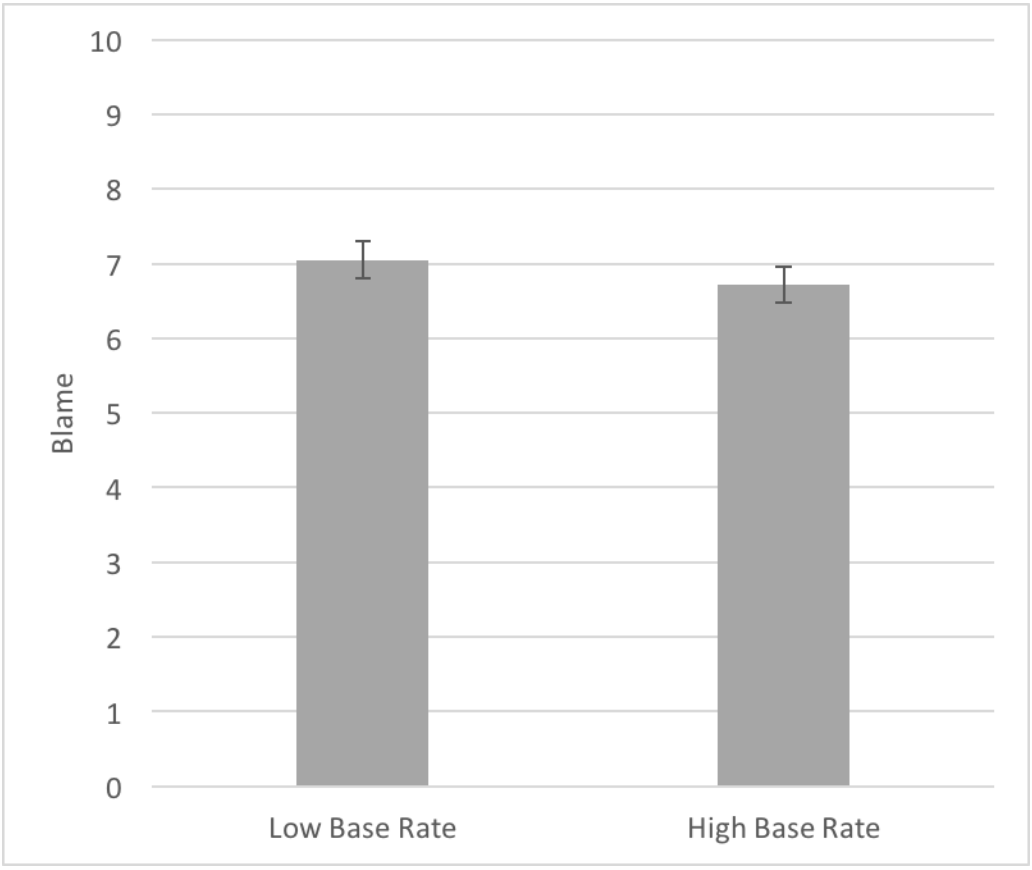

Figure 2.1c.a. Mean blame attributed for low versus high base rate conditions in Study 1c. Error bars show standard error of the mean.

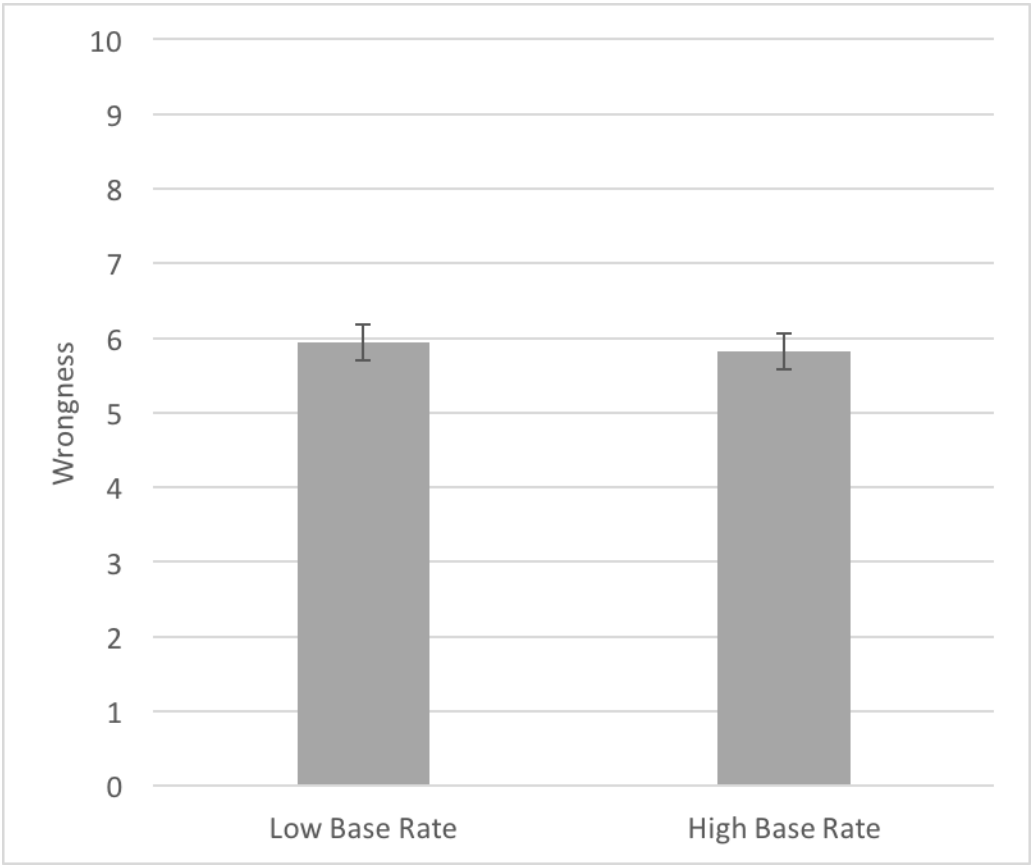

Figure 2.1c.b. Mean wrongness ratings for low versus high base rate conditions in Study 1c. Error bars show standard error of the mean.

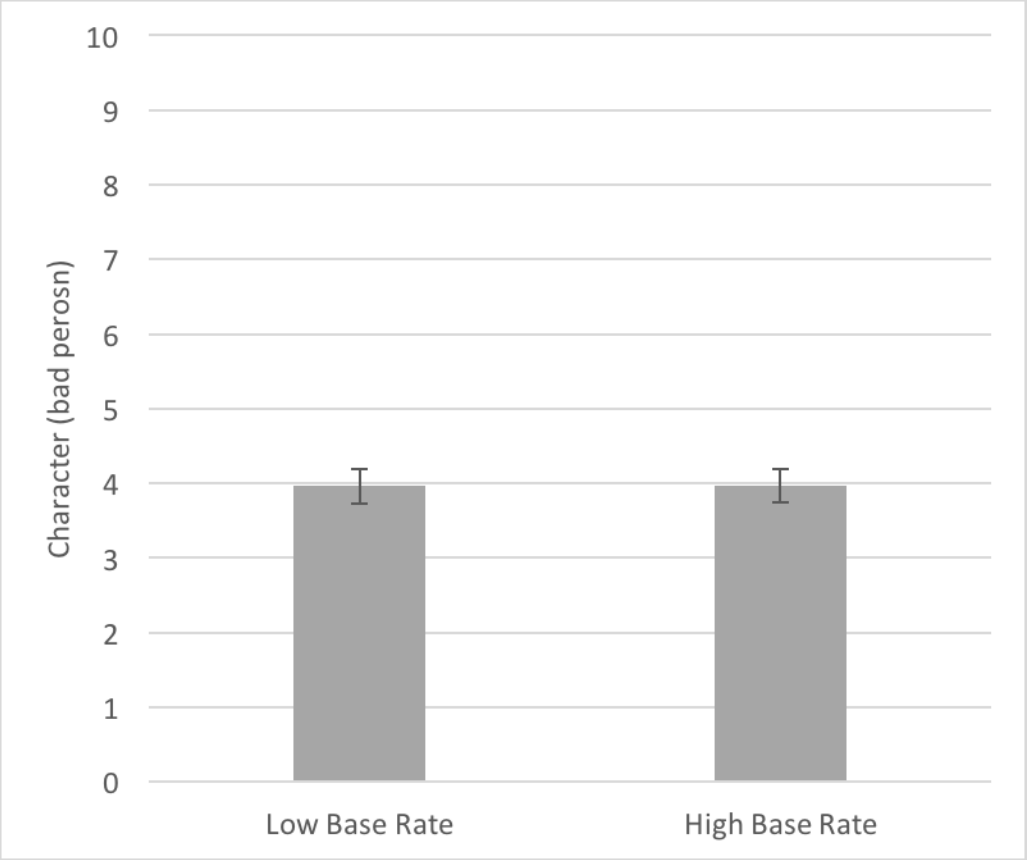

Figure 2.1c.c. Mean character ratings for low versus high base rate conditions in Study 1c. Error bars show standard error of the mean.

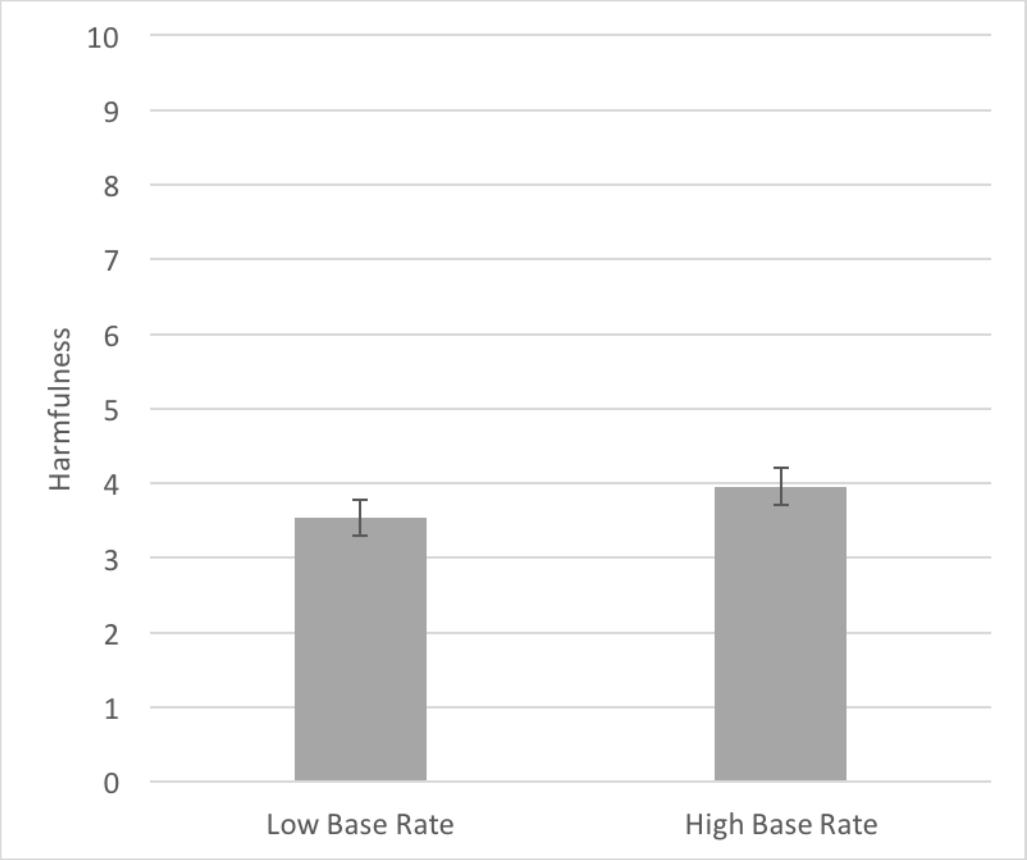

Figure 2.1c.d. Mean harmfulness ratings for low versus high base rate conditions in Study 1c. Error bars show standard error of the mean.

Discussion

With 785 participants, multiple scenarios and variations, studies 1a through 1c failed to find any evidence that behavioral base rates significantly affect moral judgments.

Study 2

In addition to arguing that "moral actions are evaluated in relation to what most other people do," Monroe and colleagues (2018) — who may have inadvertently tested the impacts of norms that were at least partly injunctive rather than purely descriptive — concluded it was important that an actor was described as being aware of the descriptive norm for that norm to influence participants' moral judgments. Thus, in the vignette for Study 2, which

builds on that used in Study 1c, we state that the actor knows the base rate. With such a specification, Study 2 no longer constitutes a straightforward test of the impact of base rates on moral judgments, but instead is a more specific test of whether an actor will knowingly violate a norm.

Methods

Participants (N = 260) were 53% female with a mean age of 37.7 ± 12.4 years.

The modified vignette read as follows:

> Every year, park visitors flock to the Rainbow Stream at Yosemite National Park.
>
> One of the stream's most notable features is its shiny pebbles. Almost all visitors would like to take a pebble when they leave, but all visitors are asked to not remove any pebbles from the site.
>
> Consider the case of Tim. *Tim knows that despite the park's request*, a small number—less than 5%— [most visitors—more than 80%—] do take a pebble with them. When visiting, Tim takes a pebble for himself."

Results

Blame attributions for the low base rate (7.096 ± 0.253) and high base rate (6.863 ± 0.265) conditions were similar ($t(258) = 0.635$, $p = .526$, $d = 0.079$). Wrongness ratings for the low base rate (6.051 ± 0.252) and high base rate (5.734 ± 0.264) conditions were similar ($t(258) = 0.871$, $p = .385$, $d = 0.108$). Character judgments for the low base rate (4.265 ± 0.228) and high base rate (3.734 ± 0.239) conditions were similar ($t(258) = 1.608$, $p = .109$, $d = 0.180$). Finally, harmfulness ratings for the low base rate (4.147 ± 0.256) and high base rate (3.790 ± 0.268) conditions were also similar ($t(258) = 0.961$, $p = .337$, $d = 0.121$).

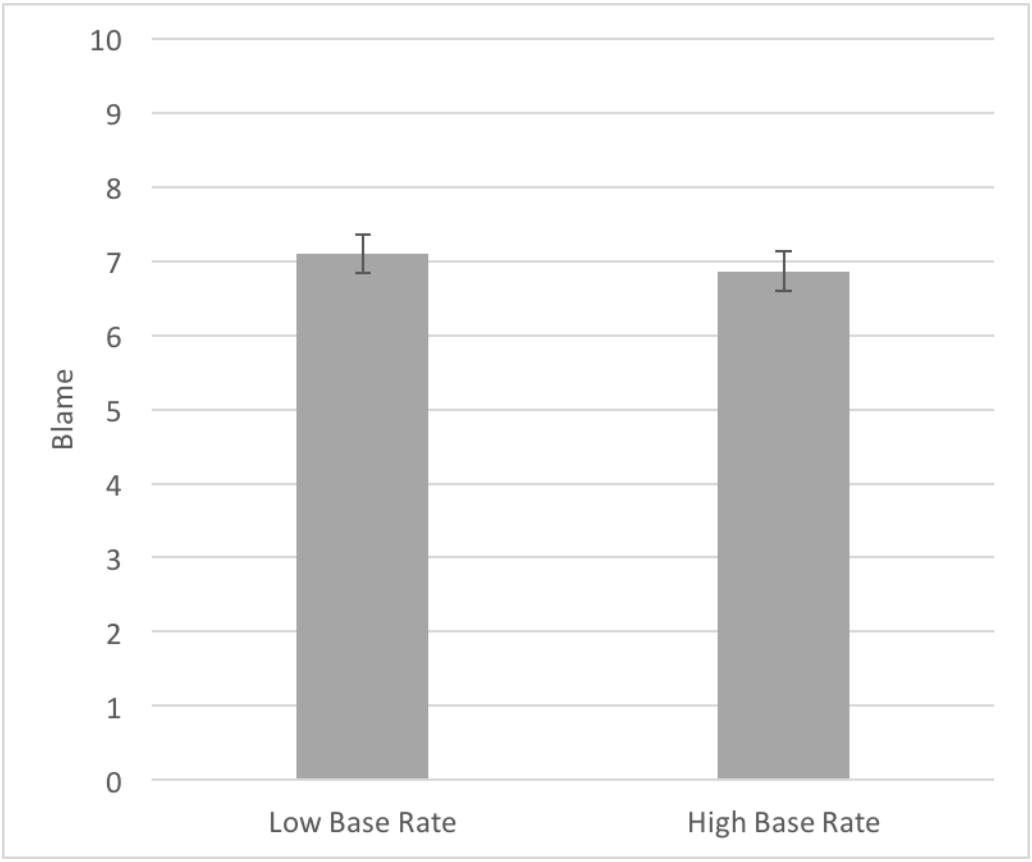

Figure 2.2.a. Mean blame attributed for low versus high base rate conditions in Study 2. Error bars show standard error of the mean.

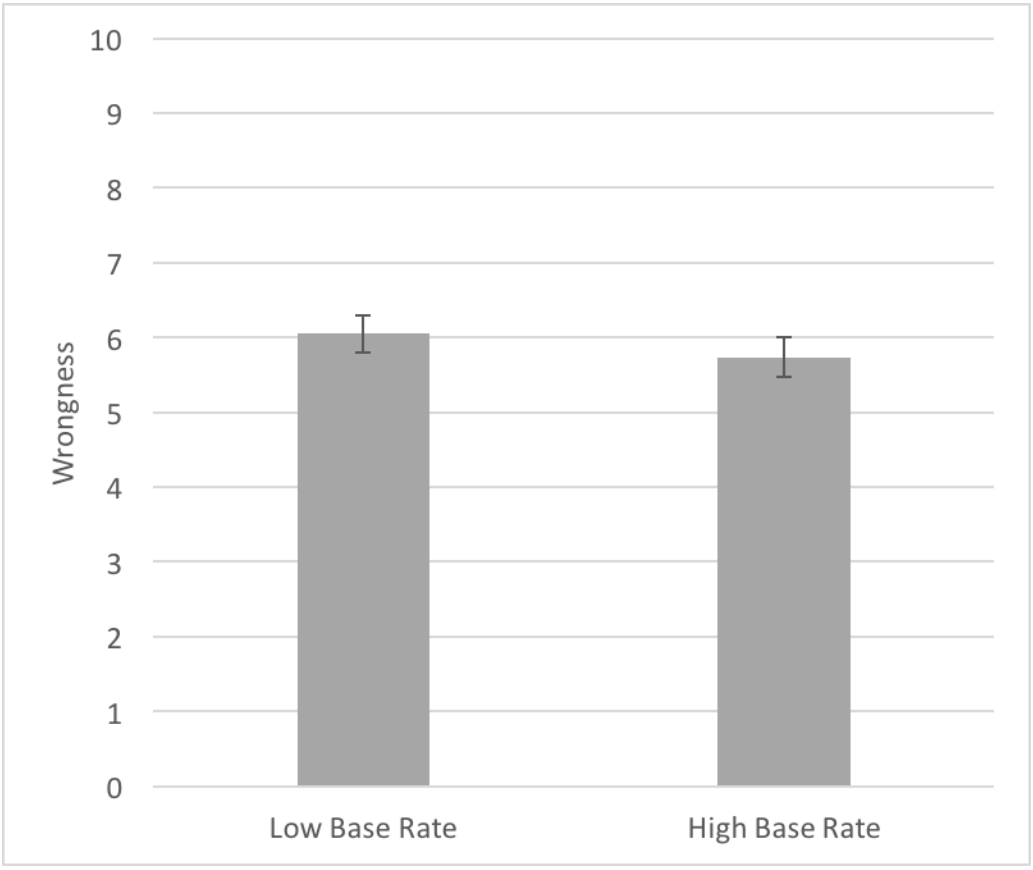

Figure 2.2.b. Mean wrongness ratings for low versus high base rate conditions in Study 2. Error bars show standard error of the mean.

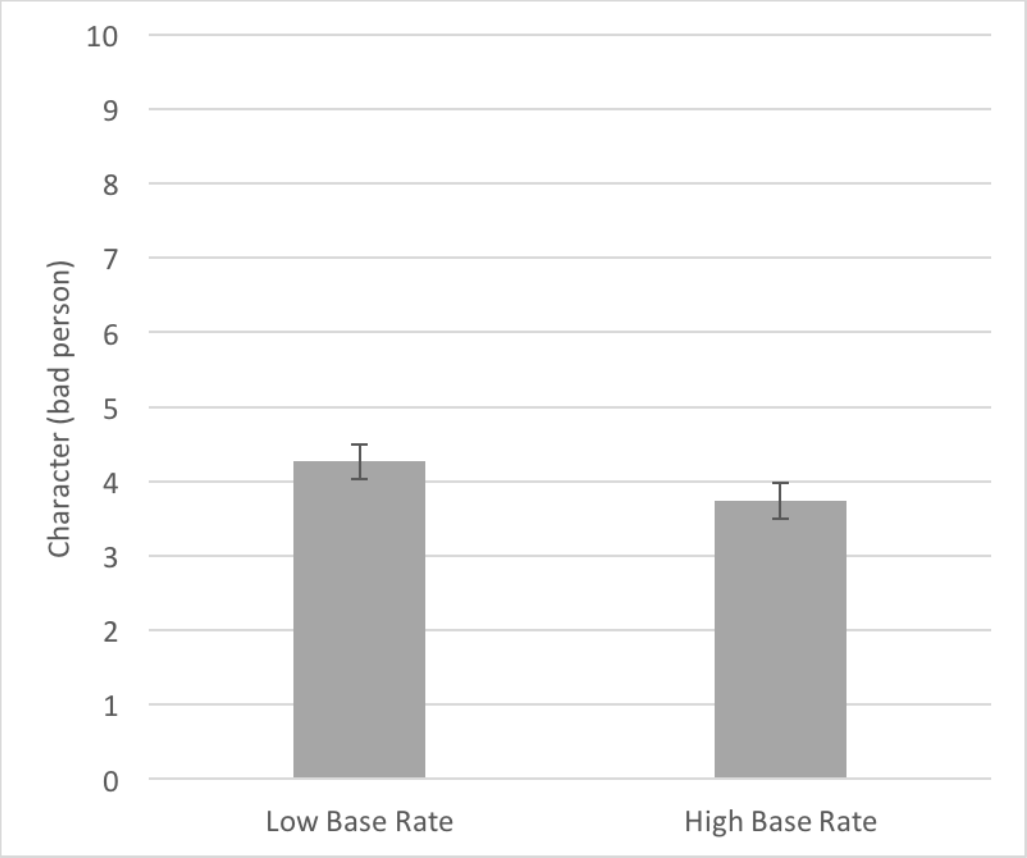

Figure 2.2.c. Mean character ratings for low versus high base rate conditions in Study 2. Error bars show standard error of the mean.

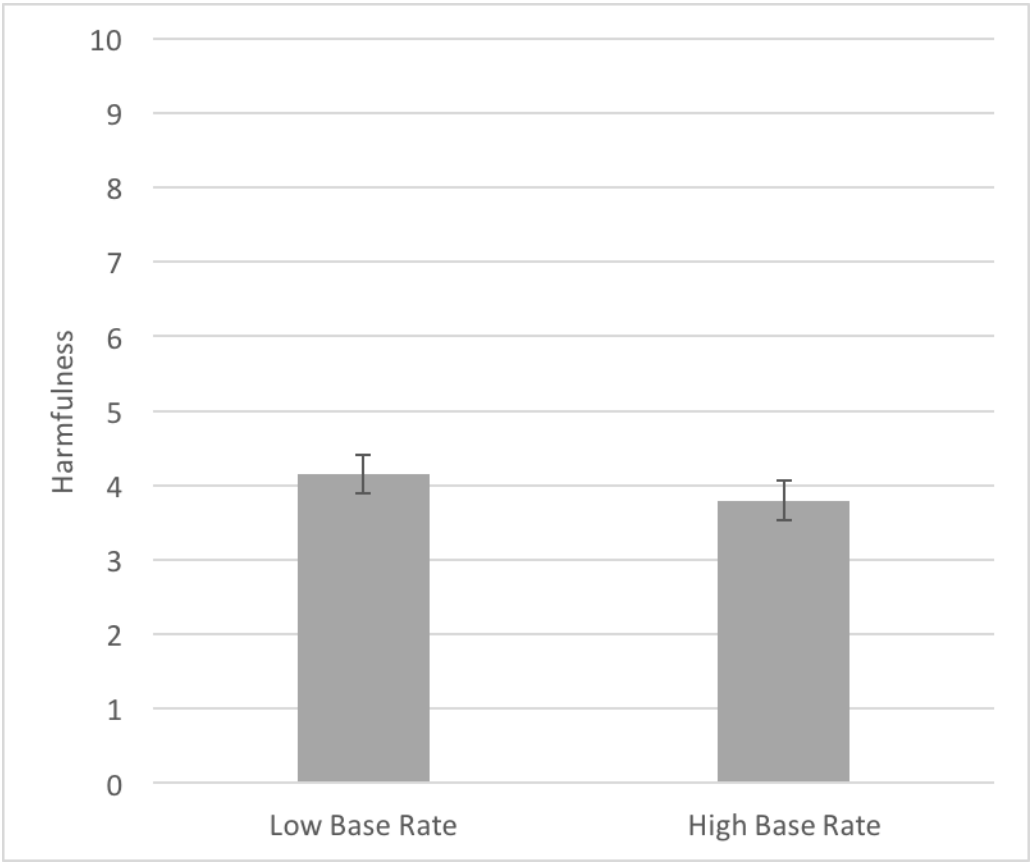

Figure 2.2.d. Mean harmfulness ratings for low versus high base rate conditions in Study 2. Error bars show standard error of the mean.

Discussion

Despite stipulating Tim's awareness of the base rate, the results resembled those of Study 1, with similar ratings and judgments across the high and low base rate conditions. The actor's knowledge, therefore, does not appear to be the critical missing aspect of our tests in Study 1. While actor awareness might be important to moral judgments in cases where the actor violates an injunctive norm, Study 2 provides no evidence that this is the case with base rates.

Study 3

In Study 3, we test the possibility that we have merely rediscovered the fundamental attribution error rather than having learned something more unique to the relationship between base rates and moral judgments. Perhaps participants generally under-emphasize base rate implications about situational explanations for a behavior (and over-emphasize dispositional explanations), including for non-moral behaviors and non-moral judgments. We therefore modified our vignette and questions to consider a non-moral behavior.

Methods

Participants (N = 265) were 59% female with a mean age of 37.1 ± 11.18 years. The modified vignette read as follows:

> Every year, park visitors flock to the Rainbow Stream at Yosemite National Park.
>
> *It is traditional for visitors to try to throw one of the stream's shiny pebbles all the way across to the others side and, if they succeed, to make a wish*. Less than 5% [More than 80%] of able-bodied adult males — that is, very few [most] men — succeed.
>
> *When Tim, a young man, visits the park, he tries but fails to throw a pebble across to the other side*.

The three questions that followed collected responses on the same scale as those in previous studies, but asked: "How bad would you say Tim is at throwing?"; "To what extent is Tim to blame for not succeeding?"; and, "How athletic is Tim likely to be in general?"

Results

Respondents were more likely to rate Tim as bad at throwing when throwing success had high base rates (6.140 ± 0.228) compared to low base rates (2.978 ± 0.222) *(*$t(263) = 9.949$, $p < .0001$, $d = 1.013$). Blame-for-failure responses for low base rate (2.404 ± 0.242) and high base rate (5.078 ± 0.249) conditions also differed ($t(263) = 7.697$, $p < .0001$, $d =$

0.856). Finally, athleticism ratings for the low base rate success (4.691 ± 0.148) and high base rate (3.271 ± 0.152) conditions were moderately different ($t(263) = 6.680$, $p < .0001$, $d = 0.455$).

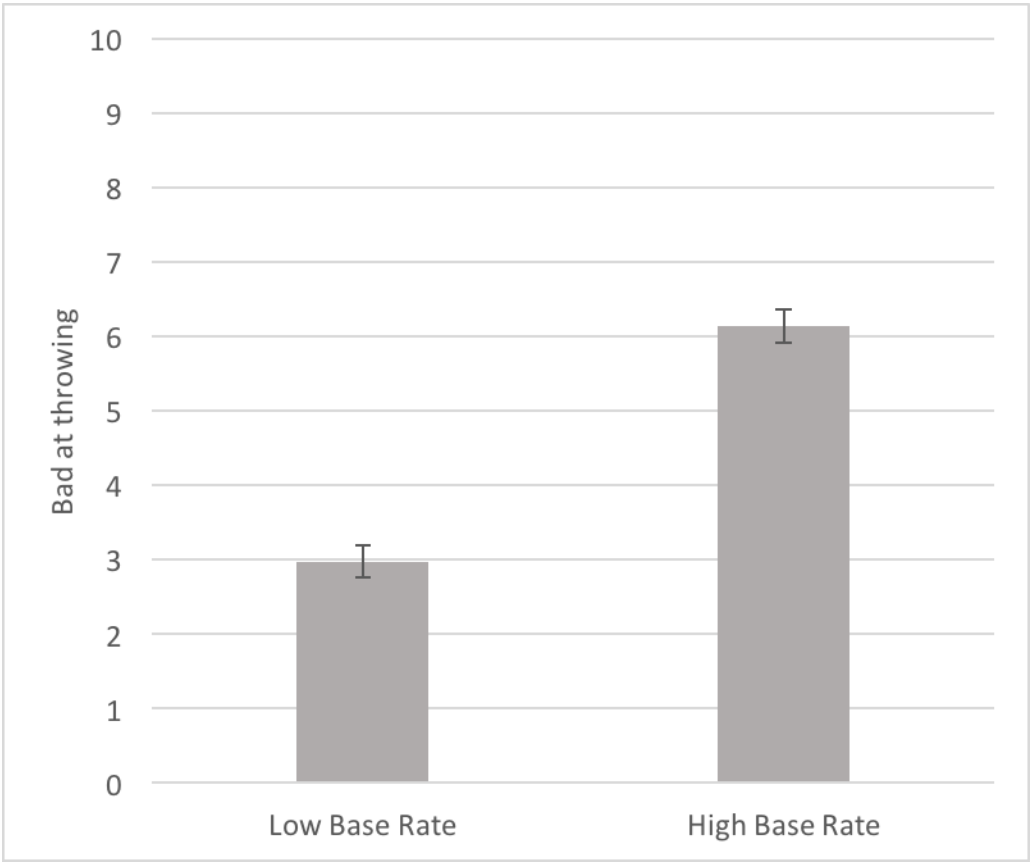

Figure 2.3.a. Mean throwing judgments for low versus high base rate conditions in Study 3. Error bars show standard error of the mean.

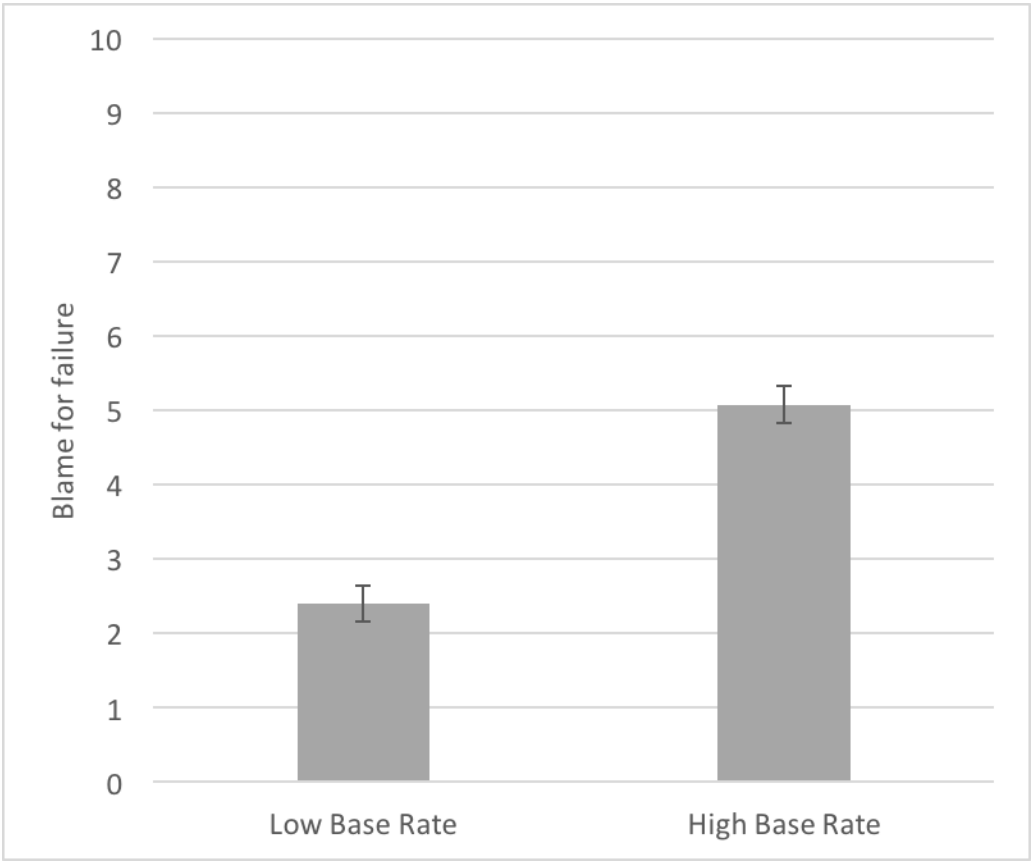

Figure 2.3.b. Mean blame for failure for low versus high base rate conditions in Study 3. Error bars show standard error of the mean.

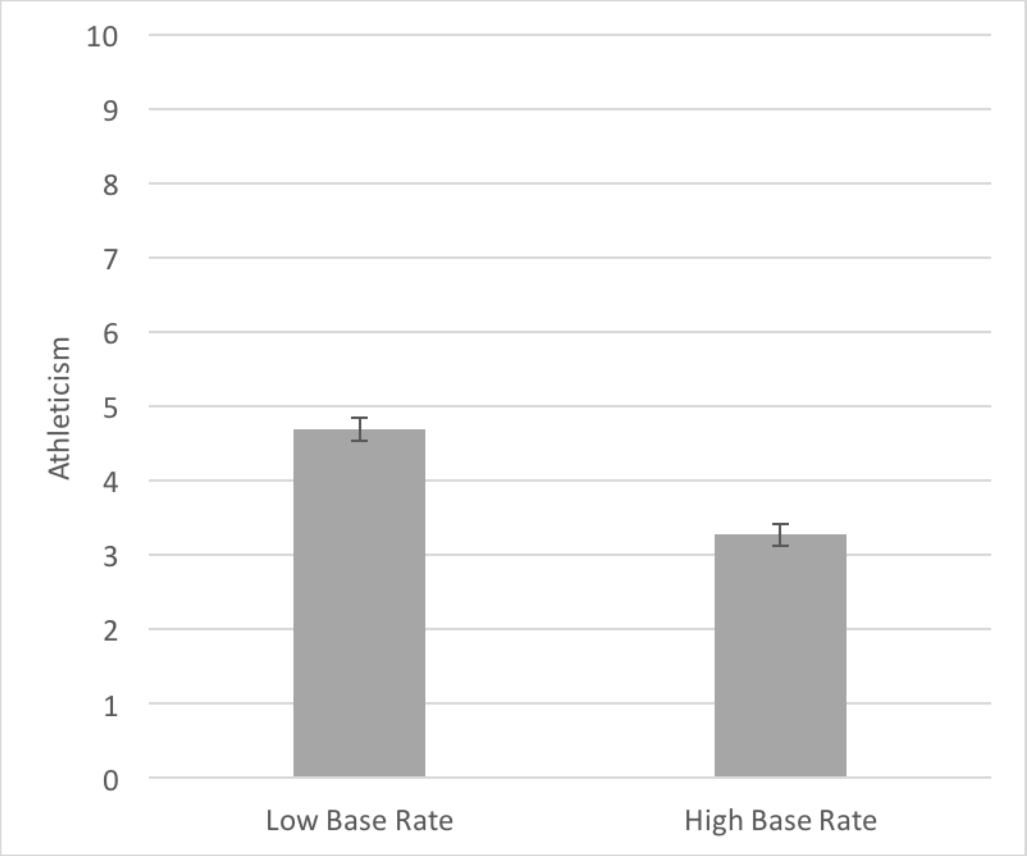

Figure 2.3.c. Mean athleticism ratings for low versus high base rate conditions in Study 3. Error bars show standard error of the mean.

Discussion

In contrast to their lack of impact on moral judgments in previous studies, success base rates significantly influenced participants' judgments of athletic performance. Thus, the lack of behavioral base-rate effects observed in Studies 1a through 2 does not appear to generalize to all types of social judgments of other people and their behaviors, and may be unique to moral judgments.

Study 4

In light of the results above, we returned to hypotheses regarding behavioral base rates and moral judgments; in particular, the matter of actor control. We next considered

whether the implications of a base rate = 100% might be too difficult to ignore. If everybody behaves the same way, one implication is that, for whatever reason, it may have been impossible, or at least difficult, to behave otherwise. To the extent that this makes harsh criticism seem inappropriate, participants could be expected to decrease the severity of their moral judgment. In addition, if, as we speculated earlier, participants make judgments not by comparison with what most others do, but instead by comparison with what well-behaved others do, a 100% base rate could reasonably be seen to include even those that are good (although this relies on a statistical inference about the likelihood that the sample contains good people). We therefore modified the vignette and base rate levels from Study 2 to compare a base rate of 0% to a base rate of 100%. Part of the vignette modification involved decreasing the number of visitors who had thus far visited so that a 100% base rate was more realistic.

Methods

Participants (N = 262) were 59% female with a mean age of 36.4 ± 12.3 years.

The modified vignette read as follows:

> Yosemite National Park has begun allowing visitors access to the park's Rainbow Stream.
>
> One of the stream's most notable features is its shiny pebbles. Almost all visitors would like to take a pebble when they leave, but the park asks all visitors not to remove any pebbles from the site.
>
> Tim is *one of 40 people to visit the stream over the course of its first few days open*. Despite the park's request, Tim takes a pebble for himself. It turns out that *he is the only one to do so, and that none of the other 40 visitors* [*every one of the other 40 visitors also*] took a pebble from the site.

Results

Blame attributions for the low base rate (8.100 ± 0.245) and high base rate (6.659 ± 0.243) conditions were moderately different ($t(260) = 4.178$, $p < .0001$, $d = 0.501$).

Wrongness ratings for the low base rate (6.220 ± 0.244) and high base rate (6.462 ± 0.245) conditions were similar ($t(260) = 0.699$, $p = .485$, $d = 0.084$). Character judgments for the low base rate (4.205 ± 0.239) and high base rate (4.638 ± 0.241) conditions were also similar ($t(260) = 1.280$, $p = .202$, $d = 0.151$), as were harmfulness ratings for the low base rate (3.985 ± 0.241) and high base rate (4.254 ± 0.243) conditions ($t(260) = 0.787$, $p = .432$, $d = 0.093$).

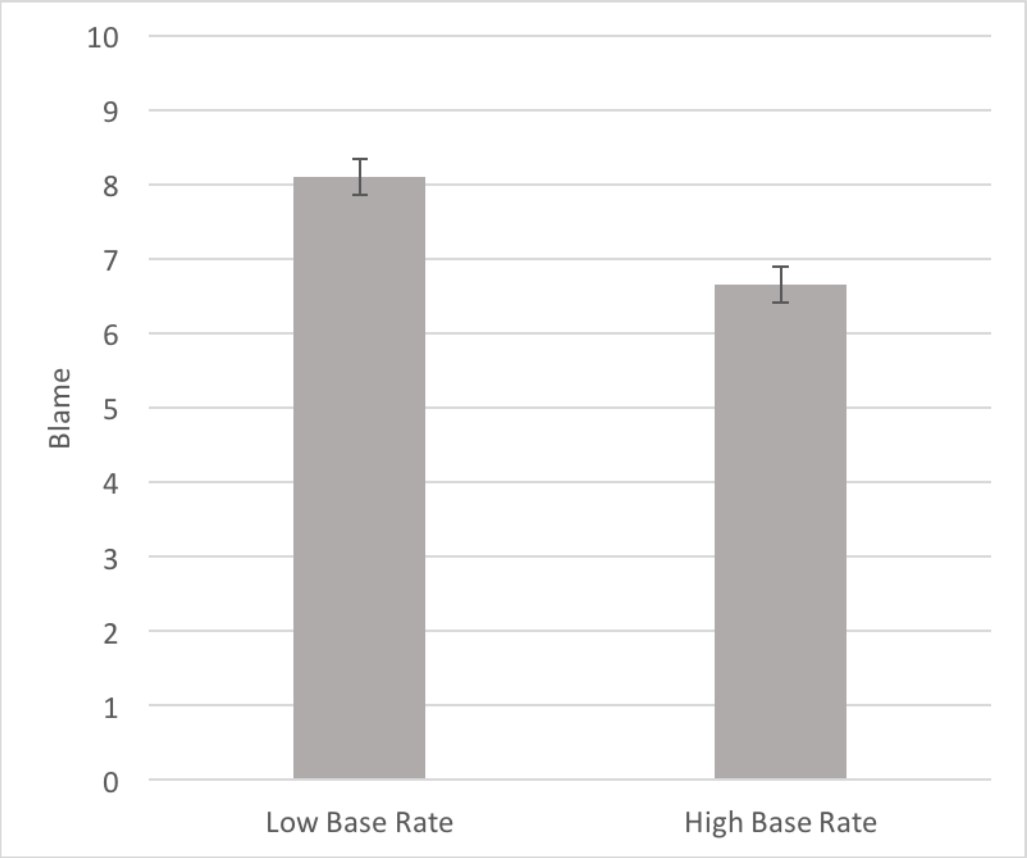

Figure 2.4.a. Mean blame attributed for low versus high base rate conditions in Study 4. Error bars show standard error of the mean.

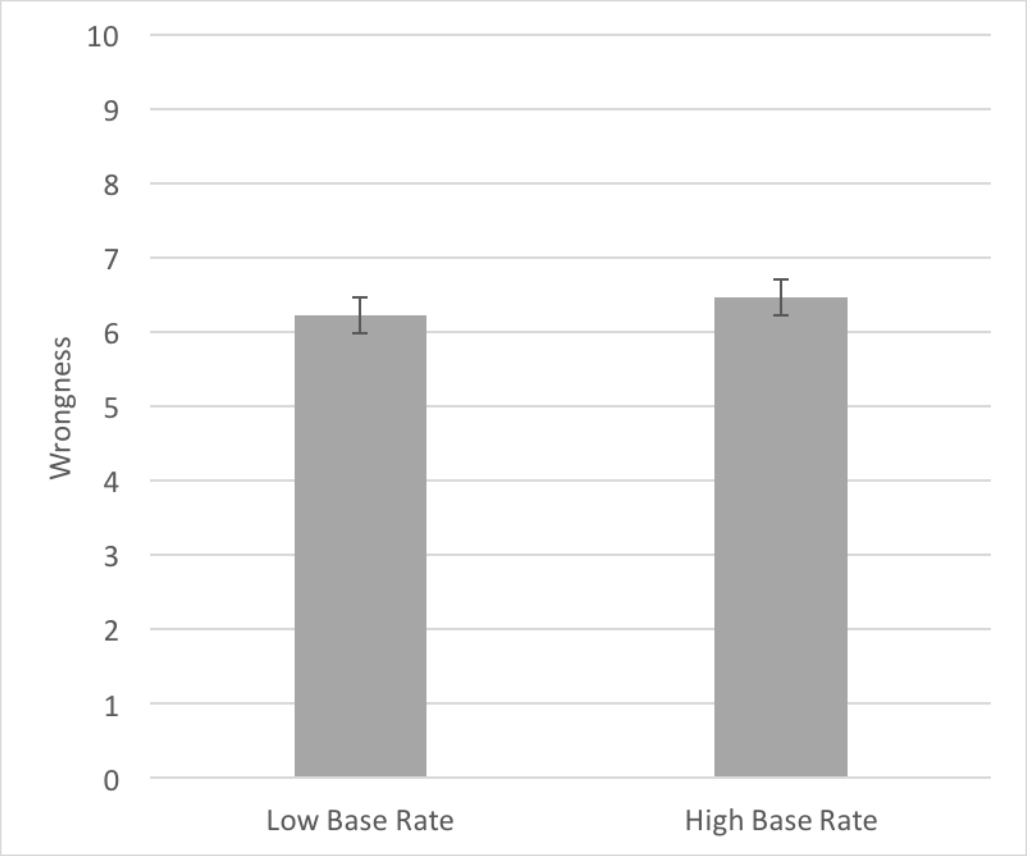

Figure 2.4.b. Mean wrongness ratings for low versus high base rate conditions in Study 4. Error bars show standard error of the mean.

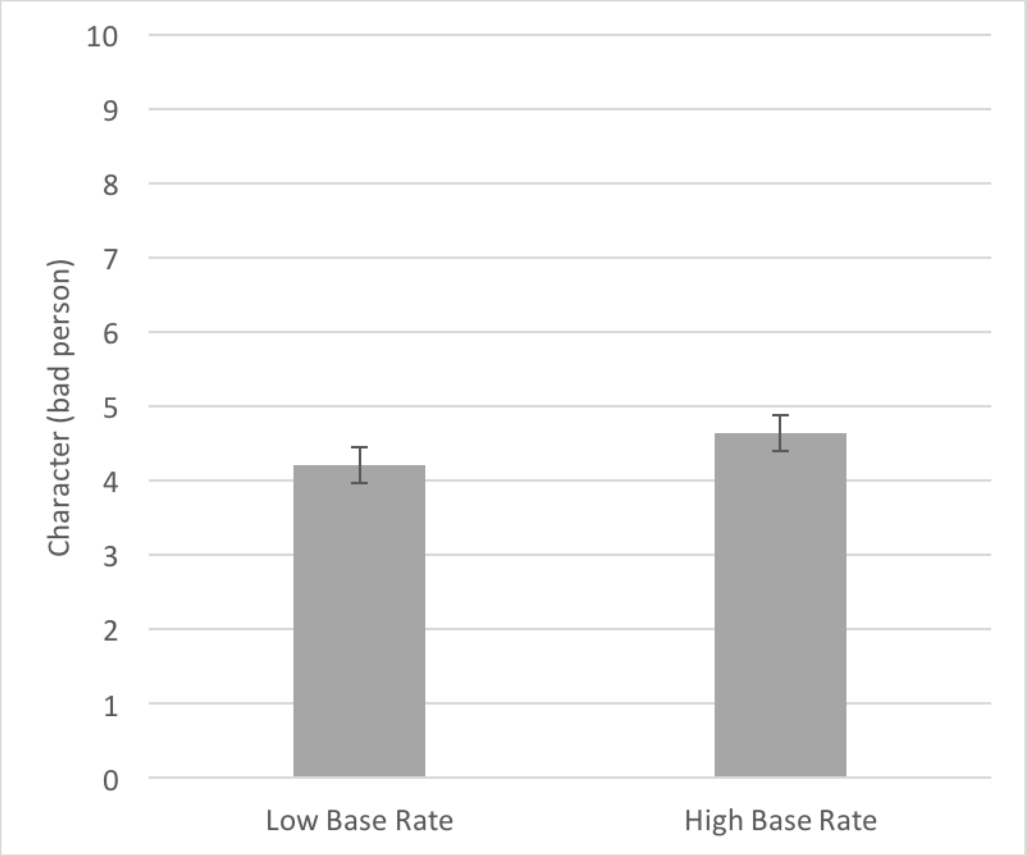

Figure 2.4.c. Mean character ratings for low versus high base rate conditions in Study 4. Error bars show standard error of the mean.

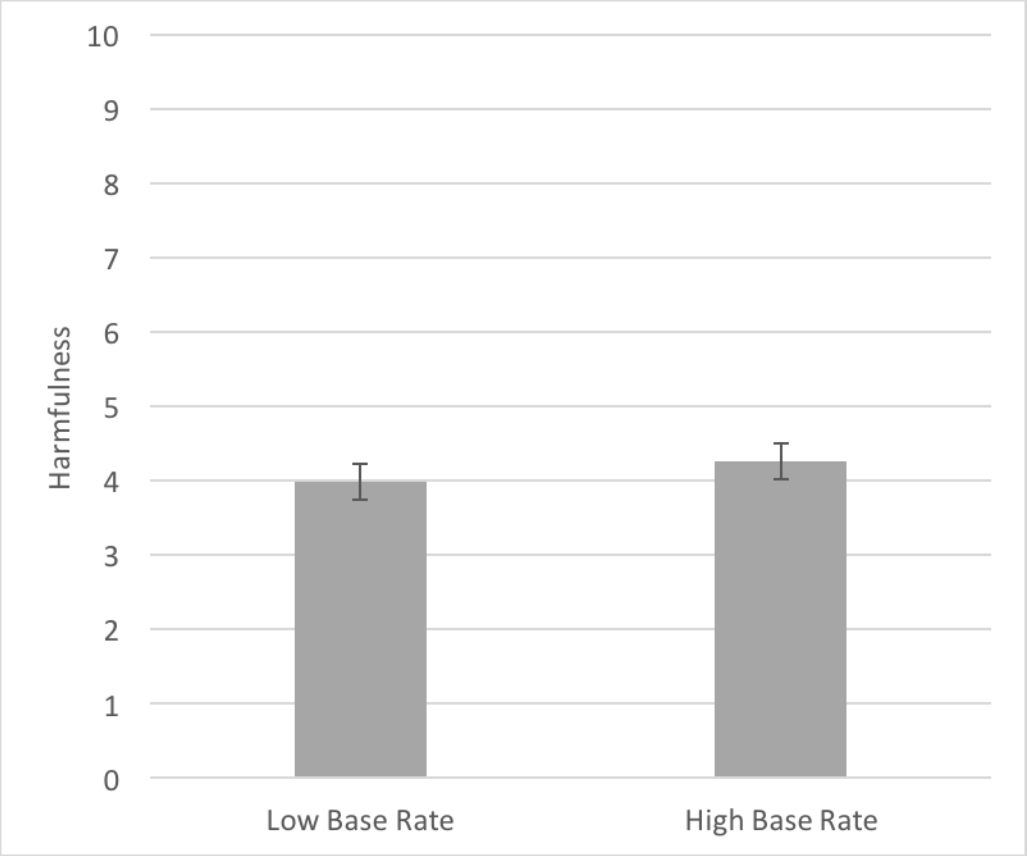

Figure 2.4.d. Mean harmfulness ratings for low versus high base rate conditions in Study 4. Error bars show standard error of the mean.

Discussion

The extreme conditions of 100% versus 0% base rate created a difference in blame judgments. However, the ratings of character, wrongness, and harmfulness remained similar across the high and low base rate conditions. Overall, thus far, base-rates have appeared, at most, impotent in their influence on moral judgments.

Combined analysis of Studies 1a-c, 2, & 4

While effect sizes are small, results of the previous studies suggest with fair consistency that higher base rates yield moral judgments with lower values. Analyzing a combined dataset gives a better overall estimate of effect size and, with increased power, a

small but reliable effect of base rate may yet prove statistically significant at the conventional alpha = 0.05.

Methods

Data from of all previous studies, except Study 3 on athleticism, were combined. This resulted in a total n = 1307 with 57% female and a mean age of 37.1 ± 12.2 years. Conditions were grouped as either low base rate or high base rate, including the 0% (low base rate) and 100% (high base rate) conditions. Results for combined ratings of blame, wrongness, and character judgment were reanalyzed as previously done for each study separately, and then again in a second reanalysis using the least square means after first statistically controlling for any differences between studies. Harmfulness ratings were not included because they were not measured in studies 1a or 1b.

Results

In the first analysis, blame attributions for the low base rate (7.576 ± 0.106) and high base rate (7.038 ± 0.110) conditions differed significantly ($t(1305) = 3.523$, $p < .001$, $d = 0.194$). The same was true in the second analysis ($t(1305) = 3.379$, $p < .001$, $d = 0.185$). In the first analysis, wrongness ratings for the low base rate (6.456 ± 0.108) and high base rate (6.228 ± 0.112) conditions were similar ($t(1305) = 1.472$, $p = .141$, $d = 0.082$). The same was true in the second analysis ($t(1305) = 1.211$, $p < .226$, $d = 0.067$). Finally, the first analysis of character judgments showed that low base rate (4.636 ± 0.103) and high base rate (4.226 ± 0.107) conditions were similar but with a statistically significant difference ($t(1305) = 2.759$, $p = .006$, $d = 0.148$). Again, the same was true of the second analysis ($t(1305) = 2.518$, $p =$

.012, $d = 0.132$). The significant effects on blame and character judgments remain significant if we apply a Bonferroni correction to account for the reanalysis (alpha = 0.05/2 = 0.025).

Discussion

Higher power analysis reveals a very small but significant influence of base rates on blame and character judgments. Wrongness judgments should be viewed with more caution due to the lack of statistical significance even with a higher power test.

Interestingly, one might imagine that when descriptive norms have any influence on moral judgments, however small, they do so by means of influencing perceived injunctive norms. However, of the three kinds of moral judgments used here, a judgment about the wrongness of a behavior is most closely aligned to the injunctive norm. And yet, of the three moral judgments, wrongness is the one for which we observed the smallest effect size and the one for which the existence of the effect is most in doubt.

Study 5

As noted, previous research reports larger and more easily-detected effects of descriptive norms on moral judgments than one would expect based on our results thus far. If this is because the descriptive norms in that previous research were confounded with injunctive qualities, then we expect that the introduction of injunctive norms to our experimental scenarios will similarly influence moral judgment here. Study 5 tests this hypothesis by incorporating into the vignette from Study 4 what is described as a widely held attitude of either acceptance or disapproval (an injunctive norm).

In addition, because the relationship between descriptive and injunctive norms is still unclear, we decided to keep the independent base rate information, creating a 2 x 2 factorial design and effectively crossing descriptive norms with injunctive norms.

Methods

Participants (N = 559) were 65% female with a mean age = 38.6 ± 12.4 years. We aimed to at least double the sample size of previous studies to account for the two-by-two factorial design with its two additional cells.

The modified vignette read as follows:

> Every year, park visitors flock to the Rainbow Stream at Yosemite National Park.
>
> One of the stream's most notable features is its shiny pebbles. Almost all visitors would like to take a pebble when they leave, but all visitors are asked to not remove any pebbles from the site.
>
> Everybody knows about the park's request, but still a few visitors (less than 5%) [many visitors (more than 80%)] do take pebbles with them. *Most people believe it's wrong [it's okay] to do so*.
>
> Consider the case of Tim. When visiting, Tim takes a pebble for himself.

Results

A two-way ANOVA revealed a significant small main effect of the injunctive norm manipulation on measures of blame ($F(1, 555) = 5.73$, $p = .017$, $\eta_p^2 = 0.010$), wrongness ($F(1, 555) = 4.25$, $p = .040$, $\eta_p^2 = 0.008$), and character ($F(1, 555) = 4.59$, $p = .033$, $\eta_p^2 = 0.008$). The main effect of the injunctive norm manipulation on judgments of harmfulness was small and not statistically significant ($F(1, 555) = 1.69$, $p = .194$, $\eta_p^2 = 0.003$)

The effect of base rates on blame ($F(1, 555) = 3.28$, $p = .071$, $\eta_p^2 = 0.006$), wrongness ($F(1, 555) = 0.066$, $p = .798$, $\eta_p^2 = 0.000$), character ($F(1, 555) = 0.552$, $p = .458$, $\eta_p^2 = 0.001$), and harmfulness ($F(1, 555) = 0.22$, $p = .643$, $\eta_p^2 = 0.000$) judgments were not statistically significant.

Likewise, the interaction between base rate and injunctive norm effects were not statistically significant: blame ($F(1, 555) = 0.21$, $p = .650$, $\eta_p^2 = 0.000$); wrongness ($F(1, 555) = 0.308$, $p = .579$, $\eta_p^2 = 0.001$); character ($F(1, 555) = 0.00$, $p = .999$, $\eta_p^2 = 0.000$); and harmfulness ($F(1, 555) = 0.061$, $p = .804$, $\eta_p^2 = 0.000$).

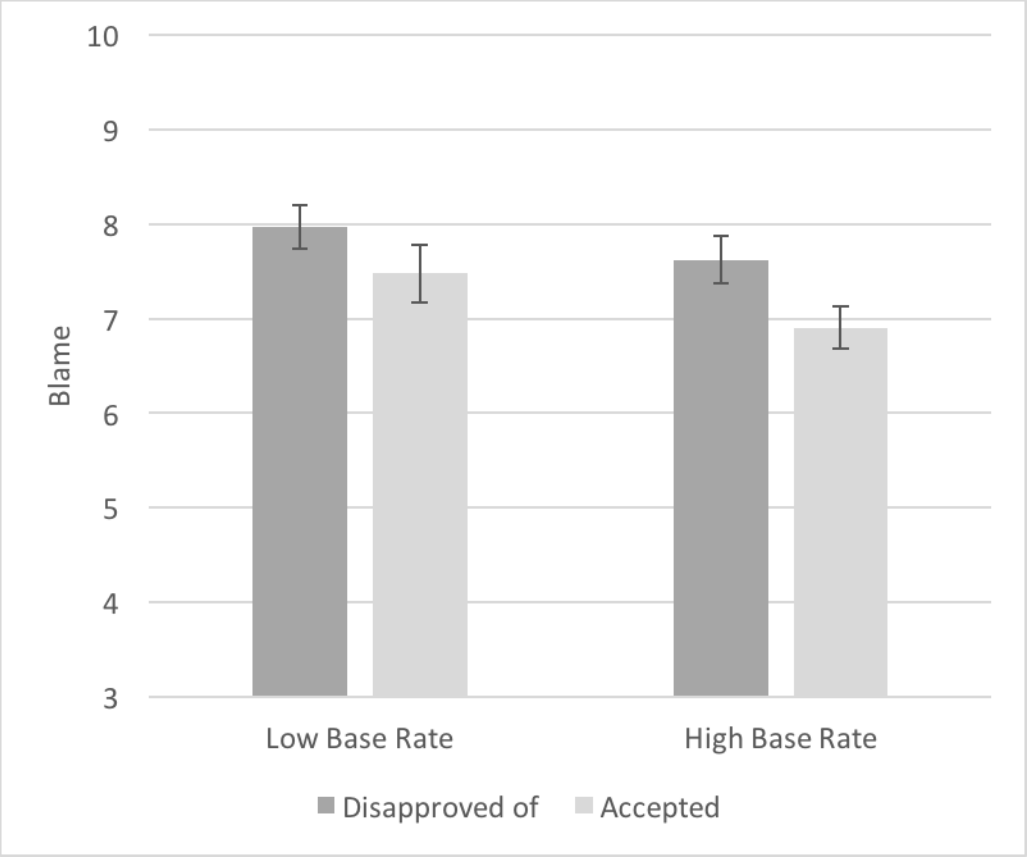

Figure 2.5.a. Blame attributed as a function of base rate and injunctive norm for Study 5. Error bars show standard error of the mean.

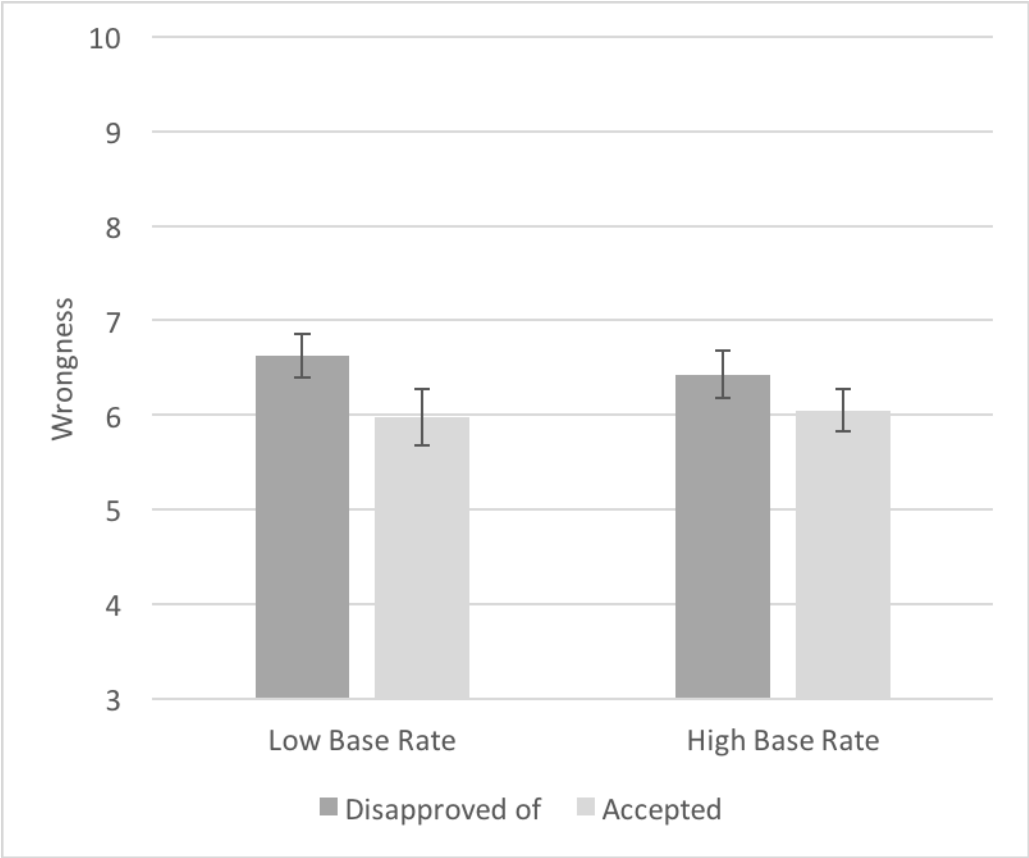

Figure 2.5.b. Wrongness rating as a function of base rate and injunctive norm for Study 5. Error bars show standard error of the mean.

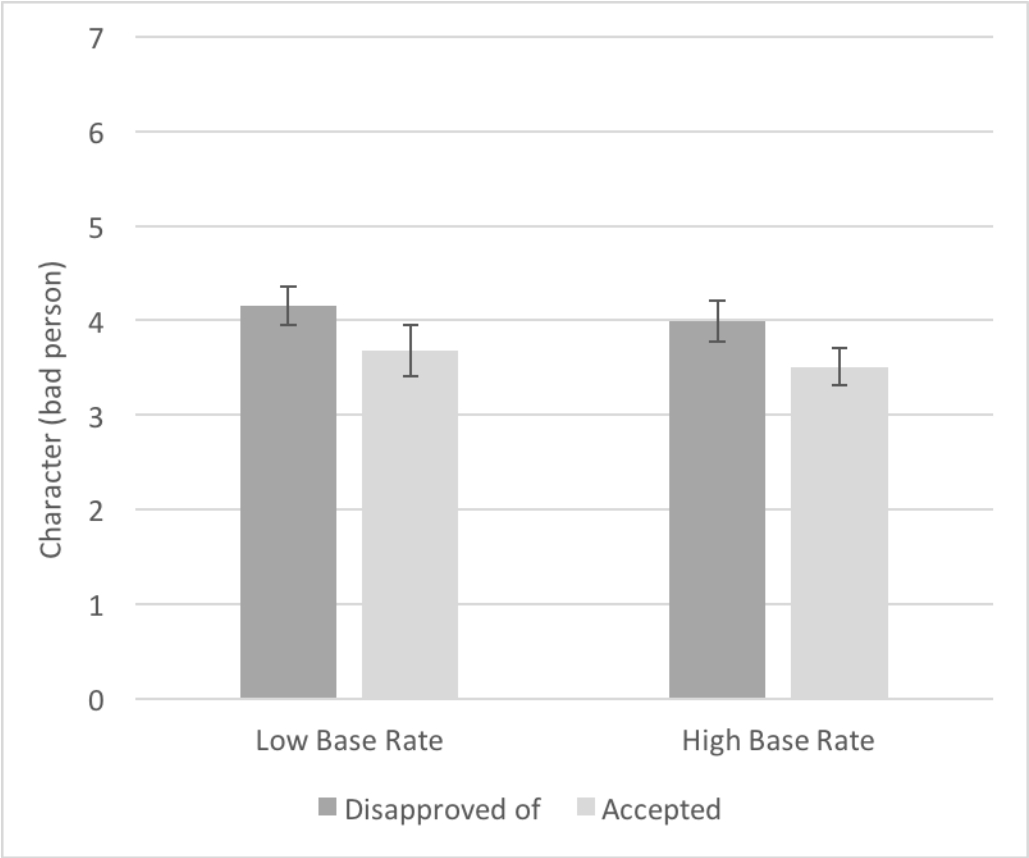

Figure 2.5.c. Character judgment as a function of base rate and injunctive norm for Study 5. Error bars show standard error of the mean.

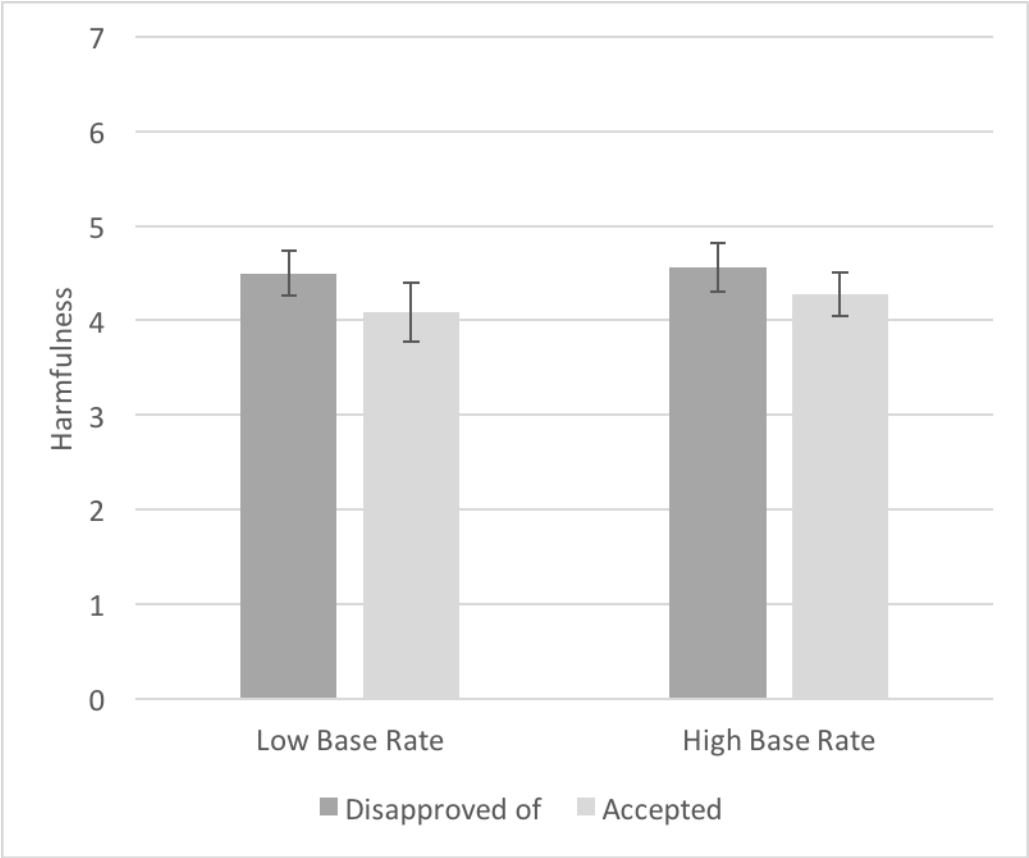

Figure 2.5.d. Harmfulness rating as a function of base rate and injunctive norm for Study 5. Error bars show standard error of the mean.

Discussion

As expected, injunctive norms had a greater influence on moral judgments than did base rates. In addition, no interactions were present between base rate and injunctive norm manipulations suggesting that the base rate neither supported nor detracted from the influence of the injunctive norm. Still, the impacts of the widespread disapproval were small.

Study 6

In light of the general resilience of moral judgments to the social factors manipulated thus far, especially in contrast to the large influence of social comparison in the non-moral case of a target individual's throwing ability (Study 3), we next test the hypothesis that it is

not what is done by others, even many others, that matters, but instead what is done by those of a more morally-relevant reference group, i.e., those perceived as good.

In Study 6, we therefore include mention of the quality of the moral character of at least some of those others who are reported as engaging in the behavior. That is, rather than simply mentioning that many people do some behavior, we now say that many people, even those of good moral character, do it. It is unclear to what extent this information about the character of some of the others speaks to the character of the target individual, and affects attributions of blame and ratings of wrongness.

Methods

Participants (N = 288) were 57% female with a mean age = 39.6 ± 12.3 years.

The modified vignette read as follows:

> Every year, park visitors flock to the Rainbow Stream at Yosemite National Park.
>
> One of the stream's most notable features is its shiny pebbles. Almost all visitors would like to take a pebble when they leave, but all visitors are asked to not remove any pebbles from the site.
>
> Despite the park's request, a few visitors, *usually those of poor moral character* [most visitors, *even those of genuinely good moral character*], do take a pebble with them.
>
> Consider the case of Tim. When visiting, Tim takes a pebble for himself.

Results

Blame attributions for the low base rate (8.387 ± 0.223) and high base rate (6.716 ± 0.210) conditions were moderately to strongly different ($t(286) = 5.455$, $p < .0001$, $d = 0.614$). Wrongness ratings for the low base rate (7.125 ± 0.234) and high base rate (5.161 ± 0.219) conditions differed strongly ($t(286) = 6.129$, $p < .0001$, $d = 0.722$). Character judgments for the low base rate (4.957 ± 0.196) and high base rate (2.799 ± 0.184) conditions also differed strongly ($t(286) = 8.010$, $p < .0001$, $d = 0.793$). Harmfulness ratings for the low

base rate (4.631 ± 0.245) and high base rate (3.924 ± 0.230) conditions were somewhat different ($t(286) = 2.105$, $p = .036$, $d = 0.260$).

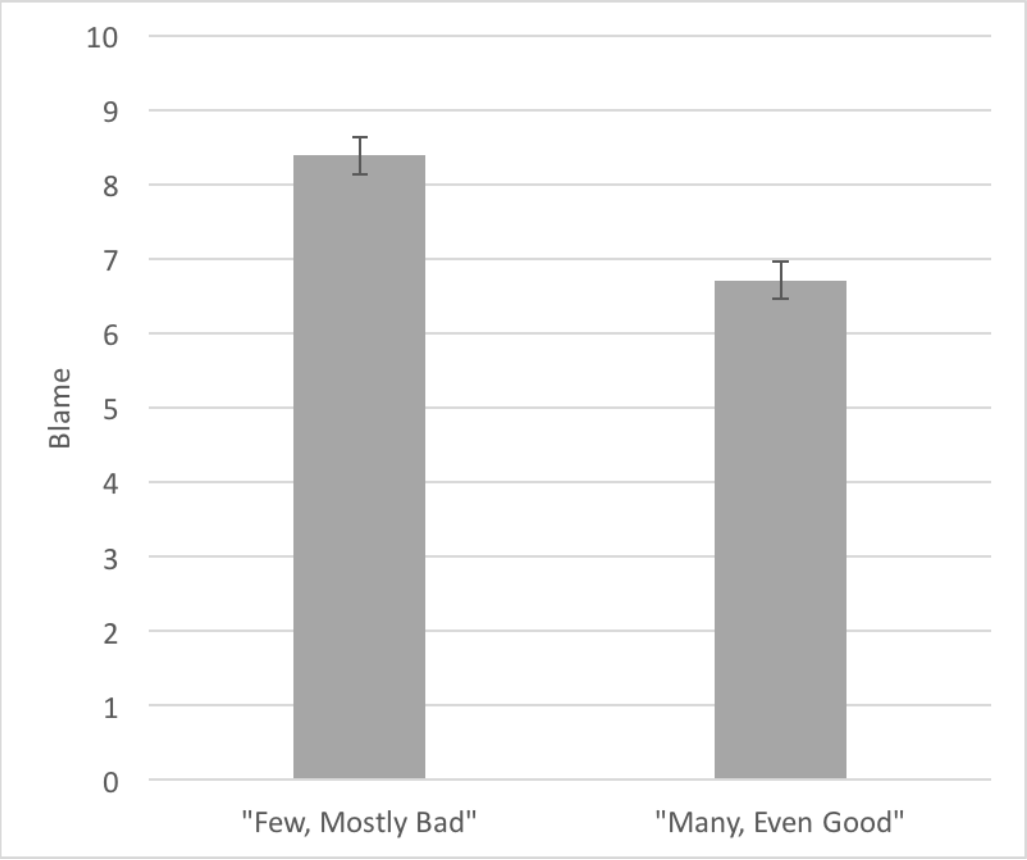

Figure 2.6.a. Mean blame attributed for low versus high base rate conditions (each with some character information) in Study 6. Error bars show standard error of the mean.

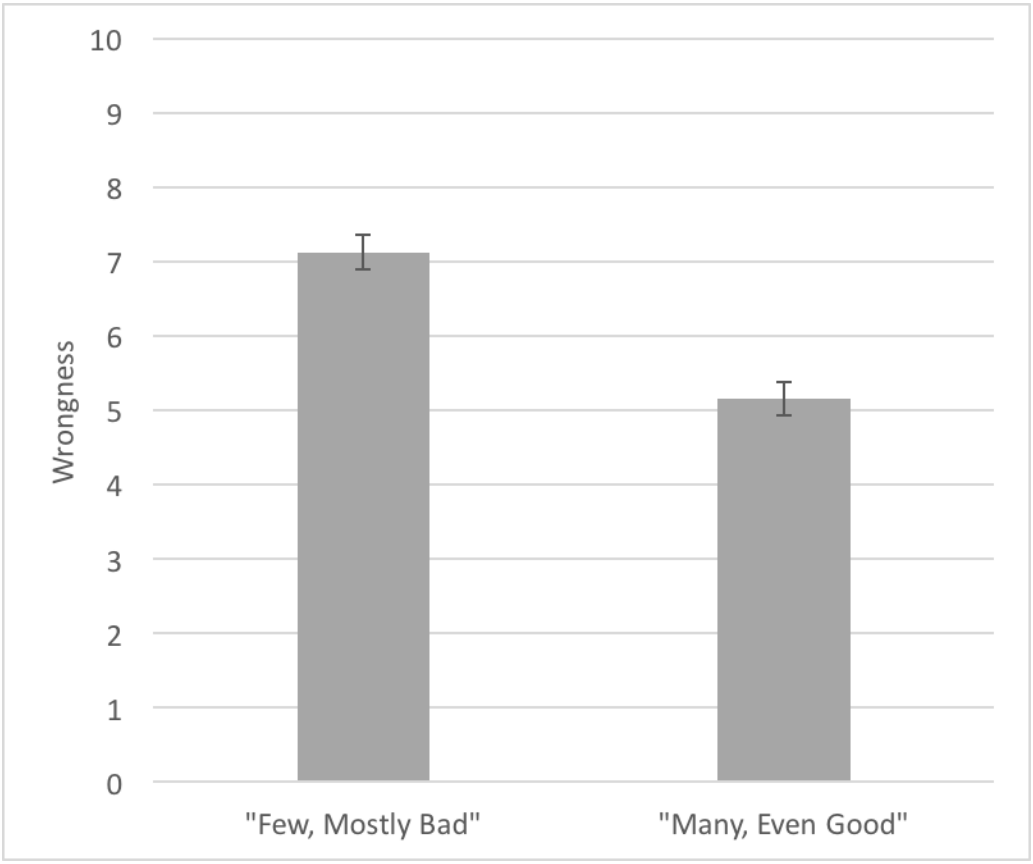

Figure 2.6.b. Mean wrongness ratings for low versus high base rate conditions (each with some character information) in Study 6. Error bars show standard error of the mean.

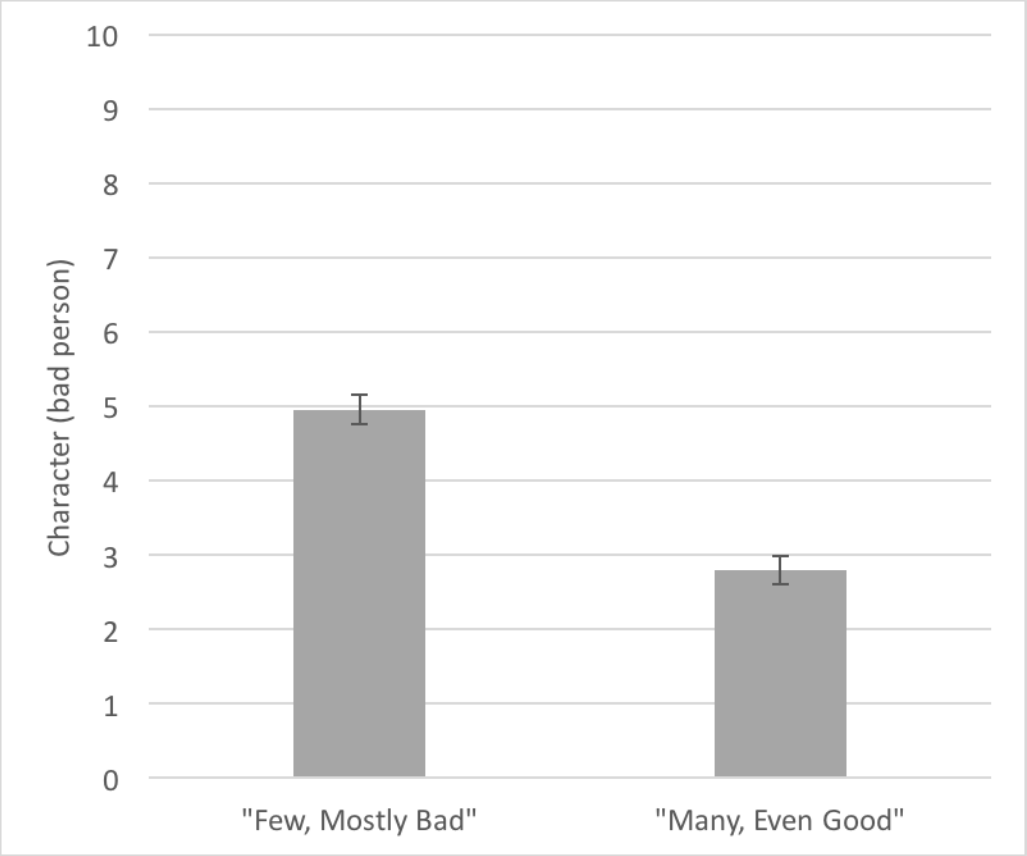

Figure 2.6.c. Mean character ratings for low versus high base rate conditions (each with some character information) in Study 6. Error bars show standard error of the mean.

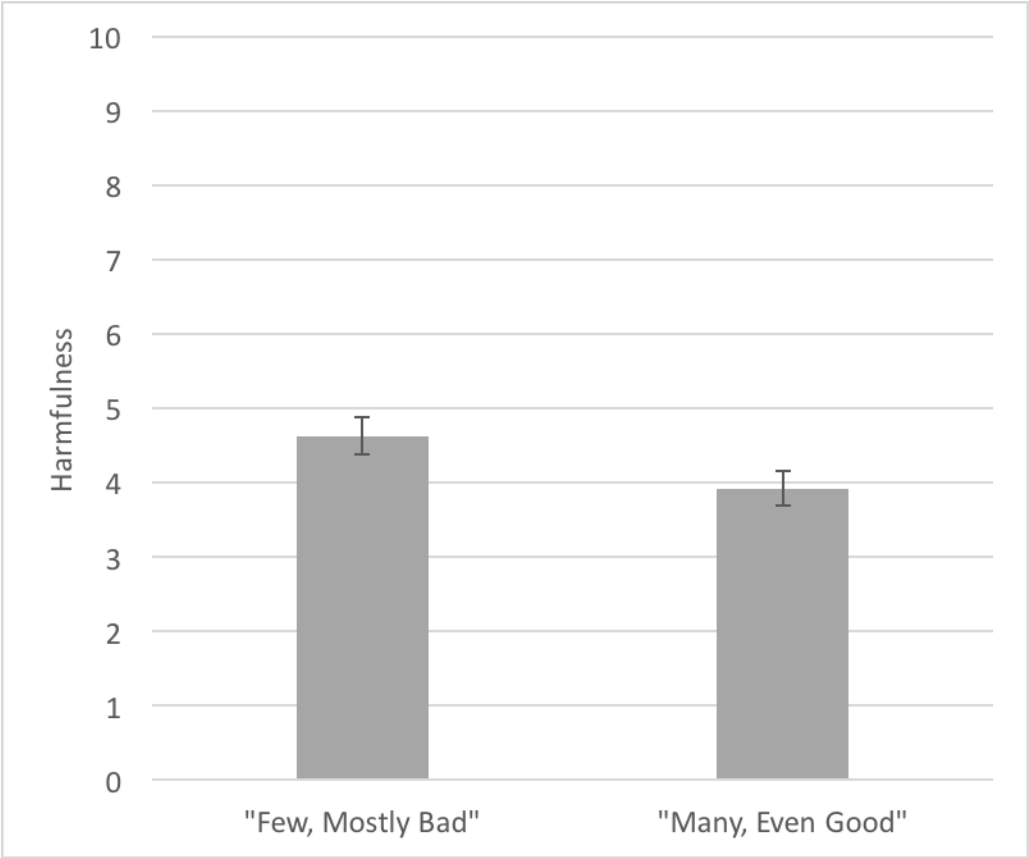

Figure 2.6.d. Mean harmfulness ratings for low versus high base rate conditions (each with some character information) in Study 6. Error bars show standard error of the mean.

Discussion

The strong impacts of others' behavior on moral judgments observed in Study 6 suggest that the extent to which others' behaviors matter depends on what we think of those others. We might infer that some behavior is not so bad if we learn that otherwise-good people engage in it. In contrast, without any notion of whether others engaging in a behavior are good or bad, their number does not seem to affect our judgment of the individual. Unless we think others who engage in some behavior are moral role-models, we might be more likely to conclude that they are only as good as their behavior than we are to conclude that their behavior is good because they engage in it.

General Discussion

In the absence of direct empirical data, the question of whether and to what extent behavioral base rates influence associated moral judgments proves challenging to answer with any confidence. In multiple experiments we found that such base rates do indeed influence judgments, but the size of the effect we observed was very small.

We considered that, in discovering moral judgments' resilience to influence from base-rates, we may have only rediscovered the fundamental attribution error. We tested this possibility by modifying our experimental materials so as to remove the moral components. We found a substantial base-rate effect on participants' judgments of performance regarding a non-moral behavior. This finding suggests that the resilience to base rates observed in the preceding studies is unlikely the result of a more general tendency like the fundamental attribution error, and may be unique to moral judgments.

We reasoned that previous research (e.g., Monroe et al., 2018) purporting to find substantial effects of base-rate information on moral judgments may have conflated that which is descriptively typical (i.e., the base rate) with that which is typically endorsed (i.e., the injunctive norm). In an experiment, we crossed prescriptive and proscriptive injunctive norms with high and low base rates and found that the impact of an injunctive norm outweighs any impact of the base rate. The implication is that the ability to draw conclusions from research failing to properly isolate the base-rate variable from injunctive norms is limited.

In an attempt to begin to reconcile our findings with well-established knowledge regarding the many powerful impacts of social-influence, we tested the hypothesis that it is not simply what is done by others, even many others, that matters much to moral judgments,

but instead what is done by those of a more morally relevant reference group. We found that simply mentioning, for example, that there were some good people among those who engaged in a high base-rate behavior, had a large influence on moral judgments; not only on judgments of the target's character, but also on judgments of blame and wrongness. As a result of that information, participants also seemed to infer that the behavior must be less harmful. This suggests that the behavior of others does indeed impact our moral judgments, but the extent of the impact, whether large or very small, depends largely on what we think of those others.

In future research, these experiments could be re-administered in populations with different cultural norms to investigate the extent to which these findings are culturally dependent. Researchers could also examine whether the influence of an injunctive norm and the influence of a role-model's behavior are additive, such that the influence of an endorsement by a role-model is even greater than that of the role-model's behavior alone or the injunctive norm alone.

References

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

Alicke, M. D., Rose, D., & Bloom, D. (2011). Causation, norm violation, and culpable control. *The Journal of Philosophy*, *108*(12), 670-696.

Bandura, A., Ross, D., & Ross, S. A. (1961). Transmission of aggression through imitation of aggressive models. *The Journal of Abnormal and Social Psychology*, *63*(3), 575.

Bear, A., & Knobe, J. (2017). Normality: Part descriptive, part prescriptive. *cognition*, *167*, 25-37.

Berofsky, B. (1966). Free will and determinism. Harper & Row

Clark, C. J., Luguri, J. B., Ditto, P. H., Knobe, J., Shariff, A. F., & Baumeister, R. F. (2014). Free to punish: A motivated account of free will belief. *Journal of personality and social psychology*, *106*(4), 501.

Cornwell, J. F., Jago, C. P., & Higgins, E. T. (2019). When group influence is more or less likely: the case of moral judgments. *Basic and Applied Social Psychology*, *41*(6), 386-395.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*(2), 353-380.

Esser, J. K. (1998). Alive and well after 25 years: A review of groupthink research. *Organizational behavior and human decision processes*, *73*(2-3), 116-141.

Goethals, G. R., & Darley, J. M. (1987). Social comparison theory: Self-evaluation and group life. In *Theories of group behavior* (pp. 21-47). Springer, New York, NY.

Goldstein, N. J., Griskevicius, V., & Cialdini, R. B. (2007). Invoking social norms: A social psychology perspective on improving hotels' linen-reuse programs. *Cornell Hotel and Restaurant Administration Quarterly*, *48*(2), 145-150.

Gray, K., & Keeney, J. E. (2015). Impure or just weird? Scenario sampling bias raises questions about the foundation of morality. *Social Psychological and Personality Science*, *6*(8), 859-868.

Grusec, J. E., & Abramovitch, R. (1982). Imitation of peers and adults in a natural setting: A functional analysis. *Child Development*, 636-642.

Holton, G. (2004). Robert K. Merton. *Proceedings of the American philosophical society*, *148*(4), 505.

Kassin, S. M. (1979). Consensus information, prediction, and causal attribution: A review of the literature and issues. *Journal of Personality and Social Psychology*, *37*(11), 1966.

Kelley, H. H. (1973). The processes of causal attribution. *American psychologist*, *28*(2), 107.

Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition*, *108*(3), 754-770.

Lynn, M. (2008). Personality effects on tipping attitudes, self-reported behaviors and customs: A multi-level inquiry. *Personality and Individual Differences*, *44*(4), 989-999.

Lynn, M., & Starbuck, M. M. (2015). Tipping customs: The effects of national differences in attitudes toward tipping and sensitivities to duty and social pressure. *Journal of Behavioral and Experimental Economics*, *57*, 158-166.

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry*, *25*(2), 147-186.

McAdams, R. H. (1997). The origin, development, and regulation of norms. *Michigan law review*, *96*(2), 338-433.

McArthur, L. A. (1972). The how and what of why: Some determinants and consequences of causal attribution. *Journal of personality and Social Psychology*, *22*(2), 171.

Milgram, S., Bickman, L., & Berkowitz, L. (1969). Note on the drawing power of crowds of different size. *Journal of personality and social psychology*, *13*(2), 79.

Miller, D. T., & McFarland, C. (1987). Pluralistic ignorance: When similarity is interpreted as dissimilarity. *Journal of Personality and social Psychology*, *53*(2), 298.

Monroe, A. E., Dillon, K. D., Guglielmo, S., & Baumeister, R. F. (2018). It's not what you do, but what everyone else does: On the role of descriptive norms and subjectivism in moral judgment. *Journal of Experimental Social Psychology*, *77*, 1-10.

Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2005). Surveying freedom: Folk intuitions about free will and moral responsibility. *Philosophical Psychology*, *18*(5), 561-584.

Nichols, S. (2002). Norms with feeling: Towards a psychological account of moral judgment. *Cognition*, *84*(2), 221-236.

Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous*, *41*(4), 663-685.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In *Advances in experimental social psychology* (Vol. 10, pp. 173-220). Academic Press.

Scanlon, T. M. (2009). *Moral dimensions*. Harvard University Press.

Schein, C., & Gray, K. (2018). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review*, *22*(1), 32-70.

Schmidt, M. F., Rakoczy, H., & Tomasello, M. (2011). Young children attribute normativity to novel actions without pedagogy or normative language. *Developmental Science*, *14*(3), 530-539.

Shaffer, L. S. (1983). Toward Pepitone's vision of a normative social psychology: What is a social norm?. *The journal of mind and behavior*, 275-293.

Sloman, S. A., Fernbach, P. M., & Ewing, S. (2009). Causal models: The representational infrastructure for moral judgment. *Psychology of Learning and Motivation*, *50*, 1-26.

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

Young, L., Nichols, S., & Saxe, R. (2010). Investigating the neural and cognitive basis of moral luck: It's not what you do but what you know. *Review of Philosophy and Psychology*, *1*(3), 333-349.

CHAPTER 3

INTUITIVE BLAMING OF PEEPING TOMS VERSUS CYBER PEEPERS

Carl P. Jago, Arseny A. Ryazanov, & Nicholas J. S. Christenfeld

Abstract

In a series of studies, we examine the existence and factors underlying a blaming discrepancy between responses to a cyber peeper and to a peeping Tom. The first study demonstrates reduced blame for a cyber peeper, and the subsequent eight studies hone in on major differences between the cyber peeping and peeping Tom scenarios that may contribute to a blaming discrepancy, such as the effort involved on the part of the peeper and the risk-taking on the part of the victim. Four of eight such factors are found to impact blame. We then compare responses to cyber peeping and peeping Tom scenarios that are roughly matched on these four factors to assess whether such matching equalizes blame. Finally, we compare blame for the viewing of a private nude photo on the Internet with blame for the viewing of a private nude photo printed on paper in order to assess whether blame is influenced by the involvement of Internet technology itself, rather than by the other factors the technology brings with it.

INTUITIVE BLAMING OF PEEPING TOMS VERSUS CYBER PEEPERS

In 2014, nude photographs of celebrities including Academy Award winning actress Jennifer Lawrence, were leaked on the Internet, widely shared, and quickly viewed by millions (Mccoy, 2014; Steinberg, 2014). Some in the media likened the entire online community to regular peeping Toms (Mendelson; 2014). Jennifer Lawrence, herself, argued that it was a sex crime, and that those who viewed the images should cower with shame for perpetuating a sexual offense (Mendelson; 2014). However, by virtue of their openness, the millions who so readily viewed, shared, and publicly commented on the images seemed not to view their cyber peeping as so serious and blameworthy as the actions of a regular peeping Tom. In a series of studies, we find that cyber peepers of this kind are indeed blamed less than peeping Toms, and we endeavor to identify the underlying causes by manipulating eight contextual factors and assessing their impact on blame attributions.

Legally, there is currently no straightforward answer as to whether the transgressions of cyber peepers and peeping Toms are comparable (Steinberg, 2014). Although laws clearly prohibit peeping Tom behaviors as they have been traditionally understood, the legality of cyber peeping is much less clear. With regard to the 2014 nude photo leak, it has been suggested that due to the combination of outdated laws from an overwhelming number of jurisdictions with different rules, the only applicable laws may be those regarding copyright, and that under these laws it is irrelevant that those portrayed in the images are naked (Steinberg, 2014). Just as MP3 audio compression and the Internet enabled the easy sharing of copyrighted music, advancing technologies generally introduce new and evolving opportunities to engage in modern analogs of illicit activities that potentially could be

regarded as morally and legally equivalent. Although legal decisions must be made about whether and to what degree the differences between analogous transgressions matter, the pace of technological developments can make it challenging for lawmakers to keep abreast (Wadhwa, 2014).

Extant theories of blame ascription are insufficiently specific to predict, a priori, a difference in blame for peeping Toms and cyber peepers. However, they help identify contextual differences between the cases that are possible causes of such a discrepancy. According to these theories, candidates for blame are generally those actors that are perceived to have caused harm either intentionally or negligently (Alicke, 2000; Cushman, 2008; Schein & Gray, 2014; Malle, Guglielmo, & Monroe, 2014). From within this framework, a peeping Tom and a cyber peeper who are perceived to have equally caused equal harms with equal intentionality would be equally blamed. On the other hand, a decrease in the perception of a perpetrator's causal contribution (Cushman, 2008; Pizarro, Uhlmann, Bloom, 2003; Searle, 1983), his intentionality (Baird & Astington, 2004; Killen, Mulvey, Knobe, 2006; Richardson, Jampol, & Woodward, 2011; Sutter, 2007), or the harm caused (Cushman, Dreber, Wang, & Costa, 2009; Gino, Moore, & Bazerman, 2008; Gino, Shu, & Bazerman, 2010; Mazzocco, Alicke, & Davis, 2004; Nadler & Rose, 2002; Horan & Kaplan, 1983) would decrease the blame attributed to that perpetrator; decreased intentionality and causal impact can be understood together as decreased control (Alicke, 2000). Although perceptions of control and harm are not the focus of this project, a search for the critical contextual differences for blame between the cases of a peeping Tom and of a cyber peeper can be narrowed to those that might plausibly impact these perceptions.

Our first study examines the existence and the magnitude of any blaming discrepancy by comparing blame for a regular peeping Tom with blame for a cyber peeper. We used, as our example of this, a cyber peeper whose actions and context mirror those of the cyber peepers in the 2014 scandal. It is worth noting that a cyber peeper who instantiates a leak – typically by hacking – is more appropriately compared with a peeping Tom who, like a cyber peeper who hacks, also performs an additional transgression, such as breaking and entering, damaging property, or trespassing (Study 9 addresses this comparison). In the same way, a peeping Tom who does not break and enter, or otherwise transgress in ways over and above that of taking an opportunity to see somebody in the nude without their consent, is best compared with a cyber peeper who similarly exploits such an opportunity but who does not also engage in hacking.

The subsequent eight studies explore differences between the cyber peeping and peeping Tom scenarios where there are theoretical reasons to suspect these differences might contribute to a blaming discrepancy. Specifically, we examine whether the cyber peeper is blamed less than the regular peeping Tom due to:

- the celebrity status of the victim – because peeping at a celebrity could be perceived as less harmful, especially for celebrities who appear naked, or nearly naked, for their careers (Spiegel, 1957);
- the increased risk-taking of the victim – because a peeper whose victim took on greater risk, perhaps by sharing a nude selfie, may be perceived as less uniquely causally responsible for bringing about the event (Karuza & Carey, 1984; Kelley, 1972);
- the lower effort involved in cyber peeping – because easy, low effort peeping permits impulsive behavior and may be perceived as less planful and thus less intentional (e.g., Pizarro, Uhlmann, & Salovey, 2003);

- a perceived increase in the likelihood that the cyber peeper was merely curious rather than pruriently motivated – because prurient peeping may be perceived as creepier and thus more harmful, or because it results in a more negative spontaneous evaluation of the perpetrator's character (Alicke, 2000; Uhlmann, Pizarro, Diermeier, 2015; e.g., Alicke, 1992);
- the large number of other cyber peepers who are also able to peep – because this may be perceived as diffused responsibility (e.g., Darley & Latane, 1968) or undermine the peeper's unique causal influence (Alicke, 2000; Kelley, 1972; e.g., Pizarro, Uhlmann, Bloom, 2003);
- the viewing of a still photograph as opposed to live viewing – because photo-viewing may be perceived as less of an invasion of privacy and thus less harmful;
- the increased distance between the peeper and the victim – because distant peeping may too be perceived as less of an invasion and thus less harmful, or because the increased spatial separation from the victim decreases perceptions of the peeper's causal influence over any harmful outcomes (Hart & Honore, 1959; Prosser & Wade, 1971);
- the involvement of a middleman who performs the initial hacking – once again this could diminish the perceived unique causal contribution of the peeper (Alicke, 2000; e.g., Kelley, 1972), divide responsibility between the hacker and peeper (e.g., Darley & Latane, 1968), or facilitate lower effort, less planful, potentially more impulsive, and thus less intentional peeping (e.g., Pizarro, Uhlmann, & Salovey, 2003).

Following the series of studies that explores these possible factors, the next study compares cyber peeping and peeping Tom scenarios that are matched on the particular factors identified as important differences between the original peeping Tom and cyber peeper cases. This study reveals whether, when so matched, blame is equalized. The final study compares blame for the viewing of a private nude photo on the Internet with blame for the viewing of a private nude photo printed on paper in order to assess whether blame is

influenced simply by the involvement of Internet technology, rather than the other factors that the technology brings with it.

Study 1: Cyber peeper versus peeping Tom

Method

Sixty American participants (33 females, M_{age} = 35.7 years, SD = 10.5 years) recruited on Mechanical Turk completed a survey in which they were randomly assigned to read a vignette describing either a case of cyber peeping or a case of a regular peeping Tom. The cyber peeping vignette mirrored the events of the 2014 scandal. The two vignettes read as follows:

> Peeping Tom:
>
> Jane is changing clothes in her hotel room. Unbeknownst to her, she can be seen from outside through a small gap in the ventilation duct. Tim, Jane's colleague, is waiting outside for Jane to change. He has heard about the gap in the ventilation duct and realizes that if he were to look through he'd see Jane in the nude. He decides to find the gap to look through, and sees Jane in the nude.
>
> Cyber Peeper:
>
> Jane, a famous actress, takes a nude photo of herself using her cell phone. The contents of her cell phone, which are backed up to the Internet, are hacked, and her nude photo is leaked on the Internet where it can be easily viewed. Tim, who knows Jane from seeing her most recent film, has heard about the nude photo and a website where it can be seen. He decides to visit the website, and he sees Jane in the nude.

Having read one of these vignettes, participants were asked rate the blameworthiness of the peeper on an 11-point scale, anchored at 0 with the label, "Not At All", at 5 with, "Moderately, and at 10 with, "Maximally."

Results and Discussion

Participants blamed the peeping Tom (M=9.02, SD=3.33) far more than the cyber peeper (M=4.47, SD=2.03); $t(58) = 6.39$, $p < 0.0001$, 95% CI = [3.13, 5.98], $d = 1.28$. However, with a mean blameworthiness rating near the middle of the scale, the cyber peeper was still considered moderately blameworthy.

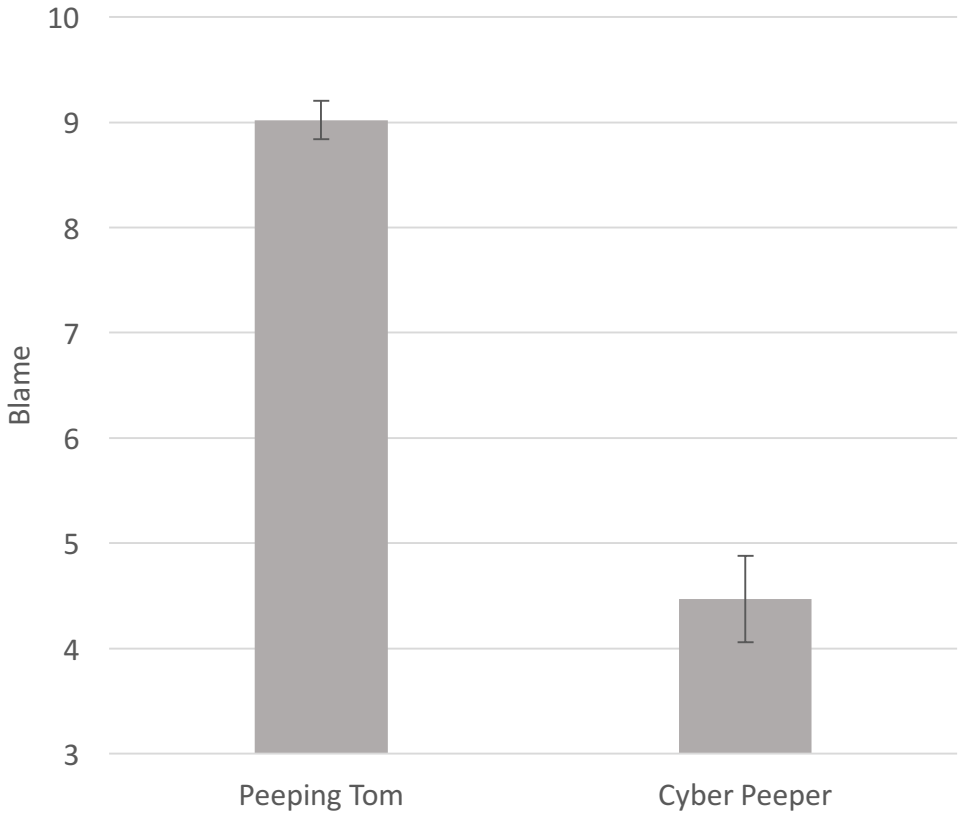

Figure 3.1. Mean blame attributed to cyber peeper versus peeping Tom. Error bars show standard error of the mean.

Given that the cases used here are gendered - that is, the perpetrator is male and the victim female - one might imagine that blaming differs across gender, but blame from female participants (M=6.75, SD=3.64) was no different than from male participants (M=6.74, SD=3.55); $t(58) = 0.01$, $p > 0.9$, 95% CI = [-1.86, 1.88], $d = 0.00$; and participant gender did not interact significantly with peeper type; $F(1, 56) = 0.25$, $p > 0.6$, $\eta_p^2 = 0.004$. Similarly,

younger adults could be thought to hold different Internet norms from older adults, however participant age had no impact on blaming; neither as a main effect, $b = 0.00$, $t(58) = 0.06$, $p > 0.9$ ($R^2 = 0.00$, $F(1, 58) = 0.00$, $p > 0.9$), nor by interacting with the type of peeper, $F(1, 56) = 0.33$, $p > 0.5$, $\eta_p^2 = 0.006$.

Cyber peeping as described thus far is clearly not judged as harshly as regular in-person peeping. However, given that there are many differences between the case of the in-person peeping Tom and that of the cyber peeper, it remains unclear which are causally responsible for the large difference in blame. Studies 2 through 9 seek to discover which differences matter in this regard. In each case, our manipulation aims to remove a condition that is present in one scenario, and add it to the other scenario where it was absent, and vice versa. For example, because the victim in Study 1 was a celebrity in the cyber peeping case and not in the peeping Tom case, in Study 2 we add a non-celebrity, cyber peeping victim and a celebrity, peeping Tom victim. This allows us to test the relevance of victim's celebrity status in blaming the peeper.

Study 2: Celebrity versus non-celebrity victim

Celebrity status itself might be treated as permission to peep (Spiegel, 1957). In the days following the 2014 celebrity photo leak, on NBC's The Tonight Show, host Jimmy Fallon referred to the event and joked that, "Jennifer Lawrence wants to sue the hacker for damages, while Kim Kardashian wants to pay the hacker for his troubles" (Fallon, 2014). Fallon was alluding to the career boost that sometimes accompanies these kinds of events. Because it is possible to see such exposure as providing some benefit to a celebrity that it would not provide to a non-celebrity, such that it is less harmful overall, the perceived blameworthiness

of a peeper might be reduced (Baron & Hershey,1988; Gino, Moore, & Bazerman, 2009; Nadler & Rose, 2002). This study tests whether peepers, both over the Internet and in person, are blamed less for viewing a celebrity than for viewing a non-celebrity.

Method

One hundred and ninety-four American participants (106 females, M_{age} = 36.1 years, SD = 12.0 years) recruited on Mechanical Turk were randomly assigned to read about either a cyber peeper or a peeping Tom, and about either a celebrity victim or a non-celebrity victim. In this and all studies that follow, with the exception of Study 6 where the critical manipulated variable is continuous rather than nominal, we aimed for 200 to 240 participants – or fifty to sixty participants per cell where the factorial design varies from the 2 by 2 standard – which is sufficient to find small-medium effects (of size $\eta_p^2 = 0.038$ at 200 participants and $\eta_p^2 = 0.032$ at 240 participants) with 80% power according to the G Power software. In this and all following studies, the measure was identical to that in Study 1. The four modified vignettes read as follows:

> Peeping Tom + Celebrity Victim:
>
> Jane, a famous actress, is changing clothes in her hotel room. Unbeknownst to her, she can be seen from outside through a gap in the ventilation duct. Tim, who knows Jane from seeing her most recent film, has heard about the gap in the ventilation duct and realizes that if he were to look through he'd see Jane in the nude. He decides to look through the gap, and he sees Jane in the nude.
>
> Peeping Tom + Non-Celebrity Victim:
>
> Jane is changing clothes in her hotel room. Unbeknownst to her, she can be seen from outside through a gap in the ventilation duct. Tim, Jane's colleague, has heard about the gap in the ventilation duct and realizes that if he were to look through he'd see Jane in the nude. He decides to look through the gap, and he sees Jane in the nude.
>
> Cyber Peeper + Celebrity Victim:

Jane, a famous actress, is changing clothes in her hotel room. Unbeknownst to her, hackers have compromised her laptop which is on the hotel room desk, and they have remotely activated her webcam. A nude photo of her is now accessible on the Internet. Tim, who knows Jane from seeing her most recent film, has heard about the nude photo and a website where it can be seen. He decides to visit the website, and he sees Jane in the nude.

Cyber Peeper + Non-Celebrity Victim:

Jane is changing clothes in her hotel room. Unbeknownst to her, hackers have compromised her laptop which is on the hotel room desk, and they have remotely activated her webcam. A nude photo of her is now accessible on the Internet. Tim, Jane's colleague, has heard about the nude photo and a website where it can be seen. He decides to visit the website, and he sees Jane in the nude.

Results and Discussion

A 2x2 ANOVA showed that participants blamed the peeping Tom (M=9.35, SD=1.25) more than the cyber peeper (M=5.83, SD=3.31); $F(1, 190) = 98.76$, $p < 0.0001$, $\eta_p^2 = 0.34$. In addition, the perpetrator was blamed less when peeping at a celebrity (M=7.25, SD=3.40) than when peeping at a non-celebrity (M=8.22, SD=2.37); $F(1, 190) = 6.80$, $p < 0.010$, $\eta_p^2 = 0.034$. The interaction between the type of peeper and the celebrity status of the victim was significant such that celebrity status influenced blame more in the cyber peeping case than in the peeping Tom case, where blame ratings were up against the ceiling; $F(1, 190) = 5.08$, $p = 0.025$, $\eta_p^2 = 0.026$.

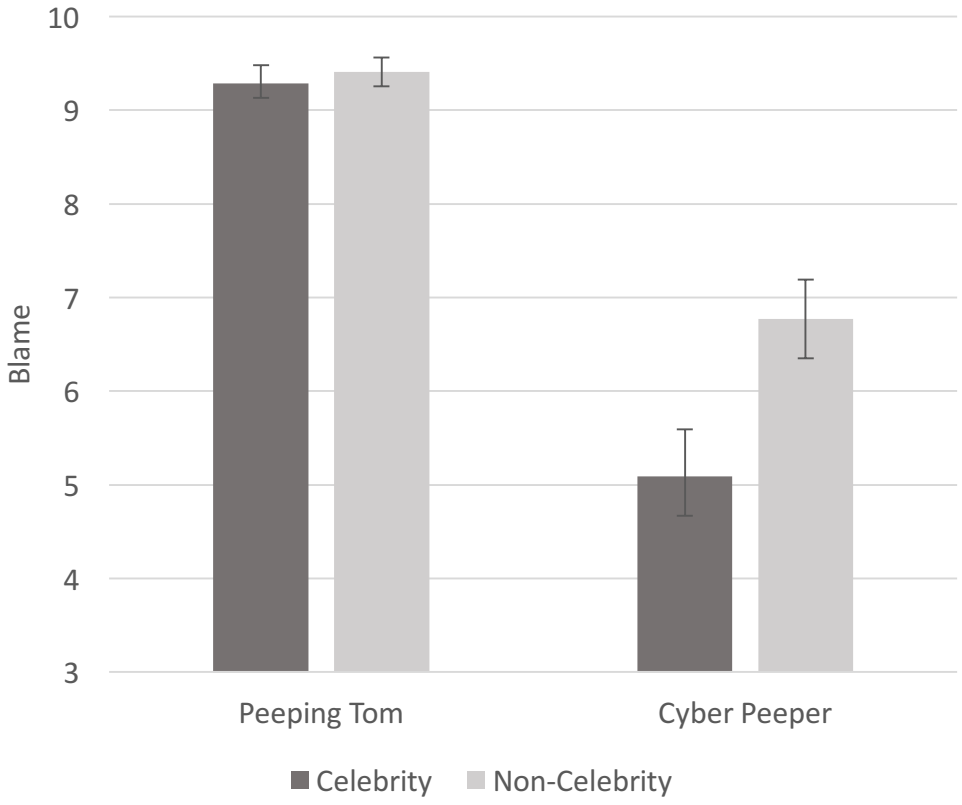

Figure 3.2. Mean blame attributed to celebrity versus non-celebrity victim. Error bars show standard error of the mean.

The medium-small effect size of the celebrity status of the victim suggests that this factor may contribute to, but not account for all of, the difference in blame for the cyber peeper and the peeping Tom that was observed in Study 1 and replicated here. The interaction, whereby celebrity status seemed to matter more in the cyber peeper case than in the peeping Tom case may be a ceiling effect due to the fact that the average blame rating for the peeping Tom was near maximum on the scale; for some participants, the maximum rating may be over-determined.

Study 3: Higher risk-taking versus lower risk-taking victim

In some cases of cyber peeping, including the case used in Study 1, victims take and sometimes even share with a partner their own nude photos. Similar risk-taking has been shown to increase victim blaming (e.g., Kanekar, Kolsawalla, & D'Souza, 1981; Karuza & Carey, 1984; Muehlenhard, Friedman, & Thomas, 1985; Schult & Schneider, 1991; Scroggs, 1976). Alicke (2000) suggests that this occurs by the consideration of process control, whereby victims are held responsible for opting in to riskier situations, or when it is believed that they should have known better. Study 3 tests whether the risk-taking, which is expected to increase blame for the victim, might also mitigate blame for the peeper.

Method

Two hundred and seventeen American participants (128 females, M_{age} = 37.0 years, SD = 12.8 years) recruited on Mechanical Turk were randomly assigned to read about either a cyber peeper or a peeping Tom, and about either a higher risk-taking victim or a lower risk-taking victim. The cyber peeping victim from Study 1, who shares a nude photo, is used as a case of higher risk, whereas the peeping Tom victim who was changing in her hotel room is used as a case of lower risk. We compare these with a low-risk cyber peeping victim who never takes or shares any nude photographs, and a higher-risk peeping Tom victim who chooses to sunbath in the nude. This risk-taking manipulation, as with other manipulations in studies that follow, is not identical for the cyber peeper and the peeping Tom. It is possible, even likely, that one operationalization might be stronger than the other. Still, this approach serves to enable the intended comparison - in this case, between the higher and lower-risk version of each - and is expected to reveal whether or not the factor is relevant in blaming, and whether any impact on blame occurs in expected direction both in person and online. The four modified vignettes read as follows:

Peeping Tom + Higher Risk-Taking Victim:

Jane is sun bathing on the private patio outside her hotel room when she decides to remove her bathing suit in order to avoid tan lines. The balcony is mostly obscured from sight but there is a small gap in the wall. She hopes that nobody will see her and goes ahead and removes her bathing suit. Tim, Jane's colleague, is outside by the hotel pool. He sees Jane's bathing suit draped over the wall surrounding her patio and realizes that Jane is sun bathing in the nude. He has heard about the gap in the wall and realizes that if he were to peek through, he'd see Jane in the nude. He decides to find the gap and to peek through, and he sees Jane in the nude.

Peeping Tom + Lower Risk-Taking Victim:

Jane is changing clothes in her hotel room. Unbeknownst to her, she can be seen from outside through a small gap in the ventilation duct. Tim, Jane's colleague, is waiting outside for Jane to change. He has heard about the gap in the ventilation duct and realizes that if he were to look through he'd see Jane in the nude. He decides to find the gap and to peek through, and he sees Jane in the nude.

Cyber Peeper + Higher Risk-Taking Victim:

Jane is changing clothes in her hotel room and decides to use her cell phone to send a nude photo of herself to her boyfriend. She hopes that the photo isn't somehow leaked and she goes ahead and sends it. Hackers have compromised the text messaging service that handles the text message and her nude photo is now publicly accessible on the Internet. Tim, Jane's colleague, has heard about the nude photo of Jane and decides to search for it at google.com. When looking through the images in the search results, he recognizes Jane's face in a thumbnail-size photo. He enlarges the photo to see Jane in the nude.

Cyber Peeper + Lower Risk-Taking Victim:

Jane is changing clothes in her hotel room. Unbeknownst to her, hackers have compromised her laptop which is on the hotel desk, and they have remotely activated her webcam. A nude photo of her is now publicly accessible on the Internet. Tim, Jane's colleague, has heard about the nude photo of Jane and decides to search for it at google.com. When looking through the images in the search results, he recognizes Jane's face in a thumbnail-size photo. He enlarges the photo to see Jane in the nude.

Results and Discussion

A 2x2 ANOVA showed again that participants blamed the peeping Tom (M=9.05, SD=1.66) more than the cyber peeper (M=6.51, SD=3.03); $F(1, 213) = 60.11$, $p < 0.0001$, η_p^2

= 0.220. The peeper was not blamed significantly more or less when peeping at higher risk-taking victim (M=7.68, SD=2.87) than when peeping at a lower risk-taking victim (M=8.04, SD=2.54); $F(1, 213) = 1.07$, $p = 0.303$, $\eta_p^2 = 0.005$. The was also no interaction between the type of peeper and victim risk-taking; $F(1, 213) = 0.66$, $p = 0.418$, $\eta_p^2 < 0.01$.

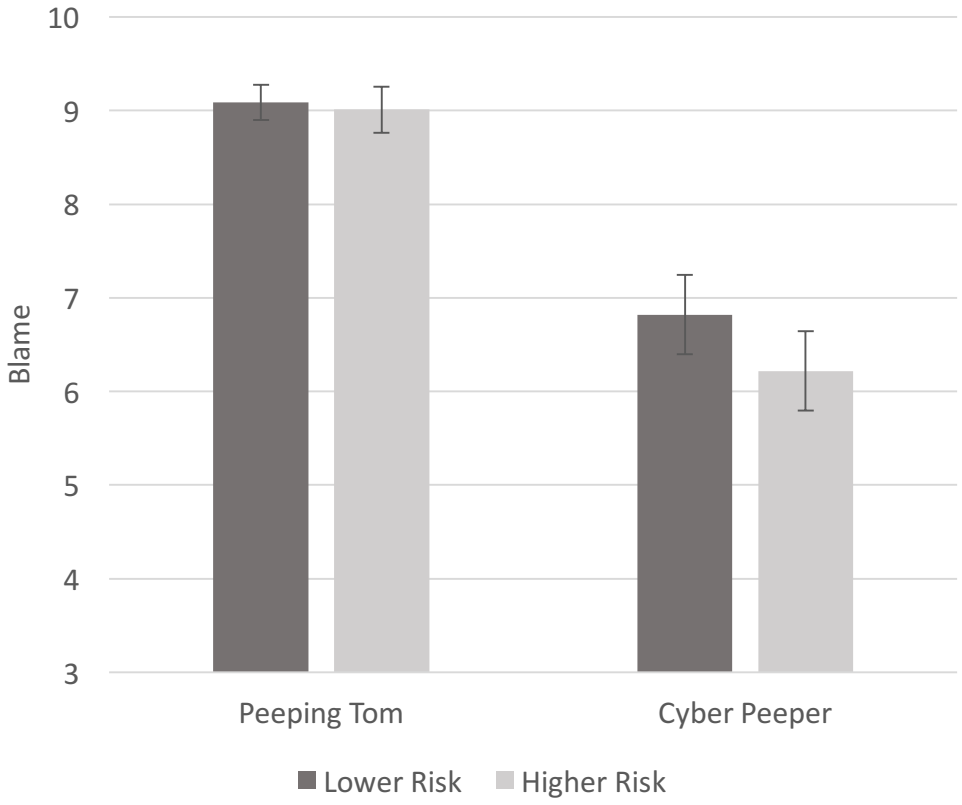

Figure 3.3. Mean blame attributed for higher versus lower risk-taking victim. Error bars show standard error of the mean.

The results suggest that the difference in victim risk-taking between the cyber peeping victim and peeping Tom victim does not seem to contribute to the difference in blame that was observed in Study 1 and persists here even when the riskiness of the victim is controlled. Additionally, blame for perpetrator was not responsive to variations in blame for

the victim. That is, an increase in blame for the victim was not accompanied by a decrease in blame for the perpetrator. Therefore, the total sum of blame appeared not to be fixed.

Study 4: Higher effort versus lower effort peeping

With the aid of Internet technology, viewing people in the nude, including those who do not wish to be so viewed, can be trivially easy. Low effort cyber peeping might be tempting in a way that generally higher-effort in-person peeping is not. The effort expended in the endeavor might indicate the robustness of the peeper's intention and be treated as reflective of his general moral character. As with perceptions of reduced intentionality, less negative indications of character mitigate blame for the cyber peeper (Bayles, 1982, Alicke, 1992). Moreover, a low effort peeping opportunity permits more impulsive behavior than does higher effort, and perhaps more planful, peeping. Impulsivity has similarly been shown to mitigate blame (e.g., Pizarro, Uhlmann, & Salovey, 2003). Study 4 tests whether the degree of effort in pursuing the outcome impacts blame for the peeper.

Method

Two hundred and seventeen American participants (135 females, M_{age} = 39.4 years, SD = 12.9 years) recruited on Mechanical Turk were randomly assigned to read about either a cyber peeper or a peeping Tom, and about either lower effort or higher effort peeping. While cyber peeping is generally low effort, we create a very low effort case where the peeper need only click a link, and we contrast it with a higher effort case that involves a degree of persistent online searching. Similarly, the high effort peeping Tom case involved searching for the gap through which to peep, whereas the low effort case involved an

immediately apparent gap and required no such searching. The four modified vignettes read as follows:

Peeping Tom + Lower Effort:

Jane is changing clothes in her hotel room. Unbeknownst to her, she can be seen from outside through a gap in the ventilation duct. Tim, Jane's colleague, is waiting outside for Jane to change. He notices the gap in the ventilation duct beside him and realizes that if he were to look through he'd see Jane in the nude. The gap is large enough and positioned so that Tim need only tilt his head a little to see through. He decides to look and sees Jane in the nude.

Peeping Tom + Higher Effort:

Jane is changing clothes in her hotel room. Unbeknownst to her, she can be seen from outside through a small gap in the ventilation duct. Tim, Jane's colleague, is waiting outside for Jane to change. He has heard about the gap in the ventilation duct and realizes that if he were to look through he'd see Jane in the nude. He decides to find the gap. It's not immediately obvious where the ventilation duct is, but after some searching he finds it and decides to peek through. The gap is small, but with some effort, he sees Jane in the nude.

Cyber Peeper + Lower Effort:

Jane is changing clothes in her hotel room. Unbeknownst to her, hackers have compromised her laptop which is on the hotel desk, and they have remotely activated her webcam. A nude photo of her is now accessible on the Internet. Tim, Jane's colleague, has heard about the nude photo of Jane, and receives an email with a link to the photo. He clicks on the link to see Jane in the nude.

Cyber Peeper + Higher Effort:

Jane is changing clothes in her hotel room. Unbeknownst to her, hackers have compromised her laptop which is on the hotel desk, and they have remotely activated her webcam. A nude photo of her is now accessible on the Internet. Tim, Jane's colleague, has heard about the nude photo of Jane and decides to search for it at google.com. It's not immediately obvious how to find the picture. After a few different searches and after looking through a few dozen images in the various search results, he recognizes Jane's face in a thumbnail-size photo. He enlarges the photo in order to see Jane in the nude.

Results and Discussion

A 2x2 ANOVA showed again that participants blamed the peeping Tom (M=7.86, SD=2.64) more than the cyber peeper (M=6.65, SD=3.23); $F(1, 213) = 9.15$, $p < 0.003$, $\eta_p^2 = 0.041$. In addition, and as predicted, the perpetrator was blamed less for low effort peeping (M=6.70, SD=2.87) than for high effort peeping (M=7.78, SD=3.06); $F(1, 213) = 7.61$, $p = 0.006$, $\eta_p^2 = 0.035$. The interaction between the type of peeper and the degree of effort was not significant; $F(1, 213) = 1.76$, $p = 0.187$, $\eta_p^2 < 0.01$.

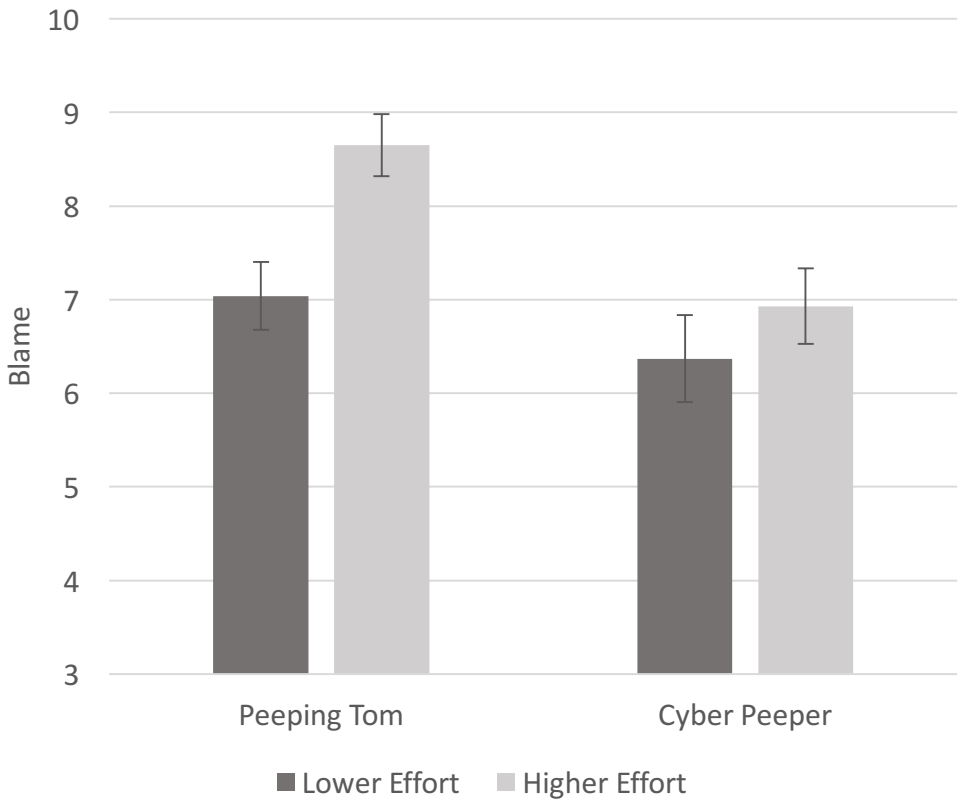

Figure 3.4. Mean blame attributed for higher effort versus lower effort peeping. Error bars show standard error of the mean.

The medium-small effect size of the effort manipulation suggests that this factor, as with victim celebrity status, may contribute to, but not account for all of, the difference in blame for the cyber peeper and the peeping Tom that was observed in Study 1 and replicated here. While blaming is responsive to how much effort the perpetrator puts in, a main effect of peeping online versus peeping in person persists.

Study 5: Prurient versus curious motivation

It is possible that peeping Toms are more likely to be perceived as acting to satisfy prurient desires than cyber peepers who may be more likely to be perceived as non-pruriently curious. Creepy, prurient peeping could be blamed more than non-prurient peeping if the perceived prurience produces a greater spontaneous, negative, affective response (e.g., Bailey & Rothblatt, 1985; Kassin & Garfield, 1991; Kerr, 1978). Study 5 tests whether prurience impacts blame for the peeper.

Method

Two hundred and seven American participants (122 females, M_{age} = 36.1 years, SD = 11.5 years) recruited on Mechanical Turk were randomly assigned to read about either a cyber peeper or a peeping Tom, and about either prurient or non-prurient peeping. The prurient peeper was described as choosing to peep because he has always been sexually attracted to the victim, whereas the non-prurient peeper was described as not being sexually attracted, but nonetheless still curious to see. The four modified vignettes read as follows:

> Peeping Tom + Non-Prurient Curiosity:
>
> Jane is sun bathing on the private patio outside her hotel room. She decides to remove her bathing suit because she believes she looks sexier when naked in front of her boyfriend if she has a full body tan. The balcony is mostly obscured from sight but there is a small gap in the wall. She hopes that nobody will see her and goes ahead

and removes her bathing suit. Tim, Jane's colleague, is outside by the hotel pool. He sees Jane's bathing suit draped over the wall surrounding her patio and realizes that Jane is sun bathing in the nude. He also notices a gap in the wall beside him and realizes that if he were to look through he'd see Jane in the nude. Tim is not sexually attracted to Jane and is only attracted to men, but he is still curious to see and so he looks through to see Jane nude.

Peeping Tom + Prurience:

Jane is sun bathing on the private patio outside her hotel room. She decides to remove her bathing suit because she believes she looks sexier when naked in front of her boyfriend if she has a full body tan. The balcony is mostly obscured from sight but there is a small gap in the wall. She hopes that nobody will see her and goes ahead and removes her bathing suit. Tim, Jane's colleague, is outside by the hotel pool. He sees Jane's bathing suit draped over the wall surrounding her patio and realizes that Jane is sun bathing in the nude. He also notices a gap in the wall beside him and realizes that if he were to look through he'd see Jane in the nude. Tim has always been sexually attracted to Jane and so he looks through to see Jane nude.

Cyber Peeper + Non-Prurient Curiosity:

Jane is changing clothes in her hotel room and decides to use her cell phone to send a nude photo of herself to her boyfriend. She hopes that the photo isn't somehow leaked and she goes ahead and sends it. A virus has compromised the text messaging service that handles the text message and her nude photo is now publicly accessible on the Internet. Tim, Jane's colleague, has heard about the nude photo of Jane. He receives an email with a link to the photo. Tim is not sexually attracted to Jane and is only sexually attracted to men, but he is still curious to see and so he clicks on the link to see Jane nude.

Cyber Peeper + Prurience:

Jane is changing clothes in her hotel room and decides to use her cell phone to send a nude photo of herself to her boyfriend. She hopes that the photo isn't somehow leaked and she goes ahead and sends it. A virus has compromised the text messaging service that handles the text message and her nude photo is now publicly accessible on the Internet. Tim, Jane's colleague, has heard about the nude photo of Jane. He receives an email with a link to the photo. Tim has always been sexually attracted to Jane and so he clicks on the link to see Jane nude.

Results and Discussion

A 2x2 ANOVA showed again that participants blamed the peeping Tom (M=6.85, SD=3.12) more than the cyber peeper (M=4.15, SD=3.26); $F(1, 203) = 38.04$, $p < 0.0001$, $\eta_p^2 = 0.158$. In addition, the perpetrator was blamed less for non-prurient peeping (M=4.96, SD=3.15) than for prurient peeping (M=6.03, SD=3.34); $F(1, 203) = 6.05$, $p < 0.015$, $\eta_p^2 = 0.029$. The interaction between the type of peeper and prurience was not significant; $F(1, 203) = 1.23$, $p = 0.269$, $\eta_p^2 < 0.006$.

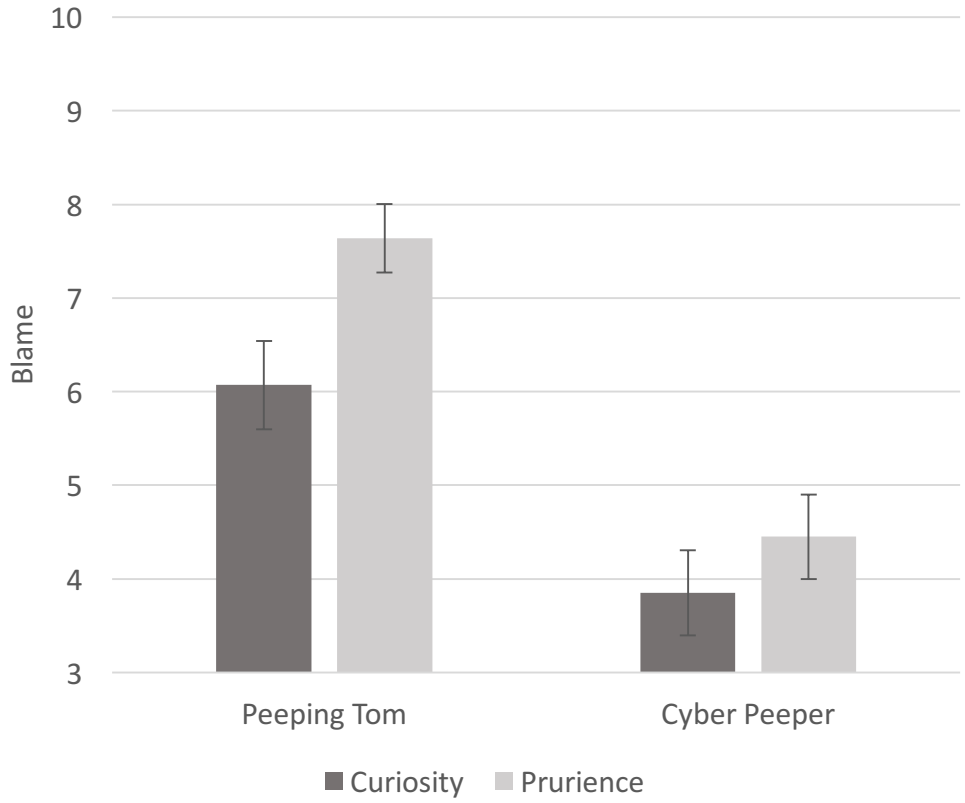

Figure 3.5. Mean blame attributed for prurient versus curious motivation. Error bars show standard error of the mean.

The medium-small effect size of the prurience manipulation suggests that this factor, as with victim celebrity status and degree of effort, may contribute to the difference in blame for the cyber peeper and the peeping Tom observed in Study 1.

Study 6: Number of peepers

Due to the involvement of the Internet, one aspect present in many cyber peeping scenarios, but which is not typical of peeping Tom scenarios, is that the peeper is just one of many. Blame for an individual peeper may be reduced as the likelihood increases that others are peeping at the same victim; perhaps due to a sharing of blame in a manner related to Darley and Latane's (1968) diffusion of responsibility, or by a normalizing of the behavior. As the attributional discounting principle suggests (Kelley, 1972), multiple agents ambiguate individual casual attributions by providing alternate casual explanations for an event and diminishing the perceived uniqueness of an individual's behavior. Study 6 tests this possibility by manipulating the number of possible peepers.

Method

Two hundred and seventy-eight American participants (160 females, M_{age} = 36.8 years, SD = 12.9 years) recruited on Mechanical Turk were randomly assigned to read about either a cyber peeper or a peeping Tom, and either that there were no other peepers, or that the peeper in the vignette was one of roughly ten, one hundred, ten thousand, or a million other potential peepers. The peeping Tom variations were capped at 100 other potential peepers to limit absurdity, while for the cyber peeping, this number can easily extend into the millions. The modified vignettes read as follows:

Peeping Tom:

Jane is sun bathing on the private patio outside her hotel room when she decides to remove her bathing suit in order to avoid tan lines. The balcony is mostly obscured from sight but there is a small gap in the wall. She hopes that nobody will see her and goes ahead and removes her bathing suit. Tim, Jane's colleague, is walking down a pathway to the hotel pool. He sees Jane's bathing suit draped over the wall surrounding her patio and realizes that Jane is sun bathing in the nude. He also notices a gap in the wall beside him and realizes that if he were to look through he'd see Jane in the nude. He knows nobody else/about 10 of his colleagues/about 100 of his colleagues will notice the gap in the wall and be able to see Jane. He decides to look through to see Jane in the nude.

Cyber Peeper:

Jane is away dressing in her hotel room. She decides to use her cell phone to email her boyfriend, Rob, a private link to a nude photo of herself changing. She gives her email the subject line: "I miss you, Rob. See my nude photo xxx". She hopes that the link isn't somehow leaked and she goes ahead and sends it. A virus has compromised the email service; it automatically sends the email, including the link, randomly to a/10/100/10,000/1,000,000 different email address/es. Tim, Jane's colleague, receives the email with the link to the photo. He sees the subject and realizes that the email was meant for Jane's boyfriend and that the link will reveal a nude photo of Jane. He can see that the virus has sent the link only to him/to 10 people/to 100 people/to 10,000 people/to 1,000,000 people. He decides to click on the link to see the photo.

Results and Discussion

Given that the range of the number of peepers was so large, extending from 0 to 1,000,000, in discrete steps of one or two orders of magnitude, and was limited in the case of the peeping Tom scenario to a maximum of 100, we computed an ANCOVA using the log (base 10) of the number of peepers as a continuous factor crossed with the nominal factor of peeper type. As expected, the analysis showed that participants blamed the peeping Tom (M=6.97, SD=2.76) more than the cyber peeper (M=5.15, SD=3.31); $F(1, 274) = 10.31$, $p = 0.0015$, $\eta_p^2 = 0.036$. The perpetrator, however, was not blamed significantly less when there was a larger number of other potential peepers than when there were fewer or no other

peepers; $F(1, 274) = 0.62$, $p > 0.4$, $\eta_p^2 = 0.002$. There was also no significant interaction between the two factors; $F(1, 274) = 2.05$, $p = 0.154$, $\eta_p^2 = 0.007$.

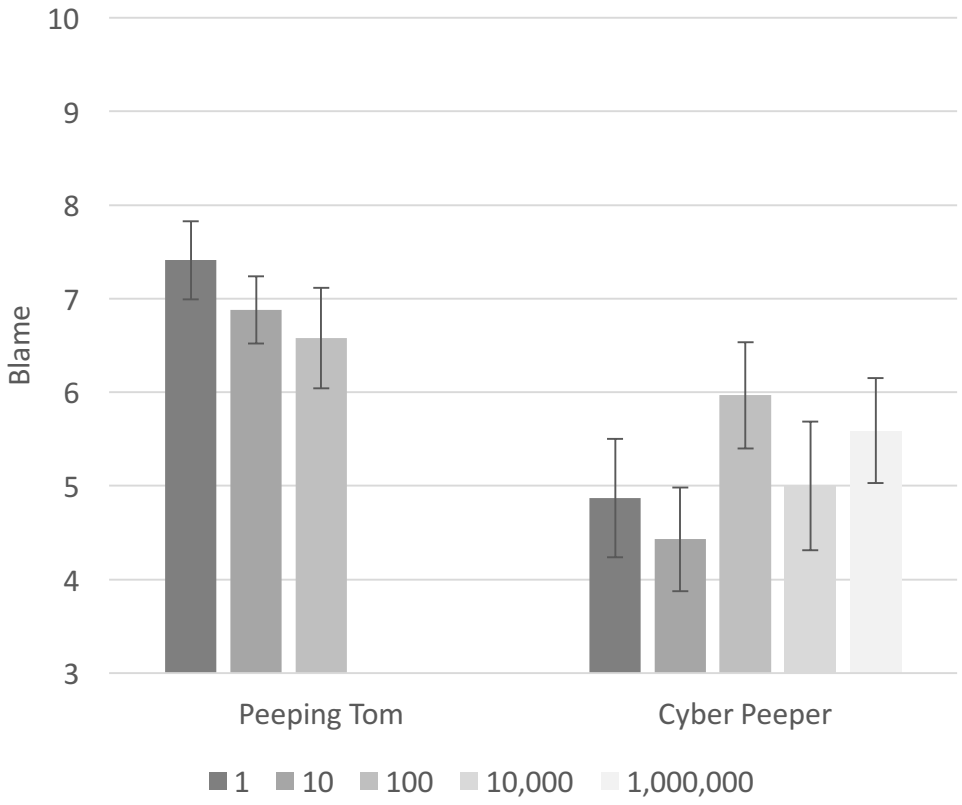

Figure 3.6. Mean blame for varying number of other peepers. Error bars show standard error of the mean.

The results suggest that the number of other potential peepers does not account for the difference in blame in Study 1 and replicated here. However, it is possible that a similar manipulation which stipulates the involvement of additional peepers rather than an increased potential for additional peepers might have different results.

Study 7: Still photo versus live viewing

The viewing of leaked photos may be judged less harshly because they are still images, removed in time, whereas peeping in person involves a live, moving victim. Peeping

live, in real time, might be considered a greater invasion of privacy and thus a more harmful transgression. If this is the case, a cyber peeper could be blamed more for viewing a live-stream than for viewing a still photograph. Study 7 tests this possibility by comparing blame for live viewing with blame for the viewing of a photograph. The study's design deviated from the previous studies because, whereas a cyber peeper scenario could involve either a still image or a live streaming video, a peeping Tom scenario cannot be similarly manipulated.

Method

One hundred and fifty-two American participants (92 females, M_{age} = 37.8 years, SD = 12.5 years) recruited on Mechanical Turk were randomly assigned to read about either a cyber peeper who views a still photograph, a cyber peeper who views a live streaming video, or a peeping Tom. There are only three cells in this study because there is no simple way to remove live, real-time viewing from a genuine peeping Tom scenario. The three vignettes read as follows:

> Peeping Tom:
>
> Jane is changing clothes in her hotel room. Unbeknownst to her, she can be seen from outside through a small gap in the ventilation duct. Tim, Jane's colleague who is waiting outside, has heard about the gap in the ventilation duct and realizes that if he were to look through he'd see Jane in the nude. He decides to find the gap to look through, and he sees Jane in the nude.
>
> Cyber Peeper + Still Photo:
>
> Jane is changing clothes in her hotel room. Unbeknownst to her, hackers have compromised her laptop which is on the hotel desk, and they have remotely activated her webcam. The hackers are able to take a nude photograph using the webcam. Soon after, the nude photo of Jane is accessible on the Internet. Tim, Jane's colleague, has

heard about the nude photo and a website where it can be seen. He decides to visit the website, and he sees the photo of Jane in the nude.

Cyber Peeper + Live Stream:

Jane is changing clothes in her hotel room. Unbeknownst to her, hackers have compromised her laptop which is on the hotel room desk, and they have remotely activated her webcam. The hackers are able to watch her in real time while she is in the nude. They make a live streaming video of Jane nude accessible on the Internet. Tim, Jane's colleague, has heard about the live stream of Jane in the nude and a website where it can be accessed. He decides to visit the website, and he sees Jane in the nude in real time.

Results and Discussion

A three level ANOVA allowed us to test three contrasts. The first contrast showed that, again, participants blamed the peeping Tom (M=9.15, SD=2.02) more than the peepers in the two cyber peeping scenarios (M=6.33, SD=2.85); $F(1, 149) = 39.96$, $p < 0.0001$, $\eta_p^2 = 0.212$. The second contrast compared the two cyber peepers. This contrast showed that the cyber peeper was not blamed significantly less for viewing a still photograph (M=6.07, SD=2.87) than for viewing a live streaming video (M=6.59, SD=2.54); $F(1, 149) = 1.01$, $p = 0.317$, $\eta_p^2 = 0.007$. A third contrast – not orthogonal to the first – confirmed that the peeping Tom was blamed significantly more (M=9.15, SD=2.02) than was the live-streaming cyber peeper (M=6.59, SD=2.96); $F(1, 149) = 24.55$, $p < 0.0001$, $\eta_p^2 = 0.141$.

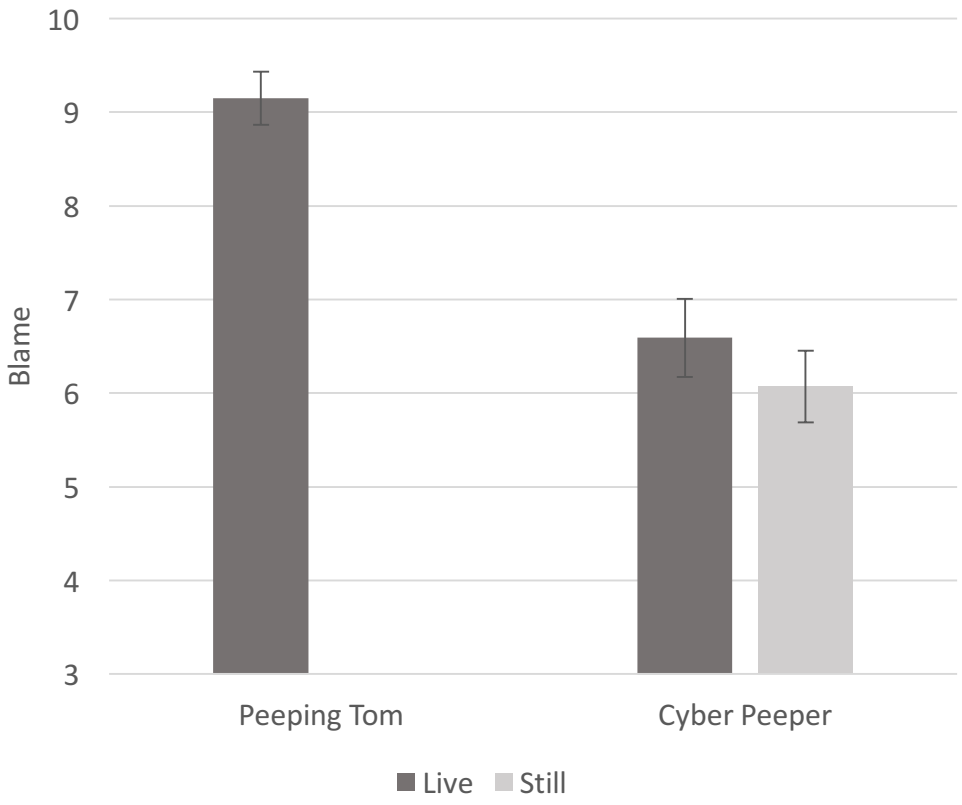

Figure 3.7. Mean blame for still photo versus live viewing. Error bars show standard error of the mean.

The results suggest that the difference between viewing a still photo and viewing a live streaming video does not meaningfully capture the blame-relevant difference between a cyber peeper and peeping Tom that has been observed in each study.

Study 8: Nearer versus more distant peeping

A cyber peeper can and usually would do his peeping at some distance from his victim, whereas a regular peeping Tom would necessarily be nearby. Study 8 tests whether spatial distance matters; that is, would the cyber peeper be considered more blameworthy if closer to the victim, and would the peeping Tom be considered less blameworthy if more distant. The spatial separation of distant peeping may enough to mitigate blame if it is

perceived as less of an invasion of privacy and thus less harmful, or it could decrease perceptions of the peeper's causal influence over any harmful outcomes (Hart & Honore, 1959; Prosser & Wade, 1971).

Method

Two hundred and fifty-nine American participants (152 females, M_{age} = 36.0 years, SD = 12.6 years) recruited on Mechanical Turk were randomly assigned to read about either a cyber peeper or a peeping Tom, and about either near or more distant peeping. Although a cyber peeping scenario, in which the Internet-aided peeping would typically occur at some distance, can be easily modified so that the peeping occurs at close proximity, peeping Tom scenarios are limited to the range of distance in which peeping is possible with a direct line of sight. For this reason, for the both the cyber peeper and the peeping Tom, distant peeping involved peeping from a room in a different hotel located across the street from the victim, whereas near peeping involved peeping from the hotel room adjacent to the victim. The four modified vignettes read as follows:

> Peeping Tom + Near:
>
> Jane is changing clothes in her hotel room. Tim, Jane's colleague who is staying in the room next door, knows that Jane is in her room undressing. Tim notices a gap in the ventilation duct to Jane's room and realizes that if he were to look through he'd see Jane in the nude. He decides to look through and sees Jane in the nude.
>
> Peeping Tom + Distant:
>
> Jane is changing clothes in her hotel room. Tim, Jane's colleague who is staying at a different hotel across the street, knows that Jane is in her room undressing. Tim notices that the previous guest left behind their binoculars. Tim realizes that with the binoculars he will be able to look out of his hotel room window, zoom in on Jane's hotel room window across the street, and be able to see Jane in the nude. Tim decides to look and sees Jane in the nude.
>
> Cyber Peeper + Near:

Jane is staying at a hotel. Tim, Jane's colleague who is staying in the room next door, learns that somebody has hacked into Jane's laptop and made her nude photos of herself accessible over the Internet. Tim decides to look and sees Jane in the nude.

Cyber Peeper + Distant:

Jane is staying at a hotel. Tim, her colleague who at a different hotel across the street, learns that somebody hacked into Jane's laptop and has made her nude photos of herself available over the Internet. Tim decides to look and sees Jane in the nude.

Results and Discussion

A 2x2 ANOVA showed again that participants blamed the peeping Tom (M=9.45, SD=1.04) more than the cyber peeper (M=6.38, SD=3.26); $F(1, 255) = 102.56$, $p < 0.0001$, $\eta_p^2 = 0.287$). The perpetrator was not blamed significantly less when peeping from a greater distance (M=7.97, SD=2.95) than when peeping from nearby (M=7.81, SD=2.81); $F(1, 255) = 0.18$, $p > 0.6$, $\eta_p^2 < 0.001$. There was also no interaction between the type of peeper and distance; $p > 0.5$.

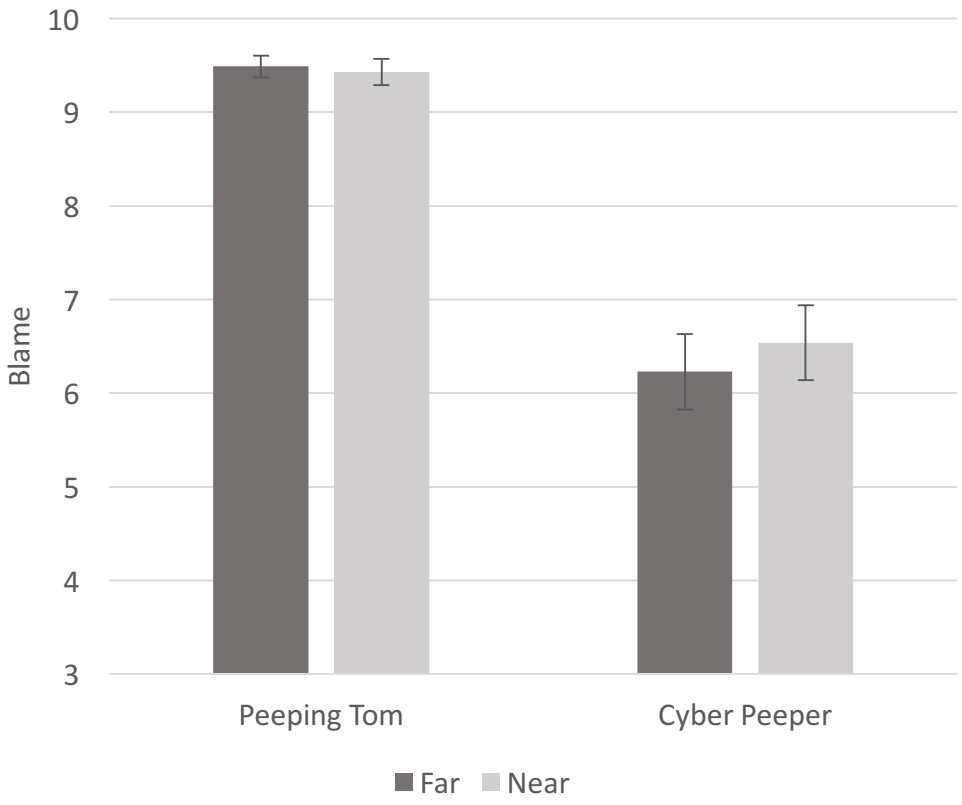

Figure 3.8. Mean blame attributed for nearer versus more distant peeping. Error bars show standard error of the mean.

The results suggest that the distance between the peeper and the victim does not account for the difference in blame for the cyber peeper and the peeping. It is possible that distance could impact blame for cyber peeping at distances greater than those possible for a peeping Tom, but putting the cyber peeper as close as a peeping Tom does not come close to equalizing blame.

Study 9: Middleman versus no middleman

Many Internet-facilitated transgressions, including the cases of cyber peeping we have used thus far, involve a middleman. For example, in the cyber peeping case somebody had hacked private accounts to gain access to the nude photos and initiate the leak. By

contrast, a peeping Tom might be more likely to need to create his own opportunity. Study 8 tests whether peepers are blamed less when a middleman facilitates or participates in bringing about the opportunity to peep. Such involvement could mitigate blame by diminishing the perceived unique causal contribution of the peeper (Alicke, 2000; e.g., Kelley, 1972), diffusing responsibility between the hacker and peeper (e.g., Darley & Latane, 1968).

Method

Two hundred and four American participants (115 females, M_{age} = 35.5 years, SD = 11.1 years) recruited on Mechanical Turk were randomly assigned to read about either a cyber peeper or a peeping Tom, and either about the involvement of a middleman who facilitates the opportunity to peep, or that no middleman was involved. The original cyber peeping case used previously involved a hacker as a middleman who makes the photos available. To create a cyber peeping case without a middleman, we created a scenario where the peeper must, on his own, exploit a security flaw in order to gain access to the photos. We introduce a middleman to the peeping Tom scenario who is described as a third person who has previously pried open a gap in the ventilation duct to the victim's room. This is contrasted with a peeping Tom scenario where the peeper pries open the gap himself. The four modified vignettes read as follows:

> Peeping Tom + Middleman:
>
> Jane is changing clothes in her hotel room. Tim, Jane's colleague, knows that Jane is in her room undressing. Tim learns that somebody has pried open a gap in the ventilation duct to Jane's room so that she can be seen from outside. Tim decides to look through and sees Jane in the nude.
>
> Peeping Tom + No Middleman:

Jane is changing clothes in her hotel room. Tim, Jane's colleague, knows that Jane is in her room undressing. Tim also knows that he can pry open a gap in the ventilation duct to Jane's room and see her in the nude. Tim decides to pry open the gap and sees Jane in the nude.

Cyber Peeper + Middleman:

Jane is staying at a hotel. Tim, her colleague who is back at their work offices, learns that somebody has hacked into Jane's laptop over the Internet by exploiting her hotel's unsecured network, and has made nude photos of Jane available on the Internet. Tim decides to access the photos and sees Jane in the nude.

Cyber Peeper + No Middleman:

Jane is staying at a hotel. Tim, her colleague who is back at their work offices, knows that he can find a way to hack into Jane's laptop over the Internet to see her private photos by exploiting her hotel's unsecured network. Tim decides to access the photos and sees Jane in the nude.

Results and Discussion

An 2x2 ANOVA showed again that participants blamed the peeping Tom (M=9.47, SD=1.38) more than the cyber peeper (M=8.36, SD=2.67); ($F(1, 200) = 16.08$, $p < 0.0001$, $\eta_p^2 = 0.074$). In addition, the perpetrator was blamed less when the opportunity to peep was facilitated by a middleman (M=8.29, SD=2.48) than when no middleman was involved (M=9.56, SD=1.64); $F(1, 200) = 21.34$, $p < 0.0001$, $\eta_p^2 = 0.096$. The interaction between the type of peeper and whether a middleman was involved was significant such that middleman involvement influenced blame more in the cyber peeping case than in the peeping Tom case $F(1, 200) = 12.09$, $p < 0.001$, $\eta_p^2 = 0.057$.

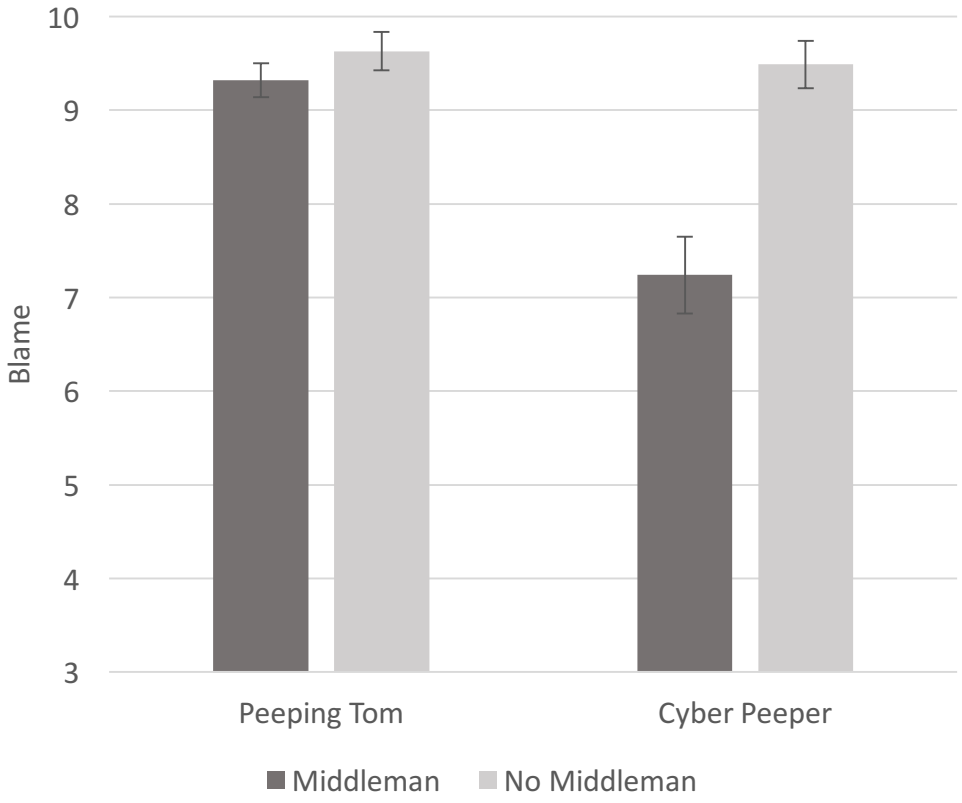

Figure 3.9. Mean blame attributed when middleman involved versus when no middleman involved. Error bars show standard error of the mean.

These results suggest that middleman involvement can measurably decrease blame for the peeper. The interaction effect may suggest that our manipulation of middleman involvement in the case of the cyber peeper had a stronger impact on blame than did our manipulation in the case of the peeping Tom. This could be because, in addition to peeping itself, the perpetrators in the cases without a middleman engaged in a second transgression - either computer hacking or vandalism - and the computer hacking could be considered more blameworthy than the vandalism.

Study 10: Matched Comparison

Study 10 aims to match a cyber peeping scenario with a peeping Tom scenario on the four contextual factors found to measurably impact blaming (i.e., celebrity, effort, prurience, & middleman) in order to see whether, when matched, the peepers are considered equally blameworthy. Although the no-middleman cyber peeper was already blamed no less that the peeping Toms in Study 9, Study 10 aims to match blame for a peeping Tom and a cyber peeper where both scenarios involve a middleman, such that a second blameworthy transgression need not be introduced for the cyber peeper.

Method

Ninety-three American participants (52 females, M_{age} = 35.5 years, SD = 13.1 years) recruited on Mechanical Turk were randomly assigned to read about either a cyber peeper or a peeping Tom, where both vignettes were matched with a celebrity victim, and low effort, non-prurient, middle-man facilitated peeping. The two modified vignettes read as follows:

> Peeping Tom:
>
> Jane, a famous actress, is sun bathing on the private patio outside her hotel room when she decides to remove her bathing suit to avoid tan lines. Tim, who has recognized Jane at the hotel after seeing her most recent film, is walking along the hotel pathway. Tim sees Jane's bathing suit draped over her patio fence and realizes that she is sunbathing nude. He becomes aware that, right beside him, somebody had pried open a gap in the patio fence through which Jane can be seen sunbathing in the nude. Tim is not sexually attracted to Jane and is only attracted to men, but he is still curious to see. Tim decides to look through and he sees Jane in the nude.
>
> Cyber Peeper:
>
> Jane, a famous actress, is changing clothes in her hotel room when she decides to use her cell phone to take a nude photo of herself. Tim, who knows Jane from seeing her most recent film, becomes aware that somebody has hacked the contents of Jane's cell phone, and that her nude photo was leaked onto the Internet. Tim is not sexually attracted to Jane and is only sexually attracted to men, but he is still curious to see. He decides to visit the website with the photo and he sees Jane in the nude.

Results and Discussion

Participants did not blame the cyber peeper (M=6.08, SD=3.29) significantly less than the peeping Tom (M=6.63, SD=3.01); $t(91) = 0.83$, $p = 0.41$, 95% CI = [-0.76, 1.84], $d = 0.173$.

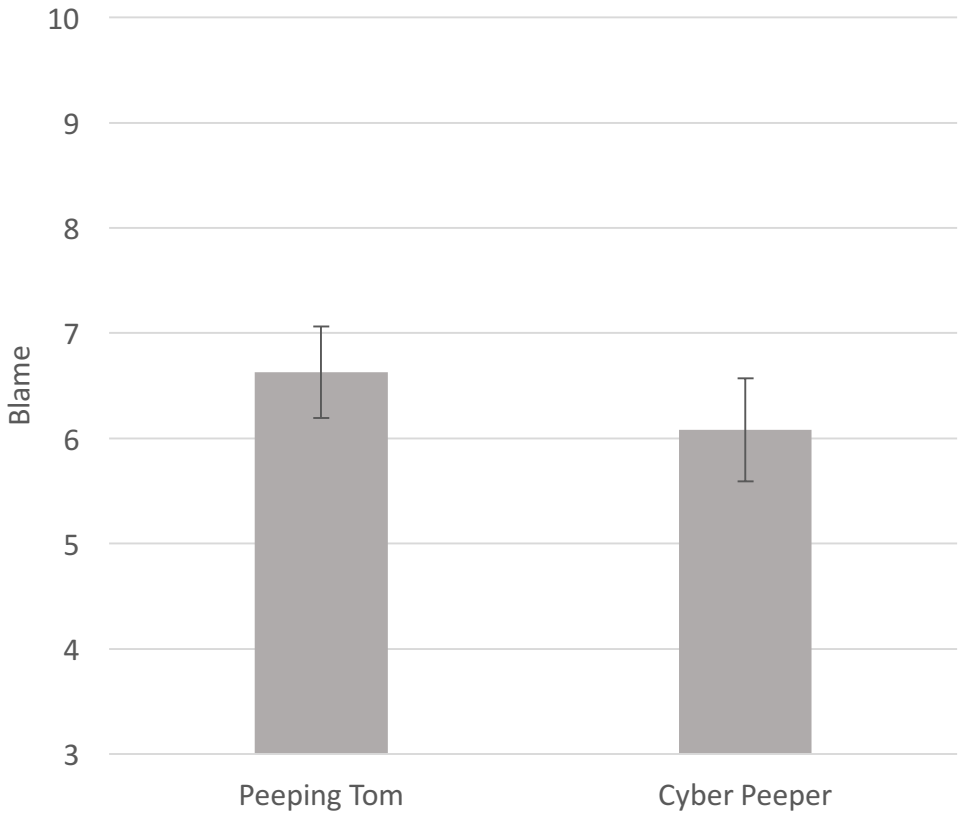

Figure 3.10. Mean blame for Peeping Tom and Cyber Peeper roughly matched on four factors. Error bars show standard error of the mean.

The results suggest that, in combination, the four factors identified in Studies 2, 4, 5, and 9 - namely, victim celebrity status, effort, prurience, and middleman involvement – are sufficient to account for the majority, if not all, of the difference in blame observed in Study 1. The modifications from Study 1 to Study 11 reduced the Cohen's d effect size of cyber

peeper versus peeping Tom from 1.28, conventionally regarded as very large, to a statistically non-significant 0.17, conventionally regarded as very small.

Study 11: Internet photo versus printed photo

The results of Study 10—which showed that a cyber peeper and peeping Tom can be considered equally blameworthy when matched on the four identified factors—suggest that it is not precisely the involvement of Internet and computer technology itself that mitigates blame for cyber peepers relative to peeping Toms, but rather other factors with which the technology is typically associated. By comparing blame for the viewing of a private nude photo on the Internet with blame for the viewing of a private nude photo printed on paper, Study 11 examines whether the involvement of the Internet itself has any impact on blaming.

Method

One hundred and twenty-one American participants (72 females, M_{age} = 36.3 years, SD = 12.4 years) recruited on Mechanical Turk were randomly assigned to read about either a cyber peeper who views an online photo or a peeper who views a physical photo printed on paper. The two modified vignettes read as follows:

> Printed Photo:
>
> Each year, on their anniversary, Jane likes to surprise her boyfriend at work with a private and sexy, nude photo of herself. Her boyfriend has never shown these pictures to anybody and never would, but he has told some of his colleagues about this anniversary tradition.
>
> As in previous years, Jane folds up her photo, then writes on it, "Dear Mike, Happy anniversary! For your eyes only xxx." Then she slides it under the door of the office to which only her boyfriend has access. Unfortunately, this time Jane mistakenly slides her photo under the wrong door and into a common space to which many employees have access.
>
> Tim, her boyfriend's colleague, sees the folded photo that was slid under the door. He sees the message and realizes that it was meant for Jane's boyfriend and that unfolding it will reveal a nude photo of Jane. He decides to unfold it to see the photo.

Online photo:

Each year, on their anniversary, Jane likes to surprise her boyfriend at work with a private and sexy, nude photo of herself. Her boyfriend has never shown these pictures to anybody and never would, but he has told some of his colleagues about this anniversary tradition.

As in previous years, Jane again sends an email to her boyfriend. The email includes a link to her nude photo and has the subject line, 'Dear Mike, Happy anniversary! For your eyes only xxx.' Unfortunately, this time Jane mistakenly mistypes her boyfriend's email address. The company's server is set up so that emails that correctly include the company domain name, @hst.com, but which are not sent to a valid email address will automatically arrive in a general inbox to which many employees have access.

Tim, her boyfriend's colleague, sees the email with the link to the photo. He sees the subject and realizes that the email was meant for Jane's boyfriend and that the link will reveal a nude photo of Jane. He decides to click on the link to see the photo.

Results and Discussion

Participants blamed the peeper who viewed the Internet photo (M=6.13, SD=3.26) no less than the peeper who viewed the printed photo (M=6.19, SD=3.57); $t(119) = 0.01$, $p > 0.900$, 95% CI = [-1.16, 1.30], $d = 0.020$.

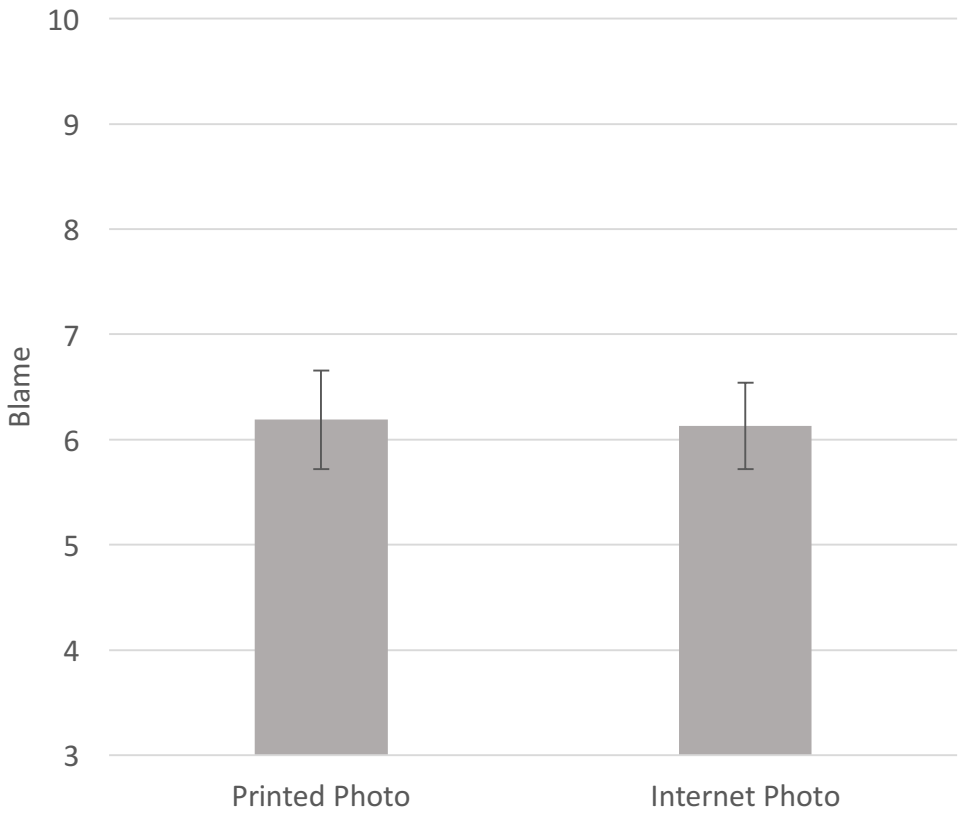

Figure 3.11. Mean blame attributed for viewing internet photo versus printed photo. Error bars show standard error of the mean.

The results suggest that while the exoneration of blame for the cyber peeper relative to the peeping Tom may be due to factors that may accompany Internet facilitated peeping - such as victim celebrity status, low effort, non-prurient curiosity, and the involvement of a middleman - it appears not to be the result of the involvement of Internet technology itself.

General Discussion

Across 11 studies, we confirmed that cyber peepers are typically blamed less than peeping Toms. We examined whether blaming was influenced by each of eight contextual factors that differ between the cases of the cyber peeper and the peeping Tom. We found no evidence that blaming is impacted by differences between higher or lower victim risk-taking,

by the number of other transgressors, by increased or decreased spatial distance, or by still photo versus real-time, live viewing. On the other hand, blaming was impacted by differences in victim celebrity status, high versus low effort peeping, prurient versus non-prurient intention, and middleman involvement. Moreover, the combination of these factors seemed to account for much of the discrepancy in ascribed blame, such that scenarios matched on these factors yield a cyber peeper and a peeping Tom that are perceived as similarly blameworthy.

In looking for commonality among the factors that successfully influenced blame, we note that both effort and prurience can be understood in relation to the perpetrator's intention and a spontaneous evaluation of his character (Alicke, 2000; Uhlmann, Pizarro, Diermeier, 2015). And among those factors that failed to significantly impact blame, both the number of other transgressors and victim risk-taking can be thought of in terms of the inclusion of blameworthy others; notably, the blameworthiness of these others did not detract from or dilute the blameworthiness of the perpetrator, as might be expected by considering the potentially reduced uniqueness of his causal contribution (Alicke, 2000), from a loose interpretation of Darley and Latane's (1968) notion of diffused of responsibility, or from the consensus component of Kelley's (1972) covariation model of causal attribution.

The results of Study 11 suggest that blame is not impacted by the involvement of the Internet itself, but rather by some of the factors with which it is associated. Similar Internet-associated factors may impact perceptions of blameworthiness for other cyber offenses, including online theft and cyber bullying.

One important limitation of the data collected was the presence of ceiling effects. These could limit our ability to assess some additional impacts of our manipulations that

might have emerged if not for the high average ratings of blameworthiness. However, the factors that were indeed deemed relevant to blame were deemed so despite any ceiling effects rather than due to any ceiling effects. Moreover, it is a feature of regular peeping Tom behaviors that they are blamed harshly rather than a peculiarity of our experimental scenarios, and to modify our scenarios too much would be to modify our research question.

This project has sought to describe and explain a discrepancy in ordinary blame ascription in terms of Internet-mediated contextual differences while neither taking nor espousing a normative position. Still, normative discussions can benefit from the understanding afforded by these and other such findings. A better understanding of intuitive ethics regarding online activities not only informs legislative and normative questions, but may be beneficial to those building the programs and platforms that connect people online. Such systems can facilitate not only communication and connection, but ethically questionable and immoral behaviors as well.

Care should be taken to avoid conflating a potential decrease in perceived perpetrator blameworthiness with a decrease in the harmfulness of an offense and the need to prevent it. Cyber peeping, along with other cyber offenses, may be no less harmful than their in-person analogs, and to the extent that they are more common and more readily engaged in, they may ultimately be even more harmful.

Future research could examine by which blame relevant perceptions these factors have their impact on blame attributions, and why only some of the tested factors influence those perceptions while others do not.

Chapter 3, in full, is currently being prepared for submission for publication of the material. Jago, Carl P.; Ryazanov, Arseny A.; Christenfeld, Nicholas J. S. The dissertation author was the primary investigator and author of this chapter.

References

Abcarian, R. (2014, October 9). Why Jennifer Lawrence is so wrong about her stolen nude photos. Los Angeles Times. Retrieved from http://www.latimes.com/local/abcarian/la-me-ra-why-jennifer-lawrence-is-wrong-20141009-column.html

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

Bailey, F. L., & Rothblatt, H. B. (1985). Successful techniques for criminal trials. Rochester, NY: Lawyers Co-Operative.

Baird, J. A., & Astington, J. W. (2004). The role of mental state understanding in the development of moral cognition and moral action. *New Directions for Child and Adolescent Development*, 103, 37–49.

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of personality and social psychology*, 54(4), 569.

Bayles, M. (1982) Character, purpose and criminal responsibility. Law and Philosophy, 1, 5-20.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, 108(2), 353-380.

Cushman, F. A., Dreber, A., Wang, Y., & Costa, J. (2009). Accidental outcomes guide punishment in a 'trembling hand' game. *PLOS One*, 4(8), e6699. doi:6610.1371/journal.pone.0006699.

Darley, J. M., & Latane, B. (1968). Bystander intervention in emergencies: diffusion of responsibility. Journal of personality and social psychology, 8(4p1), 377.

Fallon, J. (Writer). (2014, September 2). S1 E115 [Television series episode]. In The Tonight Show Starring Jimmy Fallon. New York: NBC.

Gino, F., Moore, D. A., & Bazerman, M. H. (2009). No harm, no foul: The outcome bias in ethical judgments.

Gino, F., Shu, L., & Bazerman, M. (2010). Nameless + harmless = blameless: When seemingly irrelevant factors influence judgment of (un)ethical behavior. *Organizational Behavior and Human Decision Processes*, 111(2), 93–101.

Hart, H. L. A., & Honore, A. M. (1959). Causation in the law. Oxford, England: Oxford University Press.

Horan, H. D., & Kaplan, M. F. (1983). Criminal intent and consequence severity: Effects of moral reasoning on punishment. *Personality and Social Psychology Bulletin*, 9(4), 638-645).

Kassin, S., & Garfield, D. (1991). Blood and guts: General and trial-specific effects of videotaped crime scenes on mock jurors. Journal of Applied Social Psychology, 21, 1459-1472.

Kerr, N. L. (1978). Severity of prescribed penalty and mock jurors' verdicts. Journal of Personality and Social Psychology, 36, 1431-1442.

Kanekar, S., Kolsawalla, M., & D'Souza, A. (1981). Attribution of responsibility to a victim of rape. *British Journal of Social Psychology*, 20, 165-170.

Karuza, J., Jr., & Carey, T. O. (1984). Relative preference and adaptiveness of behavioral blame for observers of rape victims. *Journal of Personality*, 52, 249-260.

Kashner, S. (2014, October 20). Both Huntress and Prey. Vanity Fair. Retrieved from https://www.vanityfair.com/hollywood/2014/10/jennifer-lawrence-photo-hacking-privacy

Kelley, H. H. (1972). Attribution in social interaction. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), Attribution: Perceiving the causes of behavior (pp. 1-26). Morristown, NJ: General Learning Press.

Killen, M., Mulvey, K. L., Richardson, C., Jampol, N., & Woodward, A. (2011). The accidental transgressor: Morally-relevant theory of mind. *Cognition*, 119(2), 197–215.

Knobe, J. (2006). The concept of intentional action: A case study in the uses of folk psychology. *Philosophical studies*, 130(2), 203-231.

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A Theory of Blame. *Psychological Inquiry*, 25(2), 147-186.

Mazzocco, P., Alicke, M., & Davis, T. (2004). On the robustness of outcome bias: No constraint by prior culpability. *Basic and Applied Social Psychology*, 26(2), 131–146.

Mendelson, S. (2014, September 03). Jennifer Lawrence Nude Photo Leak Isn't A 'Scandal.' It's A Sex Crime. Retrieved July 16, 2017, from https://www.forbes.com/sites/scottmendelson/2014/09/01/jennifer-lawrence-nude-photo-leak-isnt-a-scandal-its-a-sex-crime/#7e2089274238

Muehlenhard, C. L., Friedman, D. E., & Thomas, C. M. (1985). Is date rape justifiable? The effects of dating activity, who initiated, who paid, and men's attitudes toward women. *Psychology of Women Quarterly*, 9, 297-310.

Mccoy, T. (2014, September 3). 4chan: The Shock Post Site That Hosted the Private Jennifer Lawrence Photos. The Washington Post. Retrieved August 20, 2017, from https://www.washingtonpost.com/news/morning-mix/wp/2014/09/02/the-shadowy-world-of-4chan-the-shock-post-site-that-hosted-the-private-jennifer-lawrence-photos

Nadler, J., & Rose, M. R. (2002). Victim impact testimony and the psychology of punishment. *Cornell L. Rev*., 88, 419.

Pizarro, D. A., Uhlmann, E., & Bloom, P. (2003). Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology*, 39(6), 653-660.

Pizarro, D., Uhlmann, E., & Salovey, P. (2003). Asymmetry in judgments of moral blame and praise: The role of perceived metadesires. *Psychological Science*, 14(3), 267-272.

Prosser, W. L., & Wade, J. W. (1971). Torts: Cases and materials. Mineola, NY: Foundation Press.

Searle, J. (1983). Intentionality. Cambridge, UK: Cambridge University Press.

Schult, D. G., & Schneider, L. J. (1991). The role of sexual provocativeness, rape history, and observer gender in perceptions of blame in sexual assault. *Journal of Interpersonal Violence*, 6, 94-101.

Scroggs, J. R. (1976). Penalties for rape as a function of victim provocativeness, damage, and resistance. *Journal of Applied Social Psychology*, 6, 360-368.

Schein, C., & Gray, K. (2014). The prototype model of blame: Freeing moral cognition from linearity and little boxes. *Psychological Inquiry*, 25(2), 236-240.

Spiegel, I. O. (1957). Public celebrity v. scandal magazine--the celebrity's right to privacy. *Southern California Law Review* 30(3), 280-312.

Steinberg, J. (2014, September 24). Can You Be Arrested - Or Sued - For Viewing Or Sharing Nude Celebrity Photos? Here Are Some Relevant Laws. Retrieved August 20, 2017, from https://www.forbes.com/sites/josephsteinberg/2014/09/24/can-you-be-arrested-or-sued-for-viewing-or-sharing-nude-celebrity-photos-here-are-some-relevant-laws-2

Sutter, M. (2007). Outcomes versus intentions: On the nature of fair behavior and its development with age. *Journal of Economic Psychology*, 28(1), 69-78.

Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science*, 10(1), 72-81.

Wadhwa, V. (2014, September 19). Laws and Ethics Can't Keep Pace with Technology. Retrieved November 28, 2017, from https://www.technologyreview.com/s/526401/laws-and-ethics-cant-keep-pace-with-technology/

CHAPTER 4

WHAT WOMEN SAY THEY WANT VERSUS WHAT MEN IMAGINE THEY DO

Carl P. Jago & Karen R. Dobkins

Abstract

To appeal to the opposite gender, previous research indicates that men emphasize their wealth, status, and ambition, whereas women emphasize their physical attractiveness. Such behavior seems surprising given previous surveys in which men and women reported these traits to be less important than others such as trustworthiness, intelligence, and warmth. We addressed one potential reason for any disconnect, which is that men's and women's beliefs about what the opposite gender prefers are misguided—according to the opposite genders' self-reports. Using a new method, we asked participants to both self-report the traits they prefer in a romantic partner and to indicate what they imagine the opposite gender prefers. The results reveal striking discrepancies between what people report wanting in a potential partner and what the opposite gender imagines they want. Additionally, women appear to be better at imagining men's preferences, and we discuss several reasons why this might be the case.

WHAT WOMEN SAY THEY WANT VERSUS WHAT MEN IMAGINE THEY DO

For decades, surveys have shown that heterosexual men and women around the world report that traits such as wealth, status, or physical attractiveness in a potential romantic partner are relatively less important than traits such as trustworthiness, intelligence, and warmth (e.g. Buss & Barnes, 1986; Fletcher, Simpson, Thomas, & Giles, 1999; Lippa, 2007). However, other studies show that men and women attempt to impress each other with an emphasis on the relatively less important traits (e.g., Buss 1988). For example, rather than focusing on trustworthiness, intelligence, and warmth, men are likely to display cues indicating wealth or earning potential, such as by advertising their professional status, flashing money, or driving an expensive car. In addition, to derogate rivals, men tell women that their male rivals are poor, lack ambition, or drive cheap cars. Likewise, for women, rather than focusing on trustworthiness, intelligence, and warmth, they attempt to display their physical appeal, and derogate rivals by telling men that their female rivals are fat, have shapeless bodies, or are otherwise physically unattractive (Buss & Dedden 1990; Campos, Otta, de Oliviera Siqueira, 2002; Cicerello & Sheehan, 1995; Fletcher, Simpson, Campbell, & Overall, 2013). Men and women also expect a potential romantic partner to evaluate them more negatively after that potential partner has been exposed to others of the same gender who are, in the case of men, ambitious or independent, and in the case of women, physically attractive (Gutierres, Kenrick, & Partch, 1999). In addition, men and women appear to base their sense of self-worth on wealth/status and physical attractiveness, respectively (Crocker, Luhtanen, Cooper, & Bouvrette, 2003; Daly, Hogg, Sacks, Smith, & Zimring, 1983; Josephs,

Markus, & Tafarodi, 1992). However, outside of a dating context, both men and women indicate that traits related to warmth are more important to their self-concept than those related to either status or attractiveness (Campbell & Wilbur, 2009). It is unclear why people emphasize the less important traits in the context of romantic relationships. The potential causes for this disconnect are many, and they include the possibility that the less important traits are easier than the more important traits to signal and criticize. But perhaps the most obvious possible cause for the disconnect is that men's and women's beliefs about what the opposite gender prefers are discrepant from self-reported preferences.

To address this possibility, the primary goal of the current study is to investigate discrepancies between men's and women's self-reported preferences about traits in the opposite gender with what people *imagine* are the opposite gender's preferences. A secondary aim is to determine which gender's beliefs about the opposite gender's preferences are closest to the opposite gender's self-reported preferences. To our knowledge, the current study is the first to compare self-reported preferences with preferences imagined by the opposite gender, and to do so within the same population. In addition, our method allows us to quantify preferences in a way that supplements the existing literature's survey data by providing a more nuanced characterization of discrepancies between genders.

Method

Three hundred and fifty American participants recruited on Mechanical Turk completed an online survey designed to measure what heterosexual men and women report as preferences in romantic partners of the opposite gender, as well as what they imagine the opposite gender prefers in them. Thirty-four participants who did not identify as heterosexual were excluded leaving 316 participants (50.9% female, M_{age} = 38.3 years, SD =

11.8 years; $M_{age\ females} = 38.0$, $M_{age\ males} = 38.5$; similar average age, therefore could be considered potential dating partners) for analysis. We chose this sample size as it provides ample power (80%) to minimize the risk of type 1 and type 2 errors in detecting small to medium effect sizes ($d = 0.3$) and greater when comparing the difference between two independent means.

Before proceeding with the methods, we highlight our novel approach for characterizing preferences. First, when we measure the self-reported and imagined impacts of each trait on a person's desirability as a romantic partner, we do so at multiple levels (i.e., if the trait is "wealth", we ask how one feels about dating a partner whose wealth is far below, somewhat below, somewhat above and far above, average). Measuring impacts at multiple levels allows us to characterize non-linearities in men's and women's self-reported preferences. Second, because the levels are standardized across traits, this allows us to compare impacts on desirability across different and ordinarily incommensurable trait dimensions; where each dimension is typically measured on different scales with different standard units of measurement. Finally, we can use this method to calculate an "importance score" by averaging the absolute impacts across multiple levels of the trait, which aligns with the lay definition of importance, where a more important trait is one with a greater impact on desirability. Together, these key aspects allow us to compare the importance and any nonlinearities of different traits within gender, across gender, and between self-reported and imagined preferences.

After reporting their age and gender, each participant answered *two* blocks of questions regarding 22 different physical/personality traits (one block for self-reported preferences, another for imagining the opposite gender's preferences). Table 4.1 displays the

list of traits used in the current study. We randomized the order of the two blocks and 22 trait questions across participants. Twenty-one of the twenty-two traits were identical for men and women participants. Only one trait for each gender was different, although complementary; "penis size" as a male trait and "breast size" as a female trait. All questions were identical for men and women participants, with the exception of gender words ("he" vs. "she", "him" vs. "her", "men" vs. "women"). For simplicity, we present the instructions and questions that only female participants saw.

The block that aimed to measure participants' self-reported preferences began with general instructions: "For each of these questions, you will imagine that you learn something about how a potential romantic partner compares to average for a particular trait and how this affects your desire to date him. You should imagine that you previously expected he would be about average among men for each trait." Following that, for each of the 22 traits, we asked the following: "Imagine discovering that a potential romantic partner was _____, how would this affect your desire to date him?", with a horizontal response slider (one decimal resolution) anchored on the left side with -2 (much less likely to want to date) and on the right side +2 (much more likely to want to date) (Figure 4.1). The blank could be filled in with 4 different levels of that particular trait. For example, if the trait were *intelligence*, we asked the participants how more or less likely it would be that they would want to date someone who was 1) "far less intelligent than average", 2) "somewhat less intelligent than average", 3) "somewhat more intelligent than average", and 4) "far more intelligent than average". This block therefore had a total of 88 response sliders (22 traits x 4 levels). Note that we refer to participant responses as "impact scores", because participants are responding to how a trait would impact their willingness to date a person who possesses a certain level of

the trait. We use the term "self-reported preferences" to capture the pattern of impact scores across the four levels.

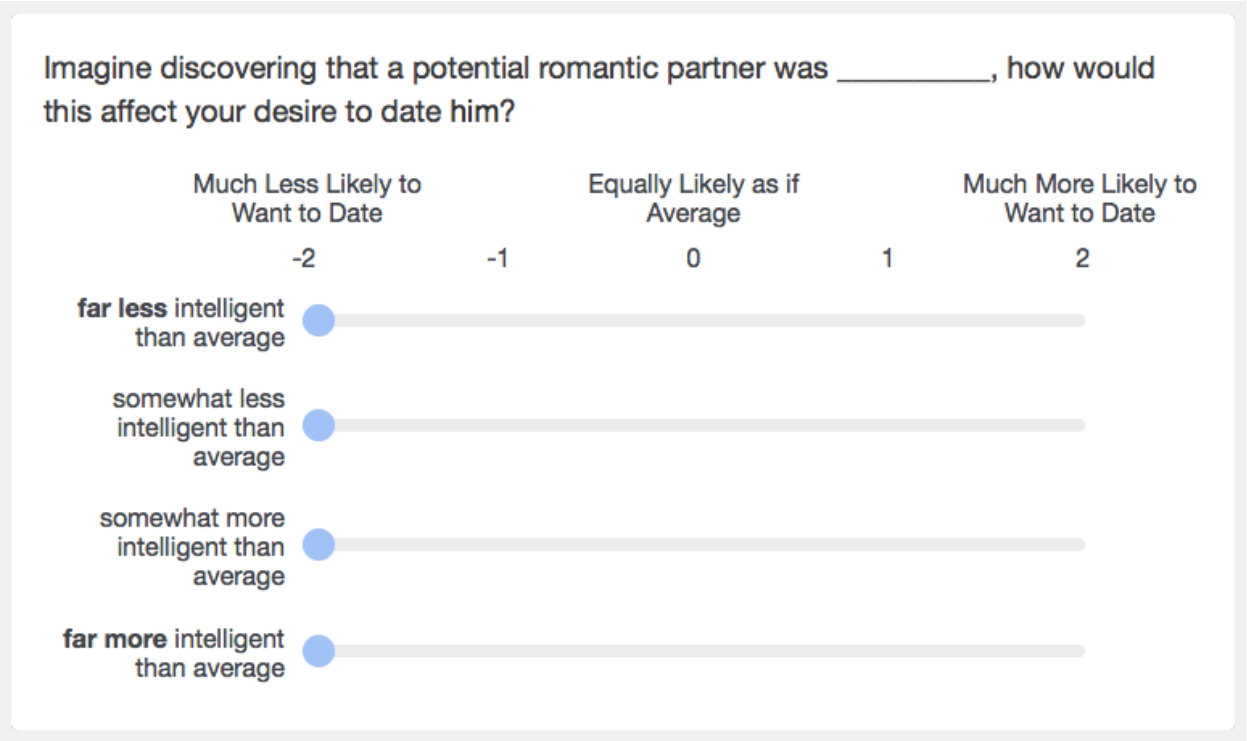

Figure 4.1. Example of question with accompanying sliders to survey the impact of a potential romantic partner's intelligence on a women's romantic interest in him.

The other block measured what participants imagine members of the opposite gender prefer in a romantic partner. This began with general instructions as follows: "Imagine how heterosexual men would respond to a woman with each trait below, compared with how they would respond to a woman who is average among women for that trait." Then, for each of the 22 traits, we asked participants: "How do you think heterosexual men's desire to date a potential romantic partner is affected if they discover that the woman is _____?" For each trait, the response slider and associated levels were identical to those in the self-reported block (see above). There were a total of 88 response sliders (22 traits x 4 levels). We use the

term “imagined preferences” to capture the pattern of imagined impact scores across the four levels.

Analysis

Calculating mean impacts and plotting zigzags

The analyses began with the participants’ raw self-reported and imagined impact scores. For each trait, we calculated the mean impact score for each of the four levels (far below average, somewhat below average, somewhat above average, and far above average), as well as for each of the four groups of preferences (those that men self-report, those that women self-report, those that men imagine for women, and those that women imagine for men). For each trait, we therefore calculated sixteen mean impact scores and used them to plot zigzags to visualize impact scores as a function of the different levels, plus average. We plotted average itself with an impact score of 0, as this is our reference point (see Figures 4.2 through 4.6 for noteworthy examples). Each plot allows for a visual comparison, and an estimate of the discrepancy, between 1) men’s and women’s self-reported preferences, or 2) men’s self-reported preferences and the preferences women imagine are true for men (and vice versa).

Calculating differences between men’s and women’s self-reported impact scores

For each trait, we calculated the absolute difference between men’s and women’s self-reported impact scores for each of the four levels. The average of these scores represents the overall difference between the zigzags, and therefore the overall difference between men’s and women’s preferences for any particular trait dimension. Analyses of importance and shape, described below, represent higher-level constructs of the zigzags.

Calculating and comparing importance scores

For each gender, we calculated the self-reported and imagined importance of each trait by averaging the absolute values of the mean impact scores across the four levels. In this way, traits with the highest mean absolute impact scores have the highest importance scores. For example, men's self-report data show that a partner's femininity is very important to them (blue line in Figure 4.2.a), while height is much less so (blue line in 4.2.c). Using the importance scores, for each trait we measured discrepancies between men's vs. women's self-reported importance scores, and discrepancies between self-reported vs. imagined importance scores. We calculated the statistical significance of each of these discrepancies using a permutation test that provides a robust alternative to inference based on parametric assumptions. For each test, there were 10,000 permutations—each with the impact scores randomly shuffled at each level across the two comparison groups so as to align with the null hypothesis by breaking any existing structure in the sample data. These tests yielded a *p* value representing the proportion of permutations wherein the resulting difference was at least as extreme as that which we observed in our data.

Comparing shapes

In addition to calculating importance scores, another construct that is visible in the zigzags is their shape, which relates to their linearity (i.e., whether they peak or plateau, and what could otherwise be considered changes in slope if they were instead functions of continuous variables). For each trait, to establish whether there was a difference in shape between the zigzags representing men's vs. women's self-reported preferences, and between the zigzags representing the self-reported and imagined preferences, we began by rescaling the impact scores from one gender to match the overall importance from the opposite gender.

Specifically, to compare men's vs. women' self-reported preferences, we rescaled men's data so that the importance score matched that of the women. This was achieved by multiplying each mean impact score for men by the ratio of women's self-reported importance over men's self-reported importance (The same rescaling was done when comparing zig-zags between self-reported preferences in one gender with imagined preferences in the other). Having removed any difference in importance, the remaining discrepancy between the zigzags represented only a discrepancy in shape. For example, shapes of self-reports for height are different for men vs. for women, whereas shapes of self-reports for financial debt are the same for men and for women, despite their difference in terms of importance (blue and red lines, respectively in Figure 4.2.a and 4.2.c, respectively). We determined statistically significant shape discrepancies using the same permutation testing paradigm discussed above. Unlike the analyses of importance (above), in which we calculate both importance scores and differences between those importance scores, in these analyses we calculate only discrepancies in shape because there is no shape score.

Asking which gender's beliefs are closer to the opposite gender's self-reports

As a first step, for each of the 22 traits we calculated the absolute difference between the self-reported and imagined mean impact scores, separately for men imagining women's preferences and women imagining men's preferences. We then asked if the mean of these values differed significantly between men and women using a Student's t-test. The gender with the lower mean discrepancy can be said to have outperformed the other, according to the opposite gender's self-reported preferences. In addition, we conducted a Pearson correlation between the self-reported and imagined importance scores for each trait, separately for men imagining women's preferences and women imagining men's preferences. This allowed us to

ask which gender appears to better estimate the relative importance of each trait to the opposite gender (according to the opposite gender's self-reports).

Results

Table 4.1 presents self-reported importance scores for men and women, as well as the discrepancy between the two, for each trait. The traits are ranked in order from most to least discrepant, regardless of the direction of the discrepancy. For each trait we also present the discrepancy between the two genders' shapes. Statistically significant discrepancies are depicted with asterisks.

Table 4.1. Importance scores of each trait dimension for women and men, the discrepancy between the two genders in importance, and in shape. Traits are ordered from most to least discrepant, regardless of the direction of the discrepancy (+ means more important to women, - means more important to men). Importance discrepancies are also each presented as a Cohen's *d* effect size. The *p*-value significance level of each importance discrepancy and of each shape discrepancy is indicated to the right of each (.1 > ~ > .05 > * > .01 > ** > .001 > *)**

Trait Dimension	Importance to Women	Importance to Men	Importance Discrepancy	Importance Discrep. *d*		Shape Discrepancy	
Height	0.754	0.191	0.563	1.119	***	0.224	***
Financial Debt	1.011	0.612	0.399	0.725	***	0.050	
Ambitious	1.016	0.672	0.344	0.685	***	0.049	
Sexual	0.538	0.835	-0.297	0.588	***	0.425	***
Romantic	0.858	0.610	0.248	0.479	***	0.013	
Charitable	0.711	0.470	0.241	0.466	***	0.047	
Breast(Penis) Size	0.513	0.279	0.234	0.430	***	0.255	***
Wealthy	0.570	0.348	0.222	0.431	***	0.053	
Social Status	0.485	0.271	0.213	0.452	***	0.056	
Outgoing	0.457	0.255	0.202	0.419	***	0.052	
Phys. Attractive	0.793	0.964	-0.171	0.346	***	0.149	*
Feminine	0.866	0.701	0.165	0.305	**	1.403	***
Funny/Playful	0.982	0.835	0.147	0.310	**	0.069	
Trustworthy	1.431	1.290	0.141	0.340	***	0.052	
Intelligent	1.072	0.931	0.141	0.291	**	0.087	
Educated	0.893	0.768	0.125	0.240	**	0.084	
Warm/Friendly	1.041	0.925	0.117	0.239	**	0.113	*
Nurturing	0.951	0.845	0.106	0.200	*	0.086	
Hygiene	1.178	1.074	0.104	0.255	*	0.075	
Aggressive	0.606	0.504	0.102	0.194	~	0.239	***
Masculine	0.734	0.809	-0.075	0.139		1.617	***
Athletic	0.587	0.630	-0.043	0.090		0.117	~

Table 4.2a presents self-reported importance scores for women alongside the importance scores that men imagine for women, and the discrepancy between the two (vice versa in Table 4.2b). Table 4.2a ranks traits from the highest to the lowest self-reported importance scores. Also presented is the discrepancy between self-reported and imagined importance scores. Statistically significant discrepancies are depicted with asterisks.

The traits most important to women are trustworthiness, hygiene, intelligence and warmth/friendliness. Least important are sexuality, penis size, social status, and least of all, extraversion ("outgoing"). The traits most important to men are trustworthiness, hygiene, physical attractiveness and intelligence. Least important are breast size, social status, extraversion, and least of all, height.

Table 4.2a Self-reported importance scores for women, importance scores that men imagine for women, and the discrepancy between the two. Traits are ranked from the highest to the lowest self-reported importance scores. Also shown are the discrepancies between self-reported shapes for woman and the shapes men imagine for women. Importance discrepancies are also each presented as a Cohen's *d* effect size. The *p*-value significance level of each importance discrepancy and of each shape discrepancy is indicated to the right of each (.1 > ~ > .05 > * > .01 > ** > .001 > *)**

Trait Dimension	Importance to Women	Imagined Importance	Importance Discrepancy	Importance Discrep. *d*		Shape Discrepancy	
Trustworthy	1.431	1.157	-0.275	0.649	***	0.109	
Hygiene	1.178	1.115	-0.063	0.155		0.136	**
Intelligent	1.072	0.886	-0.186	0.404	***	0.152	*
Warm/Friendly	1.041	0.975	-0.067	0.138		0.132	*
Ambitious	1.016	1.017	0.000	0.001		0.133	*
Financial Debt	1.011	0.714	-0.297	0.593	***	0.074	
Funny/Playful	0.982	0.998	0.016	0.035		0.099	~
Nurturing	0.951	0.799	-0.152	0.302	***	0.109	
Educated	0.893	0.914	0.021	0.040		0.078	
Feminine	0.866	0.745	-0.121	0.216	*	0.096	
Romantic	0.858	0.982	0.124	0.262	**	0.087	
Phys. Attractive	0.793	1.146	0.354	0.745	***	0.174	***
Height	0.754	0.948	0.194	0.373	***	0.106	*
Masculine	0.734	0.902	0.168	0.333	***	0.151	**
Charitable	0.711	0.621	-0.090	0.175	*	0.084	
Aggressive	0.606	0.310	-0.297	0.579	***	0.063	
Athletic	0.587	0.982	0.395	0.797	***	0.184	***
Wealthy	0.570	1.090	0.520	0.976	***	0.098	*
Sexual	0.538	0.681	0.143	0.291	**	0.192	***
Penis Size	0.513	0.824	0.311	0.576	***	0.193	***
Social Status	0.485	1.021	0.537	1.033	***	0.105	*
Outgoing	0.457	0.754	0.298	0.635	***	0.124	*

Table 4.2b. Self-reported importance scores for men, importance scores that women imagine for men, and the discrepancy between the two. Traits are ranked from the highest to the lowest self-reported importance scores. Also shown are the discrepancies between self-reported shapes for men and the shapes women imagine for women. Importance discrepancies are also each presented as a Cohen's *d* effect size. The *p*-value significance level of each importance discrepancy and of each shape discrepancy is indicated to the right of each (.1 > ~ > .05 > * > .01 > ** > .001 > *)**

Trait Dimension	Importance to Men	Imagined Importance	Importance Discrepancy	Importance Discrep. *d*		Shape Discrepancy	
Trustworthy	1.290	1.182	-0.109	0.232	*	0.079	
Hygiene	1.074	1.147	0.073	0.179	~	0.034	
Phys. Attractive	0.964	1.333	0.369	0.845	***	0.030	
Intelligent	0.931	0.506	-0.425	0.895	***	0.282	***
Warm/Friendly	0.925	0.986	0.061	0.133		0.016	
Nurturing	0.845	0.754	-0.091	0.173	*	0.059	
Funny/Playful	0.835	0.918	0.084	0.178	*	0.020	
Sexual	0.835	1.220	0.385	0.797	***	0.046	
Masculine	0.809	1.133	0.324	0.635	***	0.022	
Educated	0.768	0.452	-0.316	0.660	***	0.261	***
Feminine	0.701	0.926	0.225	0.459	***	0.071	
Ambitious	0.672	0.535	-0.138	0.278	**	0.121	
Athletic	0.630	0.796	0.165	0.368	***	0.032	
Financial Debt	0.612	0.870	0.258	0.483	***	0.025	
Romantic	0.610	0.317	-0.292	0.588	***	0.174	
Aggressive	0.504	0.780	0.276	0.546	***	0.032	
Charitable	0.470	0.326	-0.145	0.291	***	0.063	
Wealthy	0.348	0.722	0.374	0.810	***	0.032	
Breast Size	0.279	0.806	0.527	1.057	***	0.093	*
Social Status	0.271	0.654	0.382	0.837	***	0.054	
Outgoing	0.255	0.554	0.299	0.634	***	0.070	
Height	0.191	0.260	0.069	0.155		0.088	

Self-Reports: Differences between men's and women's impact scores

The four traits with the greatest gender differences in impact scores are femininity (Figure 4.2.a), masculinity (Figure 4.2.b), height (Figure 4.2.c) and financial debt (Figure 4.2.d). (Note that these differences are calculated as the mean of the absolute values of the differences between men and women across the four levels, and are not shown in Table 4.1, 4.2a, or 4.2b.) Interestingly, what drives these large differences varies across the four traits. Specifically, femininity and masculinity are very important to both men and women, but unsurprisingly, they have reversed shapes. Height is discrepant because it is important to women but hardly important at all to men. Financial debt is important to, and takes the same shape for, both genders, but is still more important to women than to men.

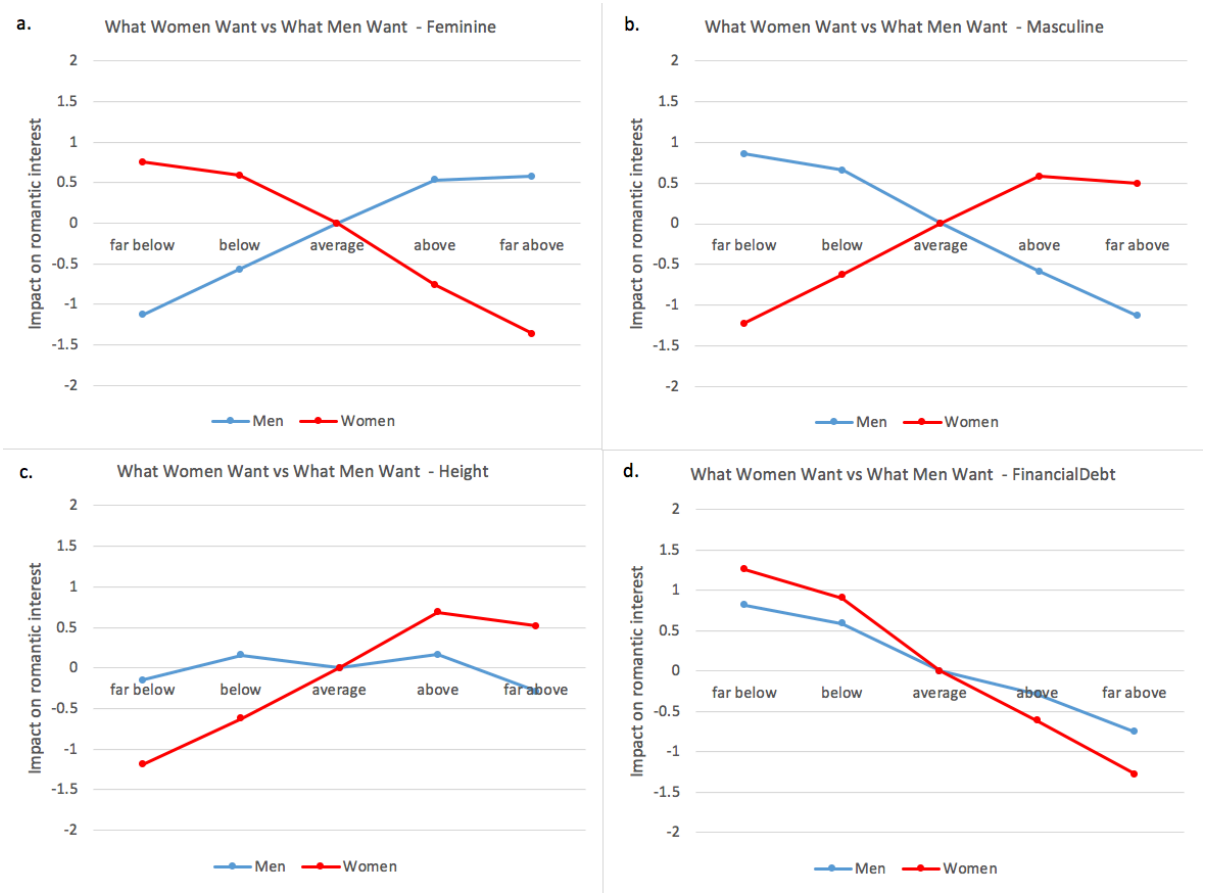

Figure 4.2. The impact on women's and men's romantic interest in a romantic partner as a function of his or her standing on four traits relative to average; as reported by women (red lines) and men (blue lines). Of the 22 traits covered, (a) through (d) show, in descending order, the four with the largest gender differences. (a) femininity, (b) masculinity, (c) height, and (d) financial debt.

Self-Reports: Discrepancies between men's and women's importance scores

The traits with the greatest gender differences in importance scores are height and financial debt (already shown in Figure 4.2), followed by how ambitious (Figure 4.3a) and sexual (Figure 4.3b) a romantic partner tends to be. Interestingly, while the shapes for ambition are the same for men and women, the shapes for sexuality differ between men and women. In Figure 4.3.d, we add in a plot for physical attractiveness because it is a trait often raised is discussions of gender differences (e.g., Li et al., 2002). Consistent with previous research, the current study finds that data from men rank physical attractiveness higher in

importance (3rd place among 22 traits) than data from women (12th place). See Table 4.2a and Table 4.2b. We return to a description of most and least important traits in the section on discrepancies between self-reported importance scores for one gender vs. those imagined by the opposite gender, below.

In addition to asking which individual traits are most discrepant, we asked whether the relative importance of each trait is similar between men and women by conducting a Pearson correlation between self-reported importance scores for men versus women. Results indicate significant correlation ($r(20) = .78$, $p < .001$). Thus, despite the above-described discrepancies for individual traits, the relative importance of traits is quite similar between genders.

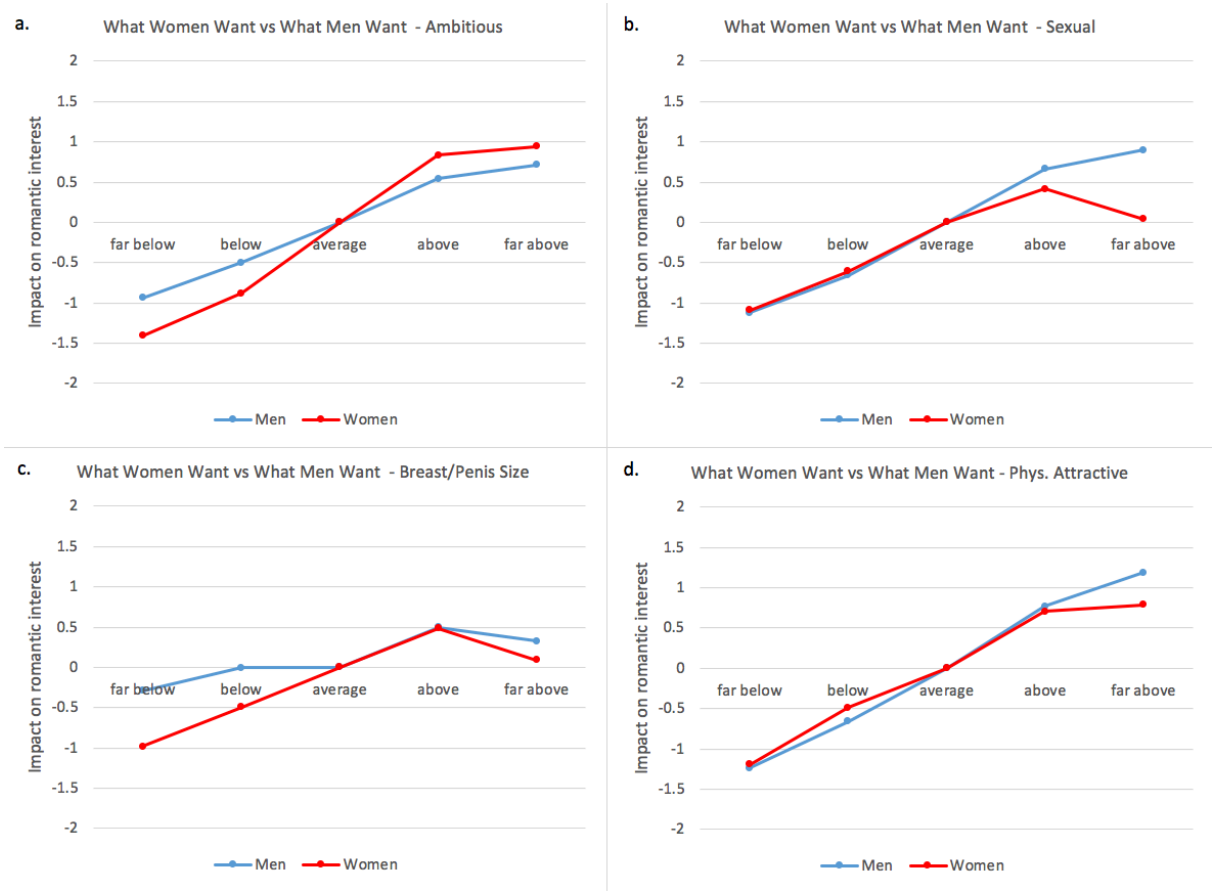

Figure 4.3. The impact on women's and men's romantic interest in a romantic partner as a function of his or her standing on four traits relative to average; as reported by women (red lines) and men (blue lines). Of the 22 traits covered, (a) through (d) show gender differences in the importance and shapes of (a) ambition, (b) how sexual, (c) height, and (d) financial debt.

Self-Reports: Discrepancies between men's and women's shapes

The traits with the greatest gender differences in shapes are masculinity and femininity (already shown in Figure 4.2), followed by sexuality (already shown in Figure 4.3) and breast/penis size (Figure 4.3.c).

Discrepancies between women's self-reported importance scores vs. those imagined by men for women

We begin with the traits most and least important to women and the traits men imagine are most and least important to women. We then examine the discrepancies in importance and shape between women's self-reported preferences and what men imagine are women's preferences. The four traits most important to women are trustworthiness (Figure 4.4.a), hygiene (Figure 4.4.b), intelligence (Figure 4.4.c), and warmth/friendliness (Figure 4.4.d). The four traits least important to women are sexuality (Figure 4.5.a), penis size (Figure 4.5.b), social status (Figure 4.5.c), and least of all, extraversion ("outgoing", Figure 4.5.d). The four traits men imagine are most important to women are trustworthiness, physical attractiveness, hygiene, and wealth. The four traits men imagine are least important to women are financial debt, sexuality, charitableness, and least of all, aggressiveness.

Note that in Figures 4.4 and 4.5, which plot data for traits that are most and least important to women, respectively, we also plot data for what men imagine for women, even though these traits may or may not be the ones that are most discrepant between the two genders. We do this simply to capitalize on showing more of our data. In the next section, which focuses on the largest discrepancies between women's self-reported importance scores vs. those imagined by men for women, we refer to already-presented figures that demonstrate these large discrepancies or we present new figures.

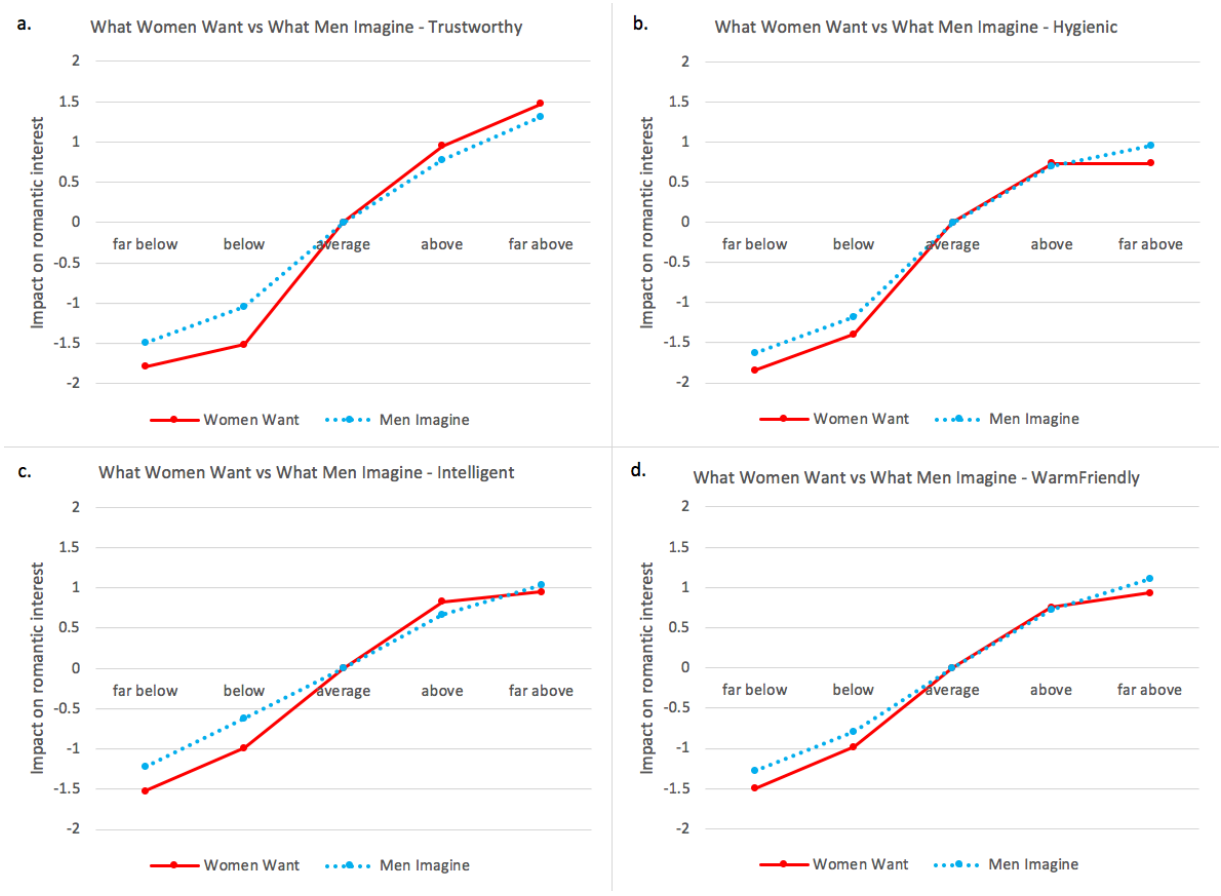

Figure 4.4. The impact on women's romantic interest in a hypothetical man as a function of his standing on four traits relative to average; as reported by women (solid lines) and as imagined by men (dotted lines). Of the 22 traits covered, (a) through (d) show, in descending order, the four that women indicated are most important. The impact of various levels of (a) trustworthiness, (b) hygiene, (c) intelligence, and (d) warmth/friendliness is shown. The differences between points on the solid and dotted lines show the discrepancy between the impacts that men imagine and that women report.

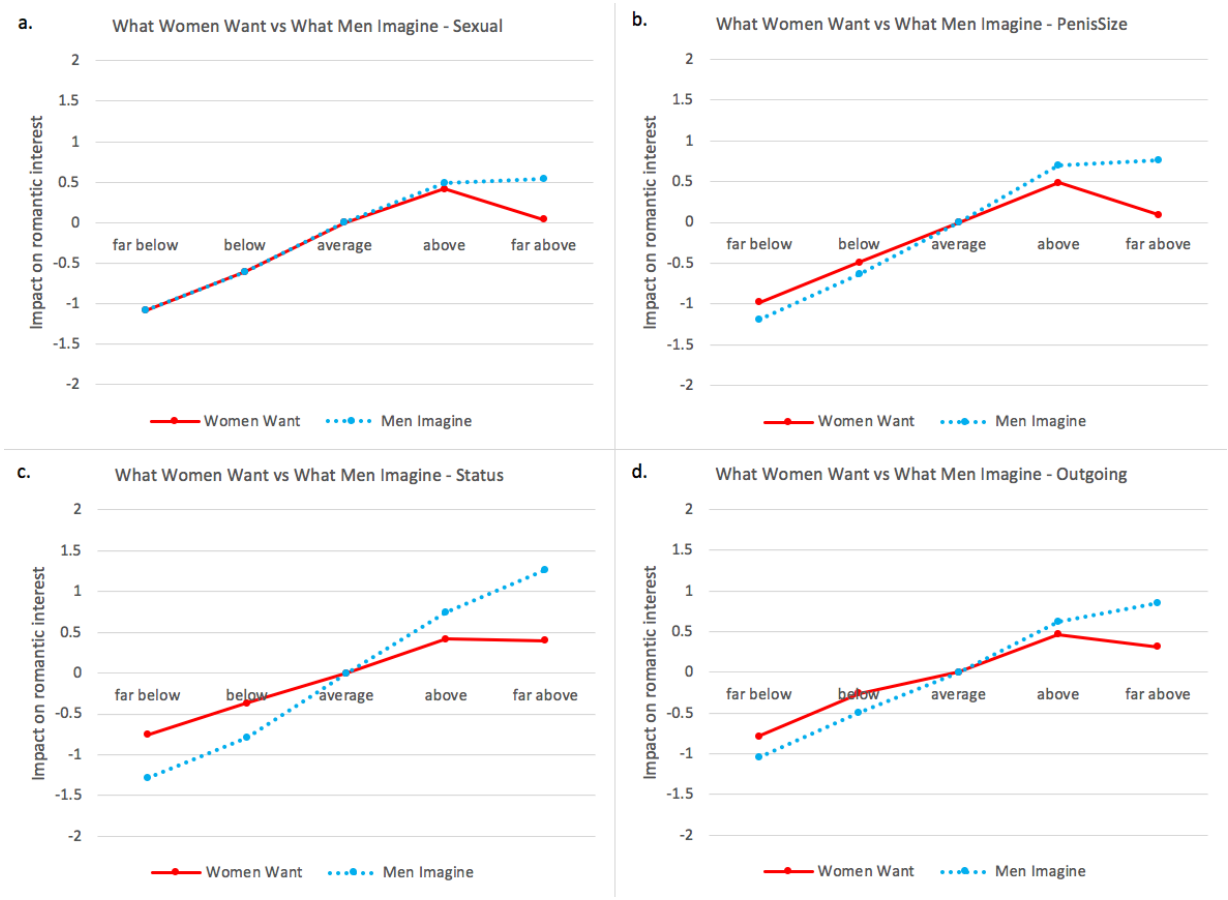

Figure 4.5. The impact on women's romantic interest in a hypothetical man as a function of his standing on four traits relative to average; as reported by women (solid lines) and as imagined by men (dotted lines). Of the 22 traits covered, (a) through (d) show, in descending order, the four that women indicated are least important. The impact of various levels of (a) sexuality, (b) penis size, (c) social status, and (d) extraversion ("outgoing") is shown. The differences between points on the solid and dotted lines show the discrepancy between the impacts that men imagine and that women report.

With regard to the traits with the largest *discrepancies* in importance, the two that men appear to *over*estimate the most (according to women's self-reports) are social status (Figure 4.5.c) and wealth (Figure 4.8.b; Figure 4.8 shows some of the largest overall or single-point discrepancies between self-reported and imagined impact that are not already included in Figures 4.4 through 4.7, which highlight importance rather than discrepancies). The two that men appear to underestimate the most are aggressiveness (Figure 4.8.a) and

trustworthiness (Figure 4.4.a). With regard to the traits with the largest discrepancies in shape, the two that men appear to most poorly estimate are penis size and sexuality (for both traits, men imagine that women's impact scores would peak and plateau between "somewhat above average" and "far above average", but women's impact scores are lower for "far above average" than for "somewhat above average").

Interestingly, when we looked at each level (ranging from far below to far above average) separately, the largest single-level discrepancies between women's self-reports and what men imagine for women involve men overestimating the positive impacts of being far above average in status (Figure 4.5.c) and far above average in athleticism (not shown in Figures). See Supplementary Materials.

Discrepancies between men's self-reported importance scores vs. those imagined by women for men

We begin with the traits most and least important to men and those that women imagine are most and least important to men. We then examine the discrepancies in importance and shape between men's self-reported preferences and what women imagine are men's preferences. The four traits most important to men are trustworthiness (Figure 4.6.a), hygiene (Figure 4.6.b), physical attractiveness (Figure 4.6.c), and intelligence (Figure 4.6.d). The four traits least important to men are breast size (Figure 4.7.a), social status (Figure 4.7.b), extraversion (Figure 4.7.c), and least of all, height (Figure 4.5.d). The four traits women imagine to be most important to men are physical attractiveness, sexuality, trustworthiness, and hygiene. The four traits women imagine are least important to men are education, charitableness, "romantic-ness", and least of all, height. Note that, as was the case in Figures 4.4 and 4.5, the plots showing traits that are most and least important to men also

plot data for what women imagine for men, in order to capitalize on showing more of our data.

With regard to the traits with the largest *discrepancies* in importance, the two that women appear to *over*estimate the most (according to men's self-reports) are breast size (Figure 4.7.a) and physical attractiveness (Figure 4.7.b). The two that women appear to most *under*estimate are intelligence (Figure 4.6.d) and education (Figure not shown). With regard to the traits with the largest discrepancies in *shape*, the two that women appear to most poorly estimate are intelligence and education (for both traits, women imagine that men's impact scores would decline between "somewhat above average" and "far above average", but men's impact scores are, in fact, greater for "far above average" than for "somewhat above average").

Interestingly, when we looked at each level (ranging from far below to far above average) separately, the largest single-level discrepancies between men's self-reports and what women imagine for men involve women overestimating the positive impacts of being far above average in breast size (Figure 4.7.a) and underestimating the positive impact of being far above average in intelligence (Figure 4.6.d). See Supplementary Materials.

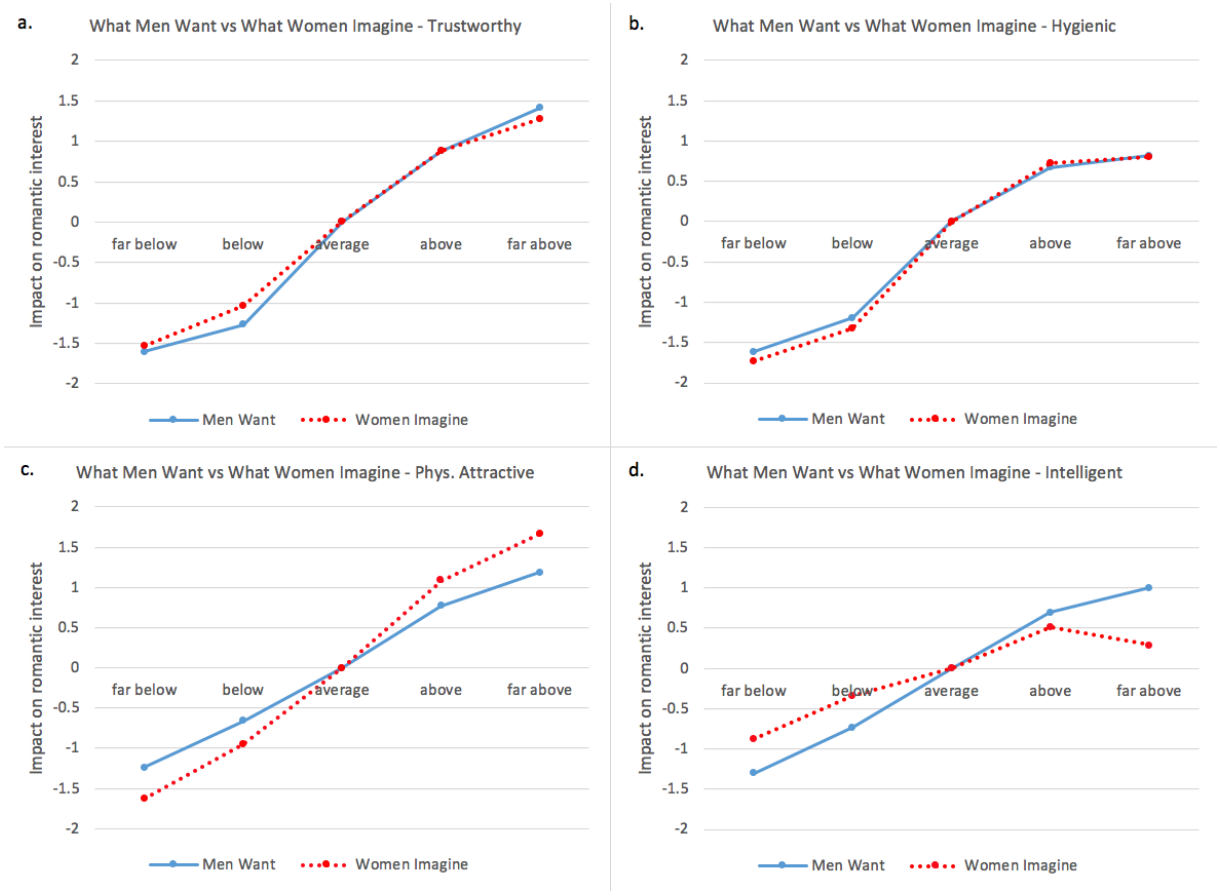

Figure 4.6. The impact on men's romantic interest in a hypothetical woman as a function of her standing on four traits relative to average; as reported by men (solid lines) and as imagined by women (dotted lines). Of the 22 traits covered, (a) through (d) show, in descending order, the four that men indicated are most important. The impact of various levels of (a) trustworthiness, (b) hygiene, (c) attractiveness, and (d) Intelligence is shown. The differences between points on the solid and dotted lines show the discrepancy between the impacts that women imagine and that men report.

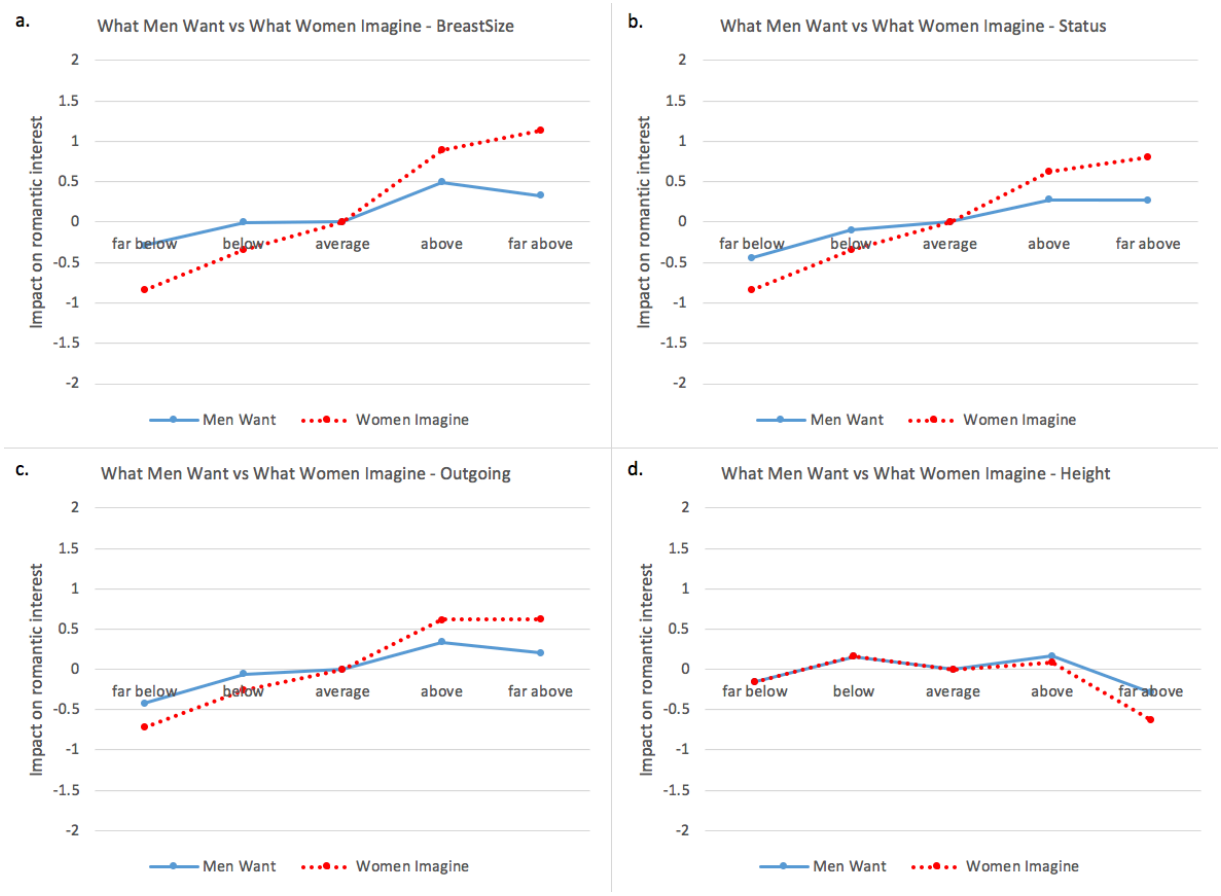

Figure 4.7. The impact on men's romantic interest in a hypothetical woman as a function of her standing on four traits relative to average; as reported by men (solid lines) and as imagined by women (dotted lines). Of the 22 traits covered, (a) through (d) show, in descending order, the four that men indicated are least important. The impact of various levels of (a) breast size, (b) social status, (c) extraversion ("outgoing"), and (d) height is shown. The differences between points on the solid and dotted lines show the discrepancy between the impacts that women imagine and that men report.

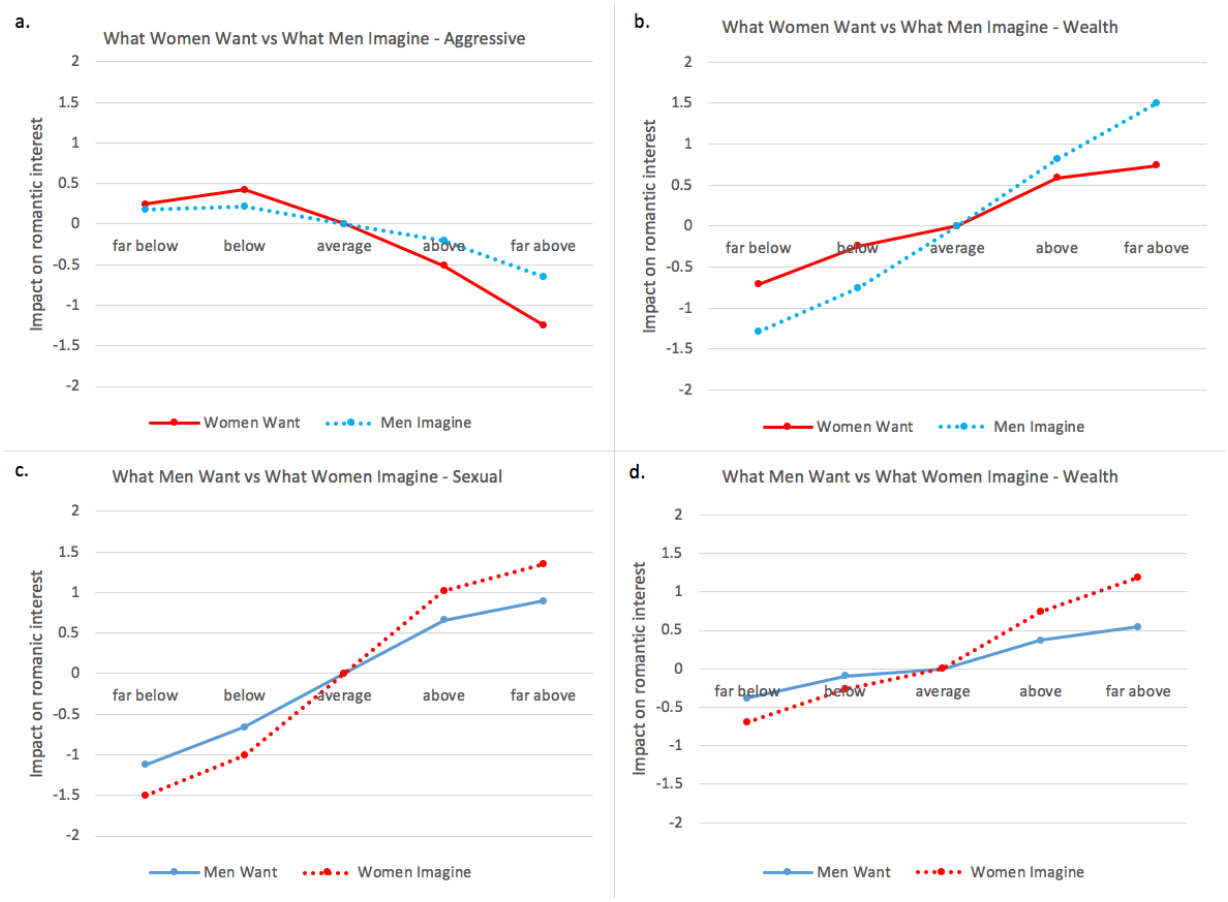

Figure 4.8. The impact on women's—(a) & (b)—and men's—(c) & (d)—romantic interest in a hypothetical romantic partner as a function of his or her standing on four traits relative to average; as self-reported (solid lines) and as imagined by the opposite gender (dotted lines). Traits (a) through (d) show some of the largest overall or single-point discrepancies between self-reported and imagined impact that are not already included in Figures 4.4 through 4.7, which highlighted importance rather than discrepancies. Graph (a) shows the impact of various levels of a man's aggressiveness on women's interest, (b) shows the impact of a man's wealth on women's interest, (c) shows the impact of a woman's sexuality on men's interest, and (d) shows the impact of a woman's wealth on men's interest. The differences between points on the solid and dotted lines show the discrepancy between the imagined and real impacts on romantic interest.

Which gender's beliefs are closer to the opposite gender's self-reports?

We begin by comparing men's and women's average absolute discrepancy between self-reported and imagined importance scores across all 22 traits (i.e., comparing the absolute values of the "importance discrepancy" scores in Table 4.2a and Table 4.2b). Averaged across the 22 traits, the mean for women (.245 +/- 1.34) and men (.210 +/- 1.53) is similar

($t(42) = .786$, p = .436, two-tailed test, $d = .237$) suggesting that one gender's imagined importance scores are no less discrepant than the other's. We then did the same analysis for shape discrepancies. Averaged across the 22 traits, the mean shape discrepancy for men imagining women (0.122 +/- .039) is significantly higher than for women imagining men (.077 +/- .073) ($t(42) = 2.51$, $p = .016$, two-tailed test, $d = .757$). Additionally, the proportion of traits for which the imagined and self-reported shapes are significantly discrepant (i.e., $p < .05$) is significantly higher for men imagining women (13 of 22, 59.1%) than women imagining men (3 of 22, 13.6%) (X^2 (1) = 9.82, $p = .002$, $\varphi = .472$). This pair of results suggests that, in terms of imagining the zigzag shape of the opposite gender's preferences, women are more accurate, according to men's self-reports.

Finally, to investigate how well each gender understands the relative importance of each trait to the opposite gender, we conducted Pearson correlations between self-reported importances of traits to men vs. those imagined for men by women, and vice versa, which is presented in Figures 4.9.A (women imagining men) and 4.9.B (men imagining women). While the correlation for women imagining men is significant ($r(20) = .62$, $p < .002$), the correlation for men imagining women is not ($r(20) = .39$, $p = .080$, although this is marginally significant). Therefore, despite the fact that women are no more or less accurate than men when it comes to imagining absolute importance scores (see above), this correlation result suggests that women may be better (according to the comparison with each gender's self-reports) at imagining the *relative importance of traits* for men.

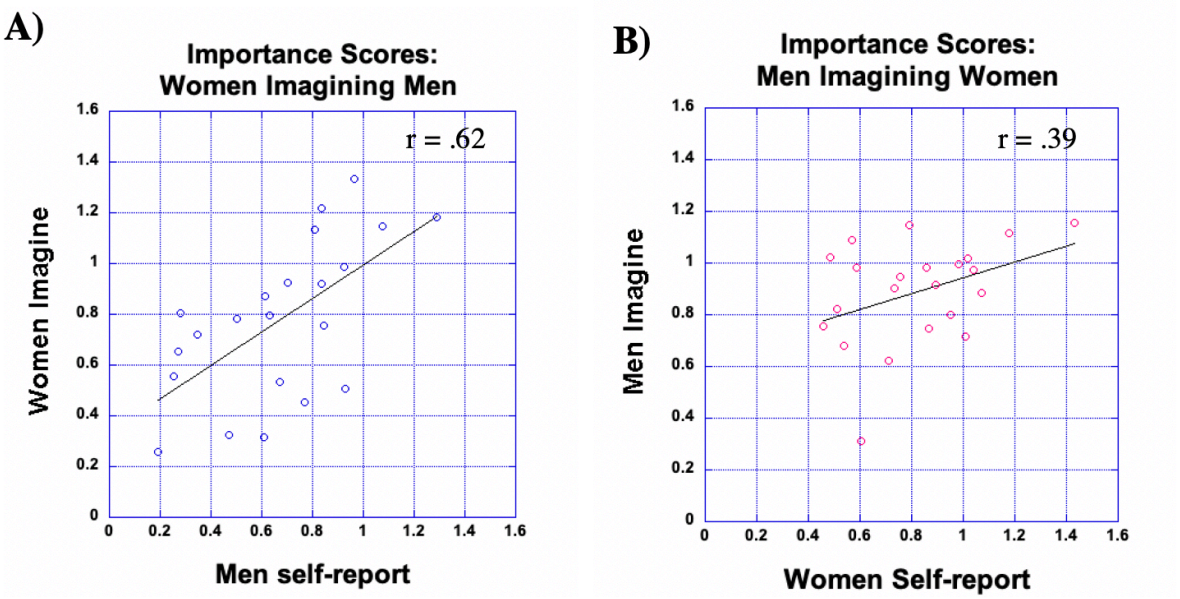

Figure 4.9. Plots showing correlation between men's self-reported importance scores vs. what women imagine for men (4.9.A) and women's self-reported importance scores vs. what men imagine for women (4.9.B).

We entertained the possibility that the larger correlation for women imagining men than men imagining women is explainable by a difference in variance between men and women in importance scores from self-report. However, the variance in men's importance scores (s^2_{men} = 0.0845) is not significantly different from that in women's importance scores (s^2_{women} = 0.0641) ($F(21)$ = .759, p = .267, one-tailed test based on the prediction that women would be lower than men). We also measured variance *within* each trait for men and women, and found that the mean variance across the 22 traits was similar for men (.249 +/- .038) and women (.254 +/- .049) (t(42) = .429, p = .670, two-tailed test, d = .129).

Interestingly, because there is a high correlation between men's and women's self-reported importance scores (r = 0.78, see above), both men and women could better imagine the *relative importance of traits* for the opposite gender if they were to substitute their own

preferences for those they imagine to be true for the opposite gender. Also, despite the correlation between men's and women's self-reported importance scores, there is no correlation between the importance scores men imagine for women and those that women imagine for men, with ($r(20) = .20$, $p = .381$).

Discussion

Consistent with previous research (e.g., Buss & Barnes, 1986; Fletcher et al. 1999, Lippa, 2007), our analyses of self-reports show that trustworthiness, warmth, and intelligence are among the most important trait dimensions to both men and women alike. Our results also show support for classic gender differences (e.g., Buss, 1989; Fletcher, Tither, O'Loughlin, Friesen, & Overall, 2004; Geary, 2010; Kenrick & Keefe, 1992; Li, Bailey, Kenrick, & Linsenmeiser, 2002; Lippa 2000; Sprecher, Sullivan, & Hatfield, 1994), wherein physical attractiveness is particularly important to men, but less so to women, and ambition is particularly important to women, but not to men. Interestingly, however, wealth and status—traits that are sometimes related to ambition (e.g., Judge & Kammeyer-Mueller, 2012)—rank low in importance to both genders. There were two other places of note where we observed something unexpected. First, extraversion ("outgoing") ranked lowest in importance to women, and second lowest in importance to men. This is despite this trait having received a relatively high importance rating in a study conducted in the late nineties (Fletcher, Simpson, & Giles, 1999). Second, despite the fact that traits such as penis/breast size and social status receive a large amount of cultural attention (e.g., Herzog, 2018; Martin, 2013), they are quite low-ranking in importance for the men and women in our study.

The novel aspect of the current study is that it compared self-reported preferences with preferences imagined by the opposite gender, within the same population. Generally,

traits that previous literature shows men and women typically emphasize to impress each other and derogate rivals (wealth and status displayed by men, physical attractiveness displayed by women: Buss 1988, Buss & Dedden 1990, Campos et al., 2002; Cicerello & Sheehan, 1995) are, in our results, the same traits whose importance to the opposite gender men and women seem to most overestimate. However, this characterization of the comparison between previous literature and our results is overly simplistic. If each gender's self-reported preferences act as our guide, then despite women overestimating the importance of physical attractiveness to men, they are certainly not misguided in their efforts to impress men along this dimension. Indeed, men's self-reports indicate that a woman's physical attractiveness is indeed among the most important traits for them. Nevertheless, by overestimating that trait's importance, women may inadvertently emphasize their physical attractiveness at the expense of opportunities to highlight other similarly important traits, the importance of which they underestimate; namely intelligence and trustworthiness. Similarly, men's efforts to signal ambition to attract women are not misguided. Indeed, women's self-reports indicate that a man's ambition is important, and men neither underestimate nor overestimate it. By contrast, however, men's efforts to signal wealth and status do seem to be misguided; men overestimate the importance of these traits to women more so than they do for any other trait. Men may emphasize their wealth and status at the expense of highlighting traits that are more important, and the importance of which they underestimate; namely intelligence and trustworthiness.

Secondary analyses revealed that what women imagine is true for men is less discrepant than what men imagine for women (based on the opposite gender's self-report). Specifically, shape discrepancies are lower for women imagining men than vice versa, and

the correlation between self-reported and imagined importance (across traits) is greater for women imagining men than vice versa. This suggests that women, compared to men, better understand the relative importance of each trait to the opposite gender, as well as the character of its impact. Additional analyses suggested that our results did not derive from differences in variance between men's and women's preferences (see results). A number of other reasons may lead to this result, not the least of which is the possibility that women are more familiar with men's preferences than vice versa because of greater interpersonal accuracy.

Although little is known about potential gender differences regarding inferring or otherwise coming to know others' preferences, in the case of measures of interpersonal accuracy, women outperform men (Hall, Gunnery, Horgan, 2016). These measures have included emotion judgments (e.g., Hall 1978; Thompson & Voyer, 2014) state and trait inferences (e.g., Murphy, Hall, & Colvin, 2003, Letzring, 2010; Vogt & Colvin, 2003) and appearance and behavior recall (e.g., Hall & Schmid Mast, 2008; Horgan, McGrath, & Long, 2009, Hall, Murphy, & Mast 2006). Across multiple modalities (e.g., voice, body, face), and across time, culture, age group, and target gender, females more accurately and more quickly infer the meanings of affective cues. In addition, research on accuracy in reading romantic interest in behavior (e.g., Abbey, 1982,), suggests that men substantially overestimate women's sexual interest, frequently perceiving interest from women where none exists (Farris, Treat, Viken, & McFall, 2008). Men's bias (Buss & Haselton, 2000) or greater relative difficulty in reading women's interest could stifle men's ability to learn about what appeals to women.

An alternative interpretation of the discrepancies in our data is based on questioning the veracity of the self-reports. It is possible that men and women do *not* misestimate, but that instead the opposite gender misreports. Misreporting may be unintentional if people are simply mistaken about their own preferences (see Perilloux and Kurzban, 2015 for a discussion of this possibility, but see Li et al., 2013 for evidence that dating behaviors reflect self-reported preferences). Alternatively, misreporting may be intentional in an effort to be coy. For example, a persistent cultural belief holds that it is "unladylike" for women to be overtly sexual (Berbary, 2012). This notion could potentially encourage women to under-report the importance to them of traits such as penis size and sexuality. In such a case, men, more so than women, would be expected to overestimate the importance of these traits to the opposite gender. Yet, according to results of the current study, men overestimate the importance of these dimensions *less* so than do women. In addition, we note that women self-report that penis size is more important than do men for breast size, and so they do not appear coy in reporting their sexual preferences (although women do report sexuality to be less important than do men). This pattern of results seems to suggest that men's underperformance is not simply due to women under-reporting the importance of traits related to sex.

Even if women do not misreport their preferences, they may still be less than forthright in interpersonally expressing interest out in the real world. One study (Place, Todd, Penke, & Asendorpf, 2009) found it to be more difficult for both males and females to accurately read romantic interest cues of women than of men, suggesting that the difference is due, in part, to women being less expressive in social settings about their interest in men.

This tendency, too, would potentially stifle men's ability to learn about what appeals to women.

Another possibility is that the media misrepresents women's preferences. Popular books and movies could mislead men by, for example, centering romance narratives around men of high status and wealth; examples include *Pretty Woman* (1990), *Fifty Shades of Gray* (2015), and *Crazy Rich Asians* (2018). However, there is no shortage of media depicting men's interest in physically attractive women, which seems a likely contributing cause of women's tendency to overestimate the importance of traits like breast-size and physical appearance more generally. Still, to account for the underperformance of men imagining women in the current study, it may be that the media's depiction of men's romantic interests is closer to the truth than is its depiction of women's interests.

The striking results and broad applicability of our methods notwithstanding, the current study has important limitations. While our set of traits is wide-reaching, it is certainly not exhaustive. There may be other traits for which the differences between men's and women's perceptions and self-reports are unlike the differences shown here. In addition, our sample of online participants, which includes residents from all over the United States, may differ in important ways from the broader population of American adults. We nevertheless would argue that the results are generalizable enough to be insightful.

A further limitation is that we use only one conception of importance while other reasonable conceptions are possible. For example, a different conception of importance could aim to provide increased weight to more central levels of each trait given that levels nearer to the mean average are likely to occur with greater frequency under a normal distribution. This

way, greater importance would reflect greater frequency of impact rather than reflecting total impact integrated uniformly across each dimension.

A final noticeable limitation is that we pose each question to participants within the context of an implicit all-else-being-equal approach. This has the advantage of isolating each variable, but it also means that we cannot capture interactions that may exist between different levels of different traits. For example, the importance of wealth might differ depending on somebody's physical attractiveness, such that a person's overall desirability could be less dependent on their wealth if they are good-looking. Interactions of this sort may be a fruitful avenue for future research in the area.

Chapter 4, in full, has been submitted for publication of the material as it may appear in Journal of Social and Personal Relationships, 2020, Jago, Carl P.; Dobkins, Karen R., Sage Publications, 2020. The dissertation author was the primary investigator and author of this paper.

References

Abbey, A. (1982). Sex differences in attributions for friendly behavior: Do males misperceive females' friendliness?. *Journal of Personality and Social Psychology*, *42*(5), 830.

Berbary, L. A. (2012). "Don't Be a Whore, That's Not Ladylike" Discursive Discipline and Sorority Women's Gendered Subjectivity. Qualitative Inquiry, 18(7), 606-625.

Buss, D. M. (1988). The evolution of human intrasexual competition: tactics of mate attraction. *Journal of Personality and Social Psychology*, *54*, 616–628.

Buss, D. M. (1989). Sex differences in human mate preferences: evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences*, *12*, 1–49.

Buss, D. M., & Barnes, M. (1986). Preferences in human mate selection. *Journal of personality and social psychology*, *50*(3), 559.

Buss, D. M., & Dedden, L. A. (1990). Derogation of competitors. *Journal of Social and Personal Relationships*, *7*, 395–422.

Buss, D.M., Abbott, M., Angleitner, A., Asherian, A., Biaggio, A., Blanco-Villasenor, A., Bruchon-Schweitzer, M., Ch'u, H.Y., Czapinski, J., Deraad, B. and Ekehammar, B., (1990). International preferences in selecting mates: a study of 37 cultures. *Journal of Cross-Cultural Psychology*, *21*, 5–47.

Campbell, L., & Wilbur, C. J. (2009). Are the traits we prefer in potential mates the traits they value in themselves? An analysis of sex differences in the self-concept. *Self and Identity*, *8*(4), 418-446.

Campos, L., Otta, E., & de Oliveira Siqueira, J. (2002). Sex differences in mate selection strategies: content analyses and response to personal advertisements in Brazil. *Evolution and Human Behavior*, *23*, 395–406.

Cicerello, A., & Sheehan, E. P. (1995). Personal advertisements: a content analysis. *Journal of Social Behavior and Personality*, *10*, 751–756.

Crocker, J., Luhtanen, R. K., Cooper, M. L., & Bouvrette, A. (2003). Contingencies of self-worth in college students: theory and measurement. *Journal of Personality and Social Psychology*, *85*, 894–908.

Daly, J. A., Hogg, E., Sacks, D., Smith, M., & Zimring, L. (1983). Sex and relationship affect social self-grooming. *Journal of Non-verbal Behavior*, *7*, 183–189.

Farris, C., Treat, T. A., Viken, R. J., & McFall, R. M. (2008). Sexual coercion and the misperception of sexual intent. *Clinical psychology review*, *28*(1), 48-66.

Fletcher, G. J., Simpson, J. A., Campbell, L., & Overall, N. C. (2013). *The science of intimate relationships* (Vol. 323). Malden, MA: Wiley-Blackwell.

Fletcher, G. J. O., Simpson, J. A., Thomas, G., & Giles, L. (1999). Ideals in intimate relationships. *Journal of Personality and Social Psychology*, *76*, 72–89.

Fletcher, G. J. O., Tither, J. M., O'Loughlin, C., Friesen, M., & Overall, N. (2004). Warm and homely or cold and beautiful? Sex differences in trading off traits in mate selection. *Personality and Social Psychology Bulletin*, *30*, 659–672.

Geary, D. C. (2010). *Male, Female: The Evolution of Human Sex Differences* (2nd ed). Washington, DC: American Psychological Association.

Gutierres, S. E., Kenrick, D. T., & Partch, J. J. (1999). Beauty, dominance, and the mating game: contrast effects in self-assessment reflect gender differences in mate selection. *Personality and Social Psychology Bulletin*, *25*, 1126–1134.

Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological bulletin*, *85*(4), 845.

Hall, J. A., Gunnery, S. D., & Horgan, T. G. (2016). 15 Gender differences in interpersonal accuracy. *The social psychology of perceiving others accurately*, 309.

Hall, J. A., Murphy, N. A., & Mast, M. S. (2006). Recall of nonverbal cues: Exploring a new definition of interpersonal sensitivity. *Journal of Nonverbal Behavior*, *30*(4), 141-155.

Hall, J. A., & Schmid Mast, M. (2008). Are women always more interpersonally sensitive than men? Impact of goals and content domain. *Personality and Social Psychology Bulletin*, *34*(1), 144-155.

Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of personality and social psychology*, *78*(1), 81.

Herzog, S. (2018, July 10). Professor speaks out after shutting down penis size study amid backlash. *USA Today*. Retrieved from https://www.usatoday.com/story/news/nation-now/2018/07/10/professor-shuts-down-penis-size-study-after-receiving-hate-mail/773429002/

Horgan, T. G., McGrath, M. P., & Long, J. A. (2009). The relevance of people versus objects in explaining females' advantage over males in appearance accuracy. *Sex Roles*, *60*(11-12), 890-899.

Josephs, R. A., Markus, H. R., & Tafarodi, R. W. (1992). Gender and self-esteem. *Journal of Personality and Social Psychology*, *63*, 391–402.

Judge, T. A., & Kammeyer-Mueller, J. D. (2012). On the value of aiming high: The causes and consequences of ambition. Journal of Applied Psychology, 97(4), 758.

Kenrick, D. T., & Keefe, R. C. (1992). Age preferences in mates reflect sex differences in human reproductive strategies. *Behavioral and Brain Sciences*, *15*, 75–133.

Letzring, T. D. (2010). The effects of judge-target gender and ethnicity similarity on the accuracy of personality judgments. *Social Psychology*.

Li, N. P., Bailey, J. M., Kenrick, D. T., & Linsenmeiser, J. A. W. (2002). The necessities and luxuries of mate preferences: testing the tradeoffs. *Journal of Personality and Social Psychology*, *82*, 947–955.

Lippa, R. A. (2000). Gender-related traits in gay men, lesbian women, and heterosexual men and women: the virtual identity of homosexual-heterosexual diagnosticity and gender diagnosticity. *Journal of Personality*, *68*, 899–926.

Lippa, R. A. (2007). The preferred traits of mates in a cross-national study of hetero- sexual and homosexual men and women: an examination of biological and cultural influences. *Archives of Sexual Behavior*, *36*, 193–-208.

Martin, R. D. (2013, January 13). Penis Size Matters. *Psychology Today*. Retrieved from https://www.psychologytoday.com/us/blog/how-we-do-it/201501/penis-size-matters

Murphy, N. A., Hall, J. A., & Colvin, C. R. (2003). Accurate intelligence assessments in social interactions: Mediators and gender effects. *Journal of Personality*, *71*(3), 465-493.

Perilloux, C., & Kurzban, R. (2015). Do men overperceive women's sexual interest?. *Psychological Science*, *26*(1), 70-77.

Place, S. S., Todd, P. M., Penke, L., & Asendorpf, J. B. (2009). The ability to judge the romantic interest of others. *Psychological Science*, *20*(1), 22-26.

Sprecher, S., Sullivan, Q., & Hatfield, E. (1994). Mate selection preferences: gender differences examined in a national sample. *Journal of Personality and Social Psychology*, *66*, 1074–1080.

Thompson, A. E., & Voyer, D. (2014). Sex differences in the ability to recognise non-verbal displays of emotion: A meta-analysis. *Cognition and Emotion*, *28*(7), 1164-1195.

Li, N. P., Yong, J. C., Tov, W., Sng, O., Fletcher, G. J., Valentine, K. A., ... & Balliet, D. (2013). Mate preferences do predict attraction and choices in the early stages of mate selection. Journal of Personality and Social Psychology, 105(5), 757.

Vogt, D. S., & Randall Colvin, C. (2003). Interpersonal orientation and the accuracy of personality judgments. *Journal of personality*, *71*(2), 267-295.